*Article*

# Hierarchical Classification of Soybean in the Brazilian Savanna Based on Harmonized Landsat Sentinel Data

Taya Cristo Parreiras [1,*], Édson Luis Bolfe [1,2], Michel Eustáquio Dantas Chaves [3], Ieda Del'Arco Sanches [3], Edson Eyji Sano [4], Daniel de Castro Victoria [2], Giovana Maranhão Bettiol [4] and Luiz Eduardo Vicente [5]

1    Institute of Geosciences, State University of Campinas (Unicamp), Campinas 13083-855, Brazil
2    Brazilian Agricultural Research Corporation (Embrapa Agricultura Digital), Campinas 70770-901, Brazil
3    National Institute for Space Research (INPE), São José dos Campos 12227-010, Brazil
4    Brazilian Agricultural Research Corporation (Embrapa Cerrados), Planaltina 73301-970, Brazil
5    Brazilian Agricultural Research Corporation (Embrapa Meio Ambiente), Jaguariúna 13820-000, Brazil
*    Correspondence: tayacristo1@gmail.com

**Abstract:** The Brazilian Savanna presents a complex agricultural dynamic and cloud cover issues; therefore, there is a need for new strategies for more detailed agricultural monitoring. Using a hierarchical classification system, we explored the Harmonized Landsat Sentinel-2 (HLS) dataset to detect soybean in western Bahia, Brazil. Multispectral bands (MS) and vegetation indices (VIs) from October 2021 to March 2022 were used as variables to feed Random Forest models, and the performances of the complete HLS time-series, HLSS30 (harmonized Sentinel), HLSL30 (harmonized Landsat), and Landsat 8 OLI (L8) were compared. At Level 1 (agricultural areas × native vegetation), HLS, HLSS30, and L8 produced identical models using MS + VIs, with 0.959 overall accuracies (OA) and Kappa of 0.917. At Level 2 (annual crops × perennial crops × pasturelands), HLS and L8 achieved an OA of 0.935 and Kappa > 0.89 using only VIs. At Level 3 (soybean × other annual crops), the HLS MS + VIs model achieved the best performance, with OA of 0.913 and Kappa of 0.808. Our results demonstrated the potential of the new HLS dataset for medium-resolution mapping initiatives at the crop level, which can impact decision-making processes involving large-scale soybean production and agricultural sustainability.

**Keywords:** Cerrado; agriculture monitoring; remote sensing; multisensor; *Glycine max* L.; HLS

## 1. Introduction

Brazil is one of the world's most important food producers and exporters, contributing significantly to meet the global growing demand [1]. For example, the national harvested area of soybeans (*Glycine max* L.) in 2020 was more than 37 million ha, higher than that from the United States, with approximately 33 million ha [2]. In the 2029/2030 crop growing season, we expect Brazil to produce 318 million tons of grains in addition to soybeans, especially maize (*Zea mays* L.), wheat (*Triticum aestivum* L.), beans (*Phaseolus vulgaris* L.), rice (*Oryza sativa* L.), coffee (*Coffea* sp.), cotton (Gossypium sp.), oatmeal (*Avena sativa*), sorghum (*Sorghum bicolor* L.), peanut (*Arachis hypogaea* L.), sunflower (*Helianthus annuus* L.), and canola (*Brassica* sp.), an increase of about 27% in relation to the 2019/2020 crop season [3].

The Brazilian tropical savanna region (Cerrado biome), occupying more than 200 million ha in the central part of the country, is the major producer of food and energy in the country [4]. It contributes to 52% of the national soybean production, 54% of maize, 96% of cotton, and 51% of sugarcane (*Saccharum officinarum* L.) [5]. Most of the grain production in the Cerrado is found in extensive flat terrains (plateaus) [6]. These are the cases, for example, of the western region of the Bahia State (Western Bahia), in the Cerrado/semi-arid transition [7,8]; of the municipalities of Lucas do Rio Verde, Sinop, and Sorriso, in the Cerrado-Amazon ecotone of the Mato Grosso State [9]; and of the MATOPIBA region, an

acronym relating the Cerrado portion of four states (Maranhão, Tocantins, Piauí, and Bahia) located in the northern part of the biome [10].

Monitoring the rainfed agricultural production, which accounts for approximately 75% of global food staple [11], using optical satellite remote sensing data is difficult because of the persistent cloud cover conditions during the wet season [12,13], the spectral similarity between some crops (for example, soybean and bean), and varying planting dates. A multisensor approach for increasing the periodicity of satellite overpasses and for taking advantage of complementary information from multiple sources of satellite data can optimize the capability of rainfed crop-type identification [14]. Perhaps one of the most well-known satellite fusion methods is the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) that was developed by Gao et al. [15]. STARFM products are made by fusing Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) data sets and have been widely used for crop monitoring [16–18]. According to Zhu et al. [19], other highly adopted fusion models are the Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model (ESTARFM) [20] and the Spatial Temporal Adaptive Algorithm for mapping Reflectance Change (STAARCH) [21], also involving fusion of Landsat and MODIS data sets.

The drawback of these models is the need for the availability of cloud-free Landsat and MODIS images on matching dates, which is not always achievable in practice [14]. Another issue is the necessity of careful image pre-processing to harmonize sets of images acquired at different spatial, spectral, and radiometric resolutions, view angles, and signal-to-noise ratios. Recently, NASA proposed the Harmonized Landsat and Sentinel-2 (HLS) initiative to produce a virtual constellation of analysis-ready, two to four day's temporal frequency, surface reflectance images acquired by the Landsat Operational Land Imager (OLI) and Sentinel-2 Multispectral Instrument (MSI) sensors [22]. This initiative relies on a set of algorithms for atmospheric correction, cloud and cloud shadow masking, spatial co-registration, common gridding, bidirectional reflectance distribution function normalization, and spectral bandpass adjustments [23–25]. Previous studies have demonstrated the potential of the HLS time series for improving crop type mapping [26], crop intensity [27] and biomass [28], crop phenological metrics estimation [29], irrigated area detection [30], livestock intensification indication [31,32], and surface phenology characterization [33,34].

Since 2020, the entire Brazilian territory has been covered by the HLS v2.0 collection. Despite its potential for providing valuable capability to monitor the spatio-temporal dynamics of tropical agriculture, to our best knowledge, detailed studies evaluating the potential and accuracy of HLS data sets to identify croplands in the Cerrado biome are still in initial stages. Within this context, our study aims to analyze the potential of HLS images to identify soybean plantations from other representative Land-Use Land-Cover (LULC) classes in the Cerrado biome using a hierarchical classification system. Ultimately, attempting to separate soybeans, the most important agricultural commodity in Brazil, from other annual crops found in the Cerrado is the main motivation of this study.

## 2. Materials and Methods

### 2.1. Study Area

The selected study area corresponds to the region located in Western Bahia, Brazil (south latitude: 11°45′; west longitude: 45°50′), surrounded by the BR-242, BA-459, and BA-460 highways and covering an area of 366 thousand ha (Figure 1). The area partially covers the municipalities of Barreiras, Luís Eduardo Magalhães, and Riachão das Neves and is known as *Soybean Ring* because of the intensive production of soybean in this region [35].

The region is characterized by the tropical continental climate, i.e., Aw in the Köppen climate classification system [36], with strong climate seasonality: six months of dry season (April to September) and six months of wet season (October to March). According to the rainfall data provided by an automatic station located in the municipality of Luís Eduardo Magalhães and recorded by the National Institute of Meteorology [37] (station code: A404; latitude = −12.15°; longitude = −45.83°; elevation = 760 m; operation date = 17 April

2002), the average annual rainfall is 885 mm. This rainfall gradient increases from east to west. According to the data gathered by the Shuttle Radar Topography Mission (SRTM), the topography of the study area is flat, with an average elevation of 788 m, while highly weathered Oxisols, with low natural fertility and high Al toxicity, are the dominant soil type [38].
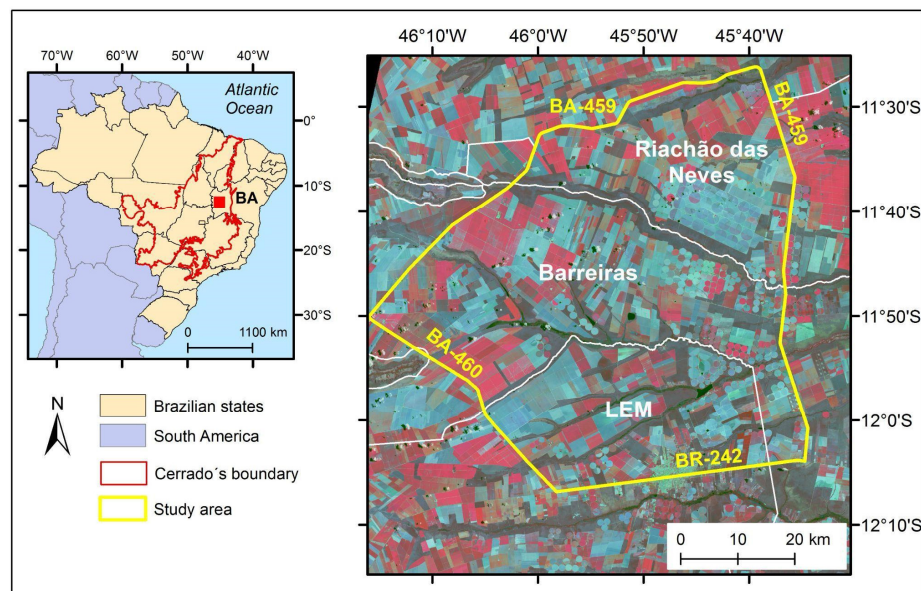


**Figure 1.** Location of the study area in the Bahia State, Brazil (inset figure), limited by the BR-242, BA-459, and BA-460 highways. The RGB color composite corresponds to the Landsat 9 OLI image (path = 220; row = 68), bands 5, 6, and 4, acquired on 1 March 2022. BA = Bahia State; LEM = municipality of Luís Eduardo Magalhães.

Since 1985, this region has faced a rapid process of conversion of native vegetation into large-scale mechanized agriculture for exportation [7]. The main rainfed crop found in the study area is soybean, followed by cotton and maize (Table 1). Coffee and bean are the most important irrigated crops by center-pivot irrigation systems. In the Landsat 8 OLI image from October 2021, we counted 349 center-pivots in the study area.

**Table 1.** Estimated crops harvested area (ha) from the municipalities of Barreiras, Luís Eduardo Magalhães, and Riachão das Neves in the 2020–2021 crop growing season.

| Municipality | Cropping Pattern | Crop Type | Harvested Area (ha) |
|---|---|---|---|
| Barreiras | Annual | Soybean | 195,500 |
| | | Maize | 18,598 |
| | | Cotton | 23,855 |
| | | Others (beans, sorghum, sugarcane) | 16,435 |
| | Perennial | Coffee, banana (*Musa* spp.), orange (*Citrus sinensis* L.), papaya (*Carica papaya* L.) | 6364 |
| Luís Eduardo Magalhães | Annual | Soybean | 162,200 |
| | | Maize | 14,600 |
| | | Cotton | 16,513 |
| | | Others (beans, sorghum, wheat) | 22,268 |
| | Perennial | Coffee, banana, orange, papaya | 1451 |
| Riachão das Neves | Annual | Soybean | 116,500 |
| | | Maize | 12,200 |
| | | Cotton | 32,895 |
| | | Others (beans, sorghum, cassava) | 7973 |
| | Perennial | Coffee, banana, orange, papaya | 1175 |

Source: Brazilian Institute of Geography and Statistics (IBGE) [35].

## 2.2. Methods

Our study can be characterized as an applied research with a qualitative–quantitative approach, using technical procedures based on a study case, and Figure 2 summarizes the proposed methodology. The Landsat 8 (L8) and the Harmonized Landsat Sentinel-2 (HLS) images acquired during the 2021–2022 crop growing season were classified by the non-parametric Random Forest (RF) classifier using training samples collected during the field campaign conducted in December 2021. The main steps of the proposed method are detailed in the following subsections.
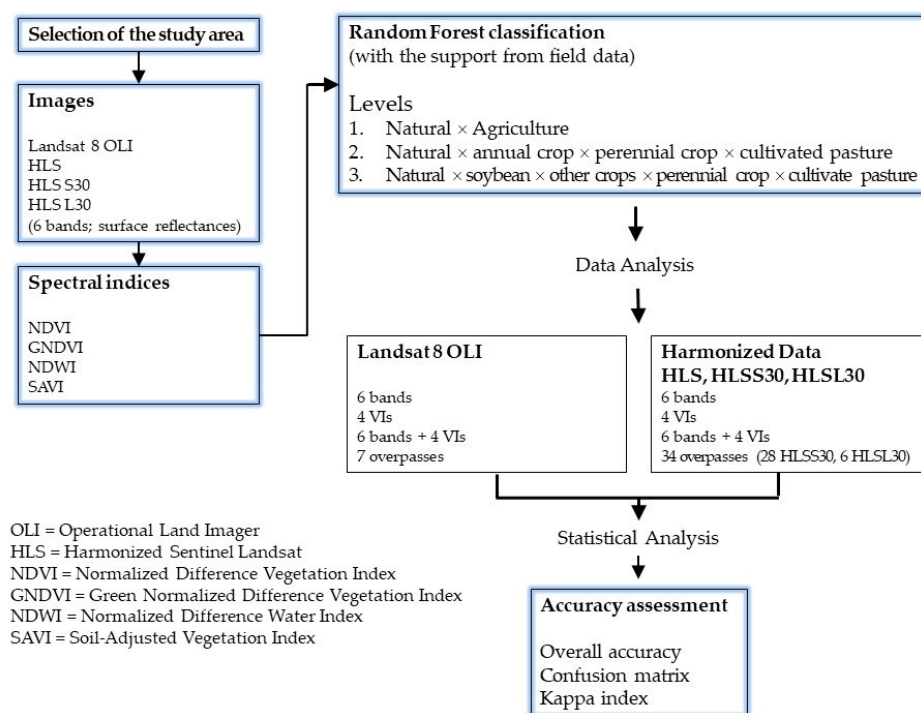


**Figure 2.** Flowchart showing the main steps of the method proposed in this study to discriminate soybean plantations from other representative land use and land cover (LULC) classes present in the study area.

## 2.3. Data Sets

In this study, we selected a set of seven Landsat 8 OLI images converted into surface reflectance (path/row = 220/68) from 1 October 2021 to 31 March 2022, which corresponds to the crop growing season in the study area. The images were obtained from the United States Geological Survey (USGS) Earth Explorer platform (https://earthexplorer.usgs.gov/ accessed on 28 April 2022). We preprocessed these images by applying a multiplicative scale factor (0.0000275) and an additive offset (−0.2) [39]. The quality assurance (QA) band was used to mask cloud, cloud shadows, water, and cirrus by using 21,824 as the integer value.

We also selected the Sentinel-2 MSI harmonized surface reflectance data resampled to 30 m over the Sentinel-2 tilling system (HLSS30; 28 images) as well as the Landsat 8 OLI harmonized surface reflectance data resampled to 30 over the Sentinel-2 tiling system (HLSL30; 6 images), version 2.0 from the same period (Figure 3). The HLSS30 and HLSL30 images are available with geometric, radiometric, and atmospheric corrections [22]. To cover the entire study area, we selected the images from four tiles: 23LLG, 23LLH, 23LMG, and 23LMH. Whenever the Landsat 8 and Sentinel-2 overpasses were coincident, we opted to select the HLSS30 data. All available images were downloaded from the EarthData website (https://search.earthdata.nasa.gov/ accessed on 1 April 2022).
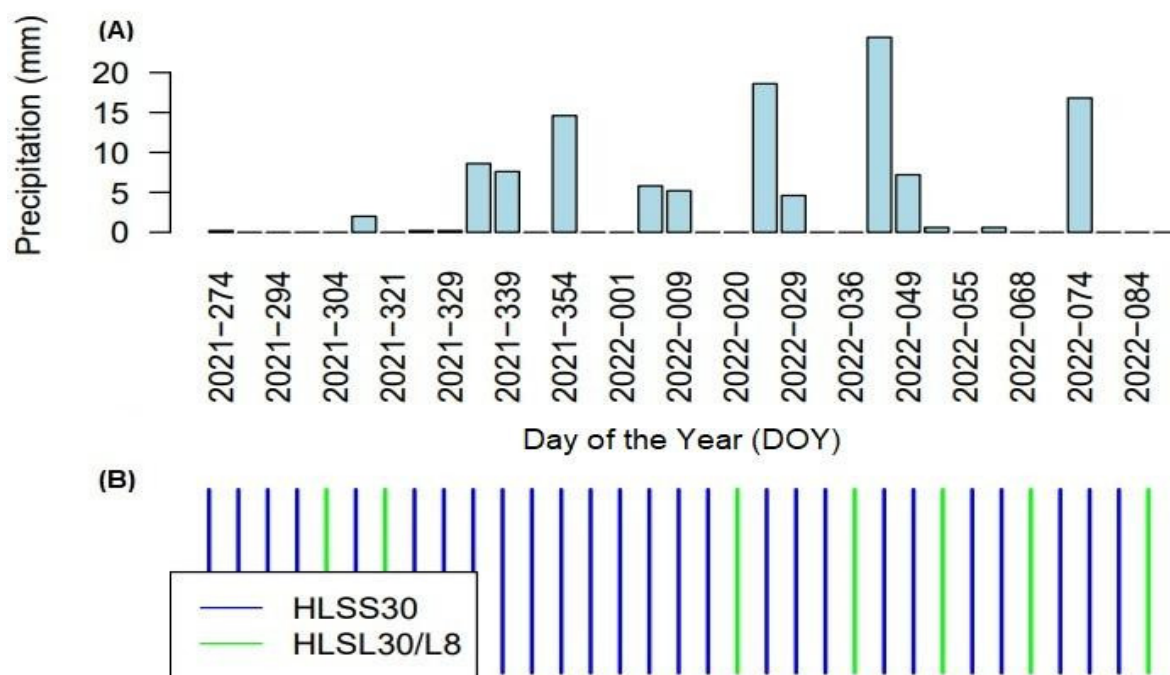
**Figure 3.** (**A**) Daily rainfall data measured by the automatic meteorological station located in the municipality of Luís Eduardo Magalhães during the 2021–2022 crop growing season over the study area. (**B**) Landsat 8 OLI, HLSL30, and HLSS30 images selected during the 2021–2022 crop growing season.

Both Landsat 8 OLI and HLS (HLSS30 and HLSL30) images were preprocessed by the following steps: (1) mosaicking the images per band and per date; (2) reprojection of the mosaics for the Southern Hemisphere (Universal Transverse of Mercator—UTM projection system and WGS84 datum); and (3) cropping the mosaics to the boundary of the study area. We also used the HLS F-mask to remove pixels contaminated by clouds, cloud shadows, water, and high aerosol level, using the digital numbers 64 and 128 (Table 2), following the indications of the HLS Users' Guide v2.0 [40]. For further image processing, as detailed in the next subsection, we selected the following spectral bands common to both Landsat 8 OLI and Sentinel-2 MSI satellites: blue; green; red; near-infrared (NIR); shortwave infrared 1 (SWIR 1); and shortwave infrared 2 (SWIR 2).

**Table 2.** Integer values selected from the F-mask, representing pixels without clouds, cloud shadows, water, and high aerosol level. The value 0 at the indicated bit means absence, while 1 indicates presence. Bit 7 and Bit 6 are the aerosol level (01 represents low level and 10 represents medium level); Bit 5 is water; Bit 4 is snow or ice; Bit 3 is cloud shadow; Bit 2 is adjacent to cloud shadow; Bit 1 is cloud; and Bit 0 is cirrus.

| Integer Value | Bit 7 | Bit 6 | Bit 5 | Bit 4 | Bit 3 | Bit 2 | Bit 1 | Bit 0 |
|---|---|---|---|---|---|---|---|---|
| 64 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Spectral Vegetation Indices (VIs)

The Landsat 8 OLI, HLSS30, and HLSL30 images were converted into the following spectral vegetation indices (VIs): Normalized Difference Vegetation Index (NDVI) [41]; Green Normalized Difference Vegetation Index (GNDVI) [42]; Normalized Water Difference Index (NDWI) [43]; and Soil-Adjusted Vegetation Index (SAVI) [44] (Table 3). The NDVI is the most traditional vegetation index and presents a high correlation with photosynthetic-activity-related parameters such as the leaf area index and leaf chlorophyll [41]. However, some plant phenological phases related to changes in leaf pigments, water content, and crop residues may not be entirely analyzed using only NDVI, especially in regions with

complex crop cultivation patterns [45]. GNDVI provides more sensitivity to chlorophyll-a content than NDVI by replacing the red band with the green band, which is more sensitive to chlorophyll [42], improving the potential to detect stressed and senescent vegetation and estimate green crops [42]. NDWI accentuates differences in the leaf water contents of several vegetation types [43]. SAVI minimizes soil background effects in the NDVI by using a correction factor in the NDVI formula [44]. The combined use of these VIs has been valuable for crop mapping purposes [8].

**Table 3.** Spectral vegetation indices (VIs) used in the study.

| VIs | Name | Equation | Reference |
|------|------|----------|-----------|
| NDVI | Normalized Difference Vegetation Index | $\frac{R\lambda nir - R\lambda red}{R\lambda nir + R\lambda red}$ | [41] |
| GNDVI | Normalized Difference NIR/Green NDVI | $\frac{R\lambda nir - R\lambda green}{R\lambda nir + R\lambda green}$ | [42] |
| NDWI | Normalized Difference Water Index | $\frac{R\lambda nir - R\lambda swir1}{R\lambda nir + R\lambda swir1}$ | [43] |
| SAVI | Soil-Adjusted Vegetation Index * | $\frac{(R\lambda nir - R\lambda red)}{R\lambda nir + R\lambda red + L} * (1 + L)$ | [44] |

* Factor for soil brightness correction ($L$) = 0.5.

### 2.4. Hierarchical Classification

The image classification was performed considering three hierarchical levels of LULC classes present in this study. The first level is composed of two groups of LULC classes: the agricultural areas and the natural vegetation. The water bodies and urban areas found in the study area were masked. The water bodies were extracted from the LULC map produced by the MapBiomas project, a multi-institutional initiative for reconstructing historical LULC maps from the Brazilian biomes since 1987, based on the cloud computing in the GEE platform [46]. The urban areas were obtained from IBGE [47]. Since the urban areas corresponded to the year 2015, we manually complemented the expanded areas larger than 1 ha. In the second level, the agricultural areas were divided into annual crops, perennial crops, and cultivated pasturelands. At this level, due to the low occurrence of coffee plantations and silviculture, both were combined into the class of perennial crops because of their similar spectral response. Finally, in the third level, the annual crops were split into soybean and other annual crops. The image classification was also generated considering other datasets: only multispectral bands (MS); only spectral vegetation indices (VI); and a combination of multispectral and VIs subsets (MS + VIs).

To perform the image classification, a set of 192 sampling points was surveyed during a field campaign conducted on 29–30 November 2021, with the support of the AgroTag software application [48] (Figure 4). More specifically, we surveyed 86 points of harvested crops in 2021–2022, 48 of soybean, 38 of other annual crops (maize, cotton, and bean), and 20 of pasturelands and Cerrados' natural vegetation. AgroTag is a software developed by the Embrapa Environment located in Campinas, São Paulo State to gather and share field information. The software enables the acquisition and storage of Global Positioning System (GPS) coordinates, field photos, and metadata of different sampling points. Seventy percent of ground truth data were used for training the classifier while 30% were used for validation.

The region is characterized by the tropical continental climate, i.e., Aw in the Köppen climate classification system [36], with strong climate seasonality: six months of dry season (April to September) and six months of wet season (October to March). According to the rainfall data provided by an automatic station located in the municipality of Luís Eduardo Magalhães and recorded by the National Institute of Meteorology [37] (station code: A404; latitude = −12.15°; longitude = −45.83°; elevation = 760 m; operation date = 17 April 2002), the average annual rainfall is 885 mm. This rainfall gradient increases from east to west. According to the data gathered by the Shuttle Radar Topography Mission (SRTM), the topography of the study area is flat, with an average elevation of 788 m, while highly weathered Oxisols, with low natural fertility and high Al toxicity, are the dominant soil type [38].
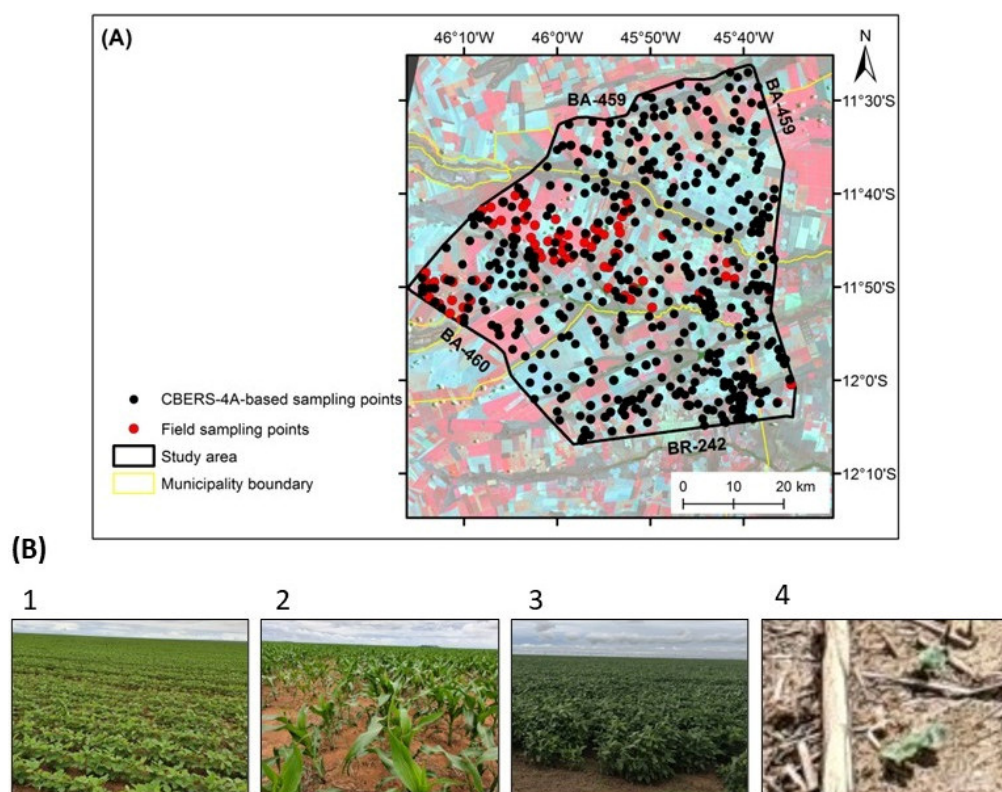
**Figure 4.** (**A**) Location of the field sampling points and remote sensing-based reference points. The field campaign was conducted on 29−30 November 2021. The image used for remote sensing-based points was obtained by the CBERS-4A satellite with the overpass on 27 October 2021. (**B**) Field photos of agricultural fields illustrating the studied crops: (1) soybean; (2) maize; (3) bean; and (4) recently planted cotton (presence of maize straw).

The ground truth data were complemented by samples extracted from a high-resolution RGB color composite of 8-m spatial resolution red, green, and blue bands of the China–Brazil Earth Resources Satellite (CBERS-4A) Wide Scan Multispectral and Panchromatic Camera (WPM) fused with the 2–meter spatial resolution panchromatic band acquired on 27 October 2021. These data were used only for the Levels 1 and 2 classification procedures since the remote sensing-based reference data are not quite feasible for complex targets [49]. At Level 1, we collected 167 samples for each class, while at Level 2, we gathered 30 samples for pasturelands, 30 samples for perennial crops, and 50 samples for annual crops.

Image Classification and Parameterization

In this study, we used the Random Forest (RF) classifier [50], which is one of the most used and successful non-parametric algorithms for image classification. It operates by combining multiple random decision trees that group decisions by average, where the most voted class among all the trees in the forest is considered the final response. It can deal with large volumes of data and imbalanced classes, supplying a ranking of the importance of variables for constructed models [51,52]. We have considered other machine learning methods at the initial stage of the study. However, some tests have shown no significant difference. Moreover, RF is a well-established method known to handle data with high dimensionality, robustness against overfitting, and high predictive capacity [53], characteristics that can be considered relevant for our purpose in this study. This situation endorsed the use of RF.

To improve the performance of the RF classification, we employed 10-fold cross-validation to adjust a set of parameters, the number of variables randomly sampled as candidates at each split (mTry), the maximum number of terminal nodes trees (maxnode), and the number of trees to be grown (nTree) for each model, using the *caret* and *randomForest*

packages available in the R environment (Table 4). The final maps were generated from the best dataset at each level and filtered with the modal function in $3 \times 3$ and $5 \times 5$ moving windows, depending on the level, using tools of the *raster* package.

**Table 4.** Results of the parametrization considering the number of variables at each split (mTry), the number of trees in the forest (nTree), and the maximum number of terminal nodes (maxnode) for each model, after 10-fold cross-validation.

| Level | Parameter | HLS | | | L8 | | | HLSS30 | | | HLSL30 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MS | VIs | MS + VIs | MS | VIs | MS + VIs | MS | VIs | MS + VIs | MS | VIs | MS + VIs |
| 1 | mTry | 10 | 6 | 14 | 18 | 16 | 19 | 13 | 11 | 12 | 6 | 5 | 6 |
| | maxnode | 9 | 14 | 8 | 11 | 9 | 7 | 12 | 9 | 9 | 14 | 8 | 6 |
| | nTree | 50 | 250 | 50 | 150 | 50 | 500 | 50 | 100 | 300 | 50 | 150 | 300 |
| 2 | mTry | 13 | 6 | 14 | 7 | 4 | 13 | 15 | 15 | 11 | 7 | 14 | 5 |
| | maxnode | 10 | 12 | 6 | 14 | 12 | 13 | 12 | 9 | 9 | 12 | 8 | 12 |
| | nTree | 50 | 200 | 500 | 250 | 350 | 50 | 100 | 100 | 150 | 450 | 250 | 300 |
| 3 | mTry | 9 | 6 | 18 | 9 | 5 | 10 | 14 | 11 | 17 | 8 | 5 | 5 |
| | maxnode | 6 | 10 | 8 | 5 | 5 | 7 | 6 | 9 | 10 | 13 | 10 | 13 |
| | nTree | 100 | 50 | 400 | 100 | 50 | 50 | 150 | 500 | 200 | 50 | 100 | 200 |

*2.5. Accuracy Assessment and Statistical Analysis*

We assessed the accuracies of the RF classifications through the overall accuracy (OA), Kappa index, producers' accuracy (PA), and users' accuracy (UA). The performances of each dataset (HLS, HLSS30, HLSL30, and Landsat 8 OLI) and each subset (MS, VIs, and MS + VIs) were evaluated to determine the best model. Therefore, along with the exploratory statistical analysis of our results, we also employed the two-sample Student's *t*-test, using the Basic Statistics and Data Analysis (BDSA) package in R [54], to assess if the HLS outcomes (OA and Kappa) were statistically different from the HLSS30, HLSL30, and L8 in a confidence interval of 0.95. This was done after we confirmed the normal distribution of the models' results for each dataset at each level using the Shapiro-Wilk test.

One default output from RF models is the variable importance ranking, where the predictive power of each variable is measured throughout a series of models' reruns with selected predictors, being a vital tool for feature selection, dimensionality reduction, and understanding variable interaction [50]. Therefore, we considered the metric of Mean Decrease in Accuracy (MDA), which measures the variable importance using *out-of-bag* samples and recording the variation in prediction error at each variable permutation, to build ranks with the top 10 most important variables for each dataset (HLS, HLSS30, HLSL30, and L8) at each level.

**3. Results**

*3.1. Influence of Cloud Cover on Satellite Data Availability*

Table 5 shows the influence of cloud cover on L8 data availability over the study area during the crop growing season. Five L8 images from a total of 11 overpasses presented 100% of cloud coverage, while another seven images presented cloud coverage ranging from 4% to 70%. Considering only the area occupied by the *Soybean Ring*, the percent of data loss due to the cloud cover interference varied from 12% to 97%, with an average of 47%.

Table 6 shows the influence of cloud cover on HLSS30 and HLSL30 data availability over the study area. The number of available scenes increased from 7 (L8) to 34 (28 HLSS30 overpasses and 6 HLSL30 overpasses). The data loss due to bad-quality pixels varied from 61% in November to 87% in December and January, with an average value of 74% during the crop growing season. Only 12 scenes presented less than 60% of data loss, while 17 scenes presented cloud coverage and cloud shadow higher than 90%. Some expected images were unavailable for download, possibly because of the high percentage of cloud coverage, which can affect the pre-processing steps. We noted that some of these missing

images in the EarthData platform were available in the Landsat 8 L1GS database in the Earth Explorer platform, indicating that these images presented insufficient locational information for terrain correction.

**Table 5.** Cloud cover and Landsat 8 OLI data loss caused by using the Quality Assessment (QA) band in the study area (Path/Row: 220/68).

| Overpass | % Cloud Cover over the Entire Tile | % Data Loss over the Study Area |
|---|---|---|
| 1 November 2021 | 69 | 97 |
| 17 November 2021 | 29 | 29 |
| 20 January 2022 | 6 | 35 |
| 5 February 2022 | 15 | 17 |
| 21 February 2022 | 51 | 93 |
| 9 March 2022 | 4 | 12 |
| 25 March 2022 | 28 | 45 |

**Table 6.** Total of overpasses, average cloud cover, and percent data loss in the Harmonized Landsat Sentinel-2 images caused using the F-mask in the study area.

| Month | Total of Overpasses | % Cloud Cover * | Data Loss (%) |
|---|---|---|---|
| October | 4 | 36 | 69 |
| November | 6 | 68 | 61 |
| December | 4 | 77 | 87 |
| January | 7 | 67 | 87 |
| February | 7 | 57 | 71 |
| March | 6 | 44 | 64 |

* Average cloud cover among the four tiles (T23LLG, T23LLH, T23LMG, and T23LMH) necessary to cover the study area.

### 3.2. Classification Results

Figure 5 depicts the classification results obtained via RF classifier based on both spectral bands and VIs of HLS, HLSS30, HLSL30, and L8 images for three hierarchical levels. The HLS Level 1 map shows that most of the study area is occupied by agricultural areas, while the remaining native vegetation mostly corresponds to riparian forests. The proportion of agricultural areas identified by our best models were 73% (HLS and HLSS30), 74% (HLSL30), and 75% (L8) of the *Soybean Ring*, respectively. The annual crops were also similarly assessed as being 86% (HLS), 83% (HLSS30), 84% (HLSL30), and 88% (L8) of the agricultural areas (approximately 60–65% of the entire study area). The L8 and HLSL30 final maps of Level 2 estimated 200% and 138%, respectively, more perennial crops (coffee under irrigation system and silviculture) than the HLS dataset, which means a difference of 80 and 55 km$^2$. Indeed, some of the coffee areas were missed by HLS at Level 1, classified as natural vegetation. HLS found 102 and 75 km$^2$ more pastures than the HLSL30 and L8 maps, respectively. The datasets differ in Level 3 classifications, as it will be approached in the next session, reflected in the soybean estimations. While the L8 and HLSL30 best models calculated that 68% and 65% of the annual crops at the first growing season were soybean, respectively, HLS estimated a rate of 77% (45% and 49% of the total area), with a difference up to 145 km$^2$ (Figure 6).
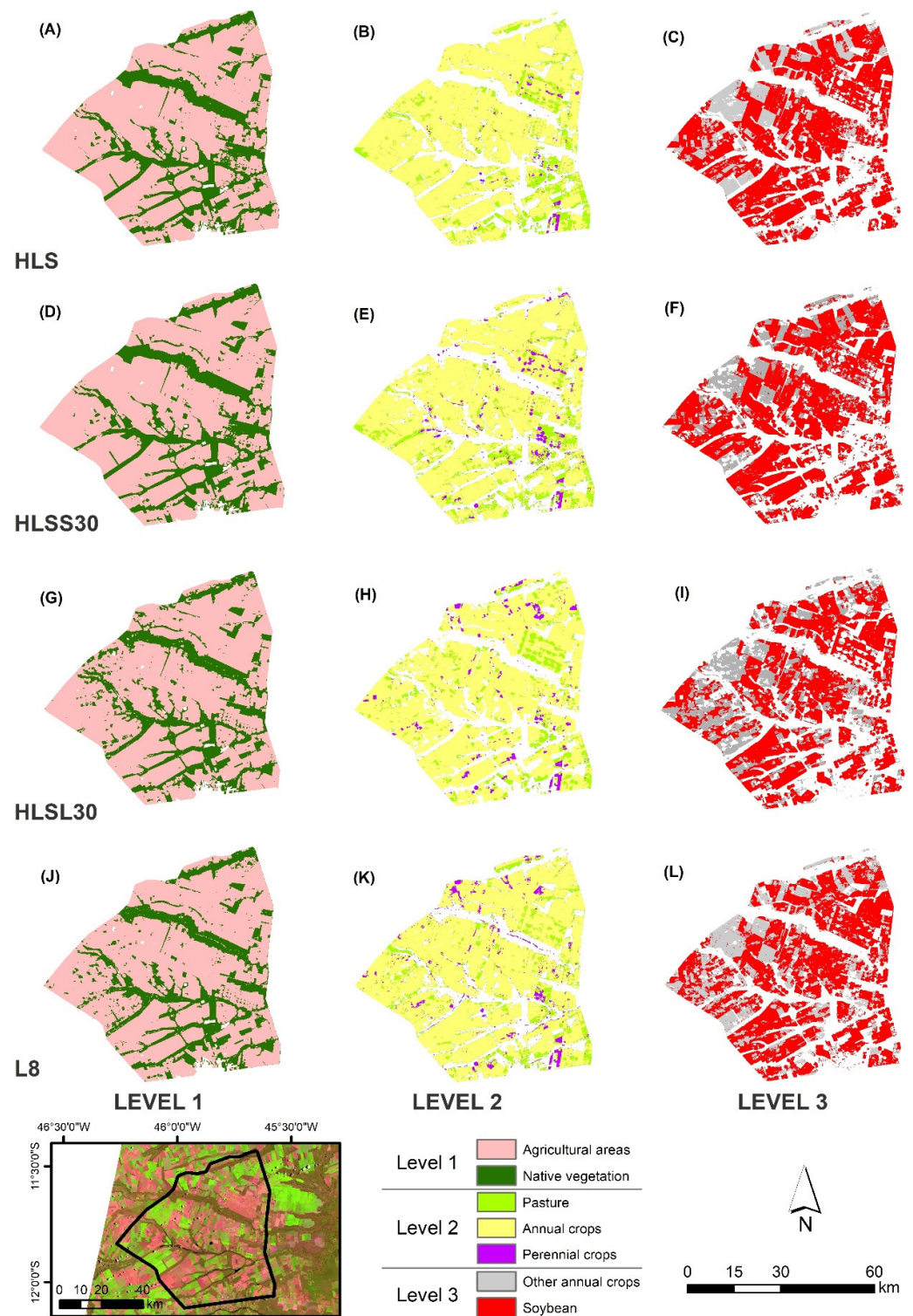
**Figure 5.** Random Forest classification results, based on Landsat 8 OLI and Harmonized Landsat Sentinel-2 (HLS) scenes, combining multispectral bands and spectral vegetation indices. (**A**–**C**) correspond to the classifications results involving the Harmonized Landsat Sentinel-2 data set; (**D**–**F**) correspond to the Sentinel-2 Multi-spectral Instrument Surface Reflectance (HLSS30) datasets; (**G**–**I**) to the three Landsat Operational Land Imager Surface Reflectance and TOA Brightness (HLSL30) datasets; and (**J**–**L**) correspond to the Landsat 8 Operational Land Imager (L8) datasets. The color-composition is a Landsat 9 OLI RGB456 acquired on 1 March 2022.
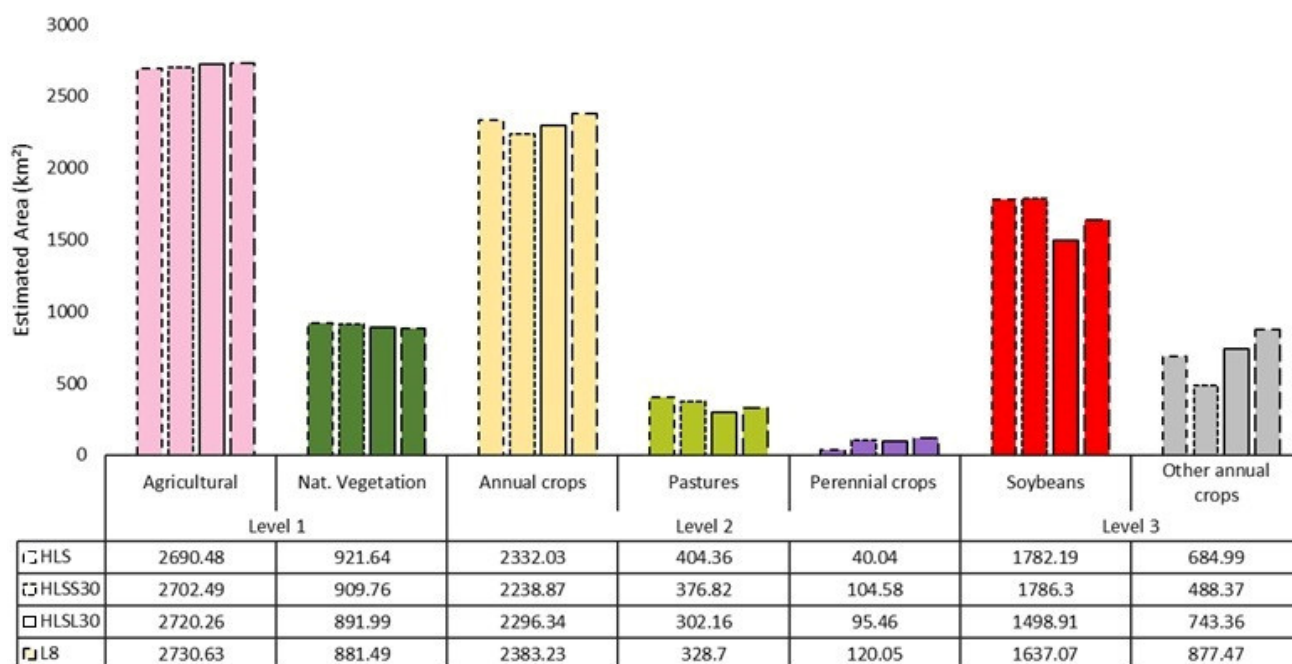
**Figure 6.** Estimated areas (km$^2$) for each class at each level, from the HLS, HLSS30, HLSL30, and L8 datasets.

| | Agricultural | Nat. Vegetation | Annual crops | Pastures | Perennial crops | Soybeans | Other annual crops |
|---|---|---|---|---|---|---|---|
| HLS | 2690.48 | 921.64 | 2332.03 | 404.36 | 40.04 | 1782.19 | 684.99 |
| HLSS30 | 2702.49 | 909.76 | 2238.87 | 376.82 | 104.58 | 1786.3 | 488.37 |
| HLSL30 | 2720.26 | 891.99 | 2296.34 | 302.16 | 95.46 | 1498.91 | 743.36 |
| L8 | 2730.63 | 881.49 | 2383.23 | 328.7 | 120.05 | 1637.07 | 877.47 |

### 3.3. Accuracy Assessment and Statistical Analysis

In general, the models' performances decrease as the hierarchical levels increase. At Level 1, except for the HLSL30, all models were able to separate agricultural areas from natural vegetation with accuracies over 0.90. The average OA/Kappa obtained by each dataset was 0.938/0.876 (HLS), 0.945/0.889 (HLSS30), 0.866/0.730 (HLSL30), and 0.948/0.897 (L8). The Student´s *t*-test showed that HLS is a better dataset than HLSL30 with statistical significance. The Cerrado is marked mainly by its heterogeneity, where different environmental conditions build a complex landscape, with several phytophysiognomies that, at the same time, reveal the high biodiversity of this biome and make mapping initiatives more challenging [6]. However, within the *Soybean Ring*, most of the remaining native vegetation is composed of gallery forests, which have a constant response over time, given the natural composition of their closed tree canopies and the absence of harvests or clear cuts during the harvest period [8].

The performances were also similar at Level 2, with average OA/Kappa of 0.903/0.837 (HLS), 0.860/0.778 (HLSS30), 0.763/0.604 (HLSL30), and 0.892/0.823 (L8). The *t*-test pointed out that the HLS models were only statistically superior from the HLSL30. At level 3, HLS had a mean OA/Kappa of 0.882/0.744, in opposition to a mean OA/Kappa of 0.840/0.633 (HLSS30), 0.725/0.441 (HLSL30), and 0.754/0.489 (L8). According to the Student´s *t*-test, the HLS has a statistically significant superiority of OA and Kappa over all datasets. The table containing the results from the hypothesis test can be seen in the Supplementary Material. Neither the HLSL30 nor the L8 data produced models with statistical significance in separating soybean from other crops at this level.

Regarding the performances of the multispectral bands (MS) and spectral vegetation indices (VIs), individually or combined, at Level 1, the combination (MS + VIs) produced higher performances in comparison with the MS and VIs subsets separately, except for the HLSL30 dataset. The combination of bands and indices improved the OA and Kappa of HLS up to 10% and 9.8%, respectively. At Level 2, the L8 VIs subset over-performed the others of this sensor, while the combined subset also generated the best HLS model. The MS subsets of L8 and HLS presented the lower scores, while the VIs and MS + VIs improved the results for these datasets. For the HLSS30 dataset, bands alone produced a model nearly identical to MS + VIs, but only with 3% higher Kappa. At Level 3, the VIs models had the

weakest performances for HLS, HLSS30, and L8, while the best overall model was obtained with the combination of bands and indices of HLS data. The overall accuracies and Kappa indices of all models are presented in Table 7, and the confusion matrices are shown in Supplementary Material along with the error matrices and classes accuracies.

**Table 7.** Overall Accuracy (OA) and Kappa Index for Level 1, 2, and 3 classifications (L8, HLS, HLSS30, and HLSL30) in datasets: multispectral bands (MS), spectral vegetation indices (VIs), and combination of MS and VI (MS + VIs) bands.

| Sensor/Data | Classifications | Datasets | OA | Kappa |
|---|---|---|---|---|
| Landsat-8 Operational Land Imager OLI (L8) | Level 1 | L8 MS | 0.938 *** | 0.877 |
| | | L8 VIs | 0.948 *** | 0.896 |
| | | L8 MS + VIs | 0.959 *** | 0.918 |
| | Level 2 | L8 MS | 0.839 *** | 0.734 |
| | | L8 VIs | 0.935 *** | 0.895 |
| | | L8 MS + VIs | 0.903 *** | 0.840 |
| | Level 3 | L8 MS | 0.782 $^{ns}$ | 0.559 |
| | | L8 VIs | 0.696 $^{ns}$ | 0.349 |
| | | L8 MS + VIs | 0.783 $^{ns}$ | 0.559 |
| Harmonized Landsat Sentinel-2 (HLS) | Level 1 | HLS MS | 0.917 *** | 0.835 |
| | | HLS VIs | 0.938 *** | 0.876 |
| | | HLS MS + VIs | 0.959 *** | 0.917 |
| | Level 2 | HLS MS | 0.839 *** | 0.726 |
| | | HLS VIs | 0.935 *** | 0.892 |
| | | HLS MS + VIs | 0.935 *** | 0.892 |
| | Level 3 | HLS MS | 0.867 ** | 0.721 |
| | | HLS VIs | 0.867 ** | 0.704 |
| | | HLS MS + VIs | 0.913 ** | 0.808 |
| Sentinel-2 Multi-spectral Instrument Surface Reflectance (HLSS30) | Level 1 | HLSS30 MS | 0.928 *** | 0.855 |
| | | HLSS30 VIs | 0.948 *** | 0.897 |
| | | HLSS30 MS + VIs | 0.959 *** | 0.917 |
| | Level 2 | HLSS30 MS | 0.871 *** | 0.788 |
| | | HLSS30 VIs | 0.839 *** | 0.746 |
| | | S30 MS + VIs | 0.871 *** | 0.800 |
| | Level 3 | HLSS30 MS | 0.869 ** | 0.721 |
| | | S30 VIs | 0.783 $^{ns}$ | 0.475 |
| | | S30 MS + VIs | 0.869 *** | 0.704 |
| Landsat-8 Land Imager Surface Reflectance and TOA Brightness (HLSL30) | Level 1 | HLSL30 MS | 0.897 *** | 0.794 |
| | | HLSL30 VIs | 0.845 *** | 0.689 |
| | | HLSL30 MS + VIs | 0.856 *** | 0.708 |
| | Level 2 | HLSL30 MS | 0.742 * | 0.568 |
| | | HLSL30 VIs | 0.774 ** | 0.622 |
| | | HLSL30 MS + VIs | 0.774 ** | 0.622 |
| | Level 3 | HLSL30 MS | 0.696 $^{ns}$ | 0.414 |
| | | HLSL30 VIs | 0.739 $^{ns}$ | 0.425 |
| | | HLSL30 MS + VIs | 0.739 $^{ns}$ | 0.485 |

Level of significance: $p > 0.05$ *; $>0.005$ **; $>0.0005$ ***; non-significant $^{ns}$.

### 3.4. Variable Importance

Based on the Mean Decrease Accuracy (MDA) scores for the HLS datasets considered in the RF classification, VIs showed to be more important than spectral bands for Levels 1 and 2 of the hierarchical classifications in all datasets. However, for the Level 3 classification, spectral bands overperformed VIs. Regarding the time of the crop growing season, the beginning of the season showed to be more relevant for the Levels 1 and 2 classifying procedure, while, to discriminate soybeans from other annual crops, the end of the growing

season is more prominent. For the L8 and HLSL30 models, there was an evident relevance of the clearer overpass, i.e., 9 March 2022, when cloud cover was lower than 5% and only 12% of the data were lost. Figure 7 displays the top 10 more important variables for the HLS datasets in the three levels of classification, while those from HLSS30, HLSL30, and L8 are shown in the Supplementary Material.
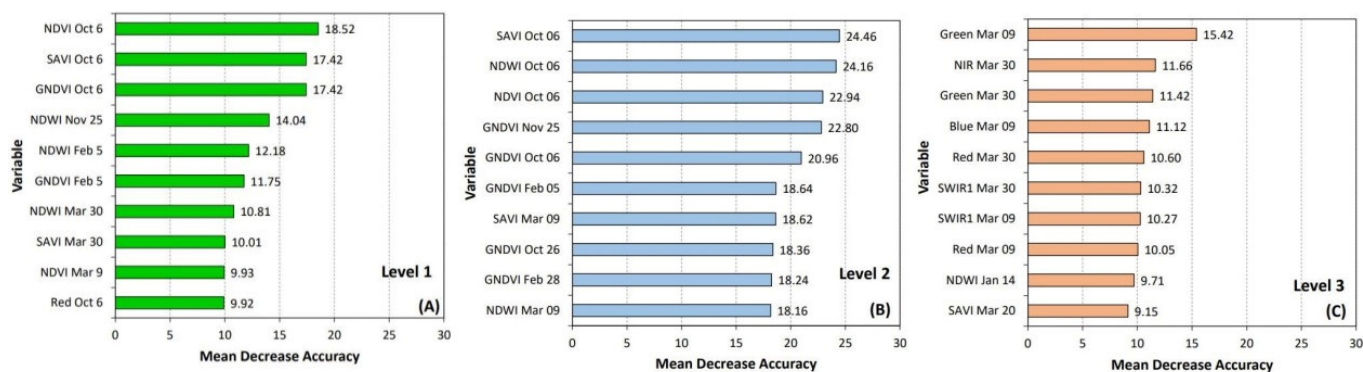


**Figure 7.** Measures of the importance of the Random Forest top 10 predictor variables for Level 1 (**A**), Level 2 (**B**), and Level 3 (**C**) hierarchical classifications based on the Harmonized Landsat Sentinel-2 (HLS) datasets. The months' names are abbreviated as Oct (October), Nov (November), Feb (February), Mar (March), and Jan (January).

Considering which HLS variables were poor predictors, we observed that, in Level 1, SWIR 1(from 15 December 2021, 5 February and 20 March 2022) and SWIR 2 (from 20 November 2022, 20 and 25 March 2022) were the poorest spectral bands, and NDVI from 10 March was the poorest VI. In Level 2, SWIR 1 from 18 February, Blue from 17 November 2021, and Red from 30 November 2021 were the poorest predictors among the spectral bands. Among VIs, GNDVI from 10 Mach 2022, NDWI from 9 January 2022, and NDVI from 18 February 2022 were the poorest. In Level 3, Red, SWIR 1, and NIR from 26 October 2021 and SWIR1 from 15 March 2022 were the poorest spectral band predictors. Among VIs, GNDVI (from 15 March and 14 January 2022) and NDVI (from 15 March and 3 February 2022) were the poorest. Specifically, at the end of the rainy season, we observed the prevalence of the observations from 10 and 15 March 2022 on those affecting the capability of different VIs and spectral bands. The average data loss among these dates was 73%, with a rate of flagged pixels between 31% and 99%, which reaffirms that having clearer observations along the crop season was determinant to agricultural and soybean mapping.

## 4. Discussion

### 4.1. Cloud Cover Interference on Satellite Image Acquisition

The rainfed crop plantation in Western Bahia includes soybean, maize, cotton, and bean. It starts in early October until late December, depending on crop type and crop variety [55]. The harvesting time usually starts from February until May. Except for the peak of the dry season (July to August), the Western Bahia is a region severely affected by persistent cloud coverage, particularly between December and February, when it usually surpasses 60–70% [13]. This study showed that we have seven L8 images with less than 70% of cloud coverage (only two with less than 10%) for the 2021–2022 crop season monitoring in the study area. This number increases to 21 when HLS images are considered (less than 70%). Therefore, regardless of the increased potential of crop production monitoring by HLS data sets related to their better spatial and spectral resolutions in relation to the L8 data sets, increasing the number of available images throughout the crop growing season is of great advantage.

As previously mentioned, we used the F-mask to remove poor-quality pixels, and we observed an average data loss of 74%. Such loss was, on average, 15% higher than the cloud cover at the acquisition moment. Such issue was also observed in South Dakota, U.S., where

38 HLS overpasses presented more than 90% of flagged pixels in less than one year [26]. In previous versions, HLS quality bands had known issues regarding bright targets [32] as well as commission and omission errors [56]. Some alternative approaches have been evaluated, such as the application of RF or decision tree classifiers over the QA bands and spectral thresholds using the NDWI [33]. Taking regional tuning and mask refinement into consideration can be a strategy to increase the number of valid HLS observations over cloud-persistent regions and therefore improve mapping accuracies; however, cloud masking still is an open matter to be continually addressed in future studies with HLS data [26].

This high cloud cover interference is a trade-off that has been part of the context of rainfed crop type identification and monitoring over the Brazilian agricultural frontiers using optical remote sensing data, especially in the Cerrado biome [57]. The development of initiatives for expanding satellite revisit frequency such as the HLS has the potential to overcome the limitation of cloud-free satellite images [22]. In our study, HLS higher data frequency proved to be fundamental for the detection of soybean production compared to HLSS30, HLSL30, and L8 datasets, providing a more detailed and precise mapping of Brazil's most important crop. They were also efficient for mapping LULC over the rainy period, even with a high rate of data loss.

Therefore, big LULC mapping projects in Brazil such as MapBiomas and TerraClass based on Landsat 30 m and MODIS 250 m data can benefit from HLS, since the temporal resolution is still a limitation, even outside of rainy periods. However, part of the trade-off is the increased amount of data; therefore, NASA HLS products should become available through cloud processing platforms such as the GEE so that regional, national, and global mapping initiatives worldwide can rely on these virtual constellations. Some authors have been developing routines to harmonize Landsat and Sentinel data using resources of the GEE [58]. However, although such processes can improve harmonization coefficients at the regional level, the amount of work and expertise in programming, along with computational limitations, mainly in developing countries, can be limiting factors for many users.

### 4.2. Impact of Parametrization on the RF Classification Performance

Parameter setting on classification with machine learning algorithms has a significant impact on models' performances [51], and though it is widely used for agricultural monitoring and mapping, the parameterization usually is less explored [52]. In this study, after a series of tests, we chose to tune three RF parameters: the number of variables at each split (mTry) from 5 to 20, the number of trees in the forest (nTree) from 50 to 500, and the maximum number of terminal nodes (maxnode) from 5 to 15. To assess the impact of parametrization on models' performances, we used the same seeds to rebuild the best models using the worst values for the three parameters tuned. At Level 1, parameterization increased HLS MS + VIs OA by 2.2% and Kappa by 4.6%, HLSS30 MS + VIs OA by 4.5% and Kappa by 9.7%, but displayed no difference for the L8 MS + VIs model (+0.02 and 0.03% of OA and Kappa). At Level 2, the combination of the worst parameters did not impact HLS and HLSS30 MS + VIs models (<1% variation) but led to an increase of 10.3% of OA and 18% of Kappa in the L8 VIs model. Level 3 was impacted by parameter setting, presenting an increase of 4.8% of OA and 12.9% of Kappa in the HLS MS + VIs model and +5.6% of OA and +13.2% of Kappa in the L8 MS + VIs model. Therefore, although the impact of parameterization in our study was discrete, it became more significant as the level increased, and led to important improvements regarding mainly Kappa coefficients.

### 4.3. LULC Mapping Challenges and Variables Importance

In general, one of the main challenges to accurately mapping the main crops on regional and national scales in Brazil is the lack of ground truth samples, as was the case in this study. Even in a landscape mainly dominated by soybean in the first growing season (possibly followed by maize and other annual crops), the number of ground samples was

smaller than intended due to the bad weather conditions, leading to poor road conditions, along with some inaccessible farms. To overcome these limitations, we adopted a hybrid sampling strategy, using high-resolution remote sensing data for the less complex levels 1 and 2, which is acceptable [49].

However, Brazil is a continental country where highly dynamic agriculture is the main economic activity. Therefore, building a representative database is not always possible, and the development of new strategies to monitor land use and main crop plantation to minimize the above-mentioned limitations is fundamental. In this context, remote sensing tools such high-resolution images and spectral indices temporal signatures are complementary and worked well in our study. The feasibility of spatial and time transference of crop samples from the U.S. Cropland Data Layer (CDL) to Brazil was evaluated by Ajadi et al. [59], using MODIS and Synthetic Aperture Radar (SAR) data. These authors produced a large-scale mapping of soybean production in Brazil in the first growing season with 88% accuracy. Unsupervised techniques can lead to up to 80% accuracy in distinguishing between soybean and maize when ground data is insufficient [60]. However, sharing and integrating databases based on the Findable, Accessible, Interoperable, and Reusable (FAIR) principles, generated by research groups, can also be an interesting way to improve agricultural monitoring where ground data acquisition is challenging [61,62].

The continuous an integrated development of Sentinel-2 and Landsat 8/9 scientific activities that enable the sensors' interoperability and the virtual constellation of HLS [63] is a successful model that could be trespassed to initiatives of collection, storage, and sharing ground data from different users, especially in developing countries, where there is a lack of cropland data and the monitoring of food production is a sensitive issue to decision-makers. The sharing of methods, data, and findings is essential in the current trend of big data on remote sensing [64]. Although acquired two years ago, the database built and shared by Oldoni et al. [65] with some coincident areas was useful for better understanding of our study area, since it was not possible to go through it entirely.

The AgroTag application presents a high potential to overcome the limitations of ground sample acquisition in Brazil if data from different users could be accessed. Finally, it is also valid to reassure the relevant role of high-resolution (2-m) pan-sharpened images from the CBERS-4A satellite and of ready-to-use MODIS NDVI temporal patterns made available by Embrapa through the SATVeg [66] in the collection of samples in less complex classes, making the process possible and more reliable.

At Level 2, areas with straw were a source of misclassification and confusion. Two situations happened: harvesting in the previous season followed by fallow until later February or March, and forest plantation areas in the harvesting process during this assessed period. Similar circumstances were found in previous studies on the agricultural activity in the Brazilian savanna [67,68], and are related to its temporal dynamics, which often propitiate regions in transition, which vary in time, space, concept, and categorization. In both situations, when using the HLS data set, these cases were classified as pasturelands, and when using the L8, they were classified as annual crops. Figure 6 illustrates such issues, showing that the classification based on the HLS and HLSS30 overestimated pastures when compared to Landsat datasets (L8 and HLSL30), while those detected up to three times more perennial crops than HLS. Unfortunately, it was not possible to add more field inspections on the *Soybean Ring*, and there are no official reference data to obtain a better understanding of some particularities present in this area. During the field work, it was noted that, apart from the annual crop predominance, there is a complex agricultural dynamic, as in the Cerrado itself. An example is the large variety of crops (soybean, bean, maize, cotton, and coffee or pastures for seeds) and the irrigation pivots occupied with more than one crop, increasing the challenge of mapping and monitoring crop types in this region.

Moreover, we observed several fields in the conversion process from natural vegetation to possible annual cropping or pasture, as well as silviculture areas being converted into pasturelands, or even abandoned after harvesting. Considering the dynamics of LULC

conversion in this modern agricultural frontier, where LULC change decisions are mainly driven by the agribusiness sector [10], this situation is quite common [7].

These challenges are present in the natural and anthropic Cerrado dynamics. Although some strategies that were not used in this study can be further employed with HLS datasets to assess how they can help reduce confusion and produce even better results, such as the estimation of phenometrics and the use of larger time series [69], it is important to state that the 2021–2022 crop season was unusual for the region in terms of precipitation, since the accumulated rainfall between October and March was 1068 mm, while the average for the same period was 790 mm in the last two decades [37]. This caused even more cloud cover, and therefore, harmed the total of useful remote sensing optical observations.

Regarding the spectral data, earlier HLS images (mainly October) and the VIs showed to be more appropriate to discriminate between agricultural lands and natural vegetation (Level 1), as well as to discriminate between annual crops, perennial crops, and pasture-lands (Level 2). This situation was somewhat expected, since at the beginning of the wet season, most of the agricultural lands are dominated by bare soils and straws [60,67,68], highlighting the spectral contrast with the evergreen gallery forests.

The discrimination between soybean and other annual crops (Level 3) in HLS models was more effective at the end of the wet season with multispectral bands alone or combined with vegetation indices (MS + VIs). The HLS, HLSS30, and L8 VIs model also presented the weakest performance among the three subsets at this level. Adding the MS subset increased the results up to 12% in OA and 60% in Kappa, and seven out of the 10 more important variables were from March. Using L8 data to analyze the spectral-temporal response of different types of crops, such as maize, soybean, and sugarcane, Montibeller et al. [70] also found spectral bands more efficient to retrieve such responses than the spectral vegetation indices. These patterns can be explained by the differences in the growing season between soybean and other annual crops, especially cotton, the second most-cultivated crop in this harvest period. In March, most of the soybean is entering harvesting time, while cotton is still in the growing season. This difference in crop calendars is a proxy for separating both crop types, as each spectral profile pattern varies in time and amplitude, as previously reported [71–75].

Overall, MS + VIs subsets outperformed the others, enhancing the importance of VIs for LULC mapping purposes. Capable of expressing differences in plant responses under different soil, weather, environment, and management conditions [76], VIs were essential to detect soybean production throughout our hierarchical classification system. HLS-based results showed their relevance for Levels 1 and 2. In Level 1, NDVI, SAVI, and GNDVI from 6 October 2021 were the three most relevant features to the RF algorithm, while NDWI from 25 November 2021 and 5 February 2022 also were significant. In Level 2, SAVI, NDWI, and NDVI from the same date (6 October 2021) were the three most relevant features, with GNDVI being important on different dates (25 November 2021, 6 October 2021, 5 February 2022, 26 October 2021, and 28 February 2022). L8-based results showed the prevalence of spectral bands for Levels 1 and 3. In both cases, Green, SWIR1, and Red from the observation acquired on 9 March 2022 were the three most relevant features, with Blue, SWIR2, and NIR presenting complementary relevance in different observations of January and March 2022, highlighting the period of maximum vegetative vigor. Level 2 was the only L8-based result where VIs were the most relevant features: GNDVI, NDVI, SAVI, and NDWI from 9 March 2022. As it was mentioned, the top 10 most important variables of the HLSS30, HLSL30, and L30 datasets are show in the Supplementary Material.

*4.4. HLS Applications in Agricultural Monitoring*

The impact of the higher satellite revisit provided by HLS, although it was not evident in all three levels of models, was noticeable in the mapping results (Figure 5). At Level 3, our main subject, the HLS datasets outperformed HLSS30, HLSL30, and L8 with statistical significance. This difference was evident in the final maps, where the HLS MS + VIs model provided a more confident representation of crops' spatial distributions over the study area.

In contrast, for example, the L8 MS + VIs final map was visibly harmed by the lack of data, presenting a higher occurrence of misclassifications and granular effects (Figure 8). These circumstances may have been due to the spectral similarity between soybean, maize, and cotton. Regionally, all are rainfed and dryland crops with relatively close growth times. Confusion involving soybean and maize can derive from the intercropping between them commonly adopted by local farmers [77] and the similarity of crop type phenology and their spectral responses [69].
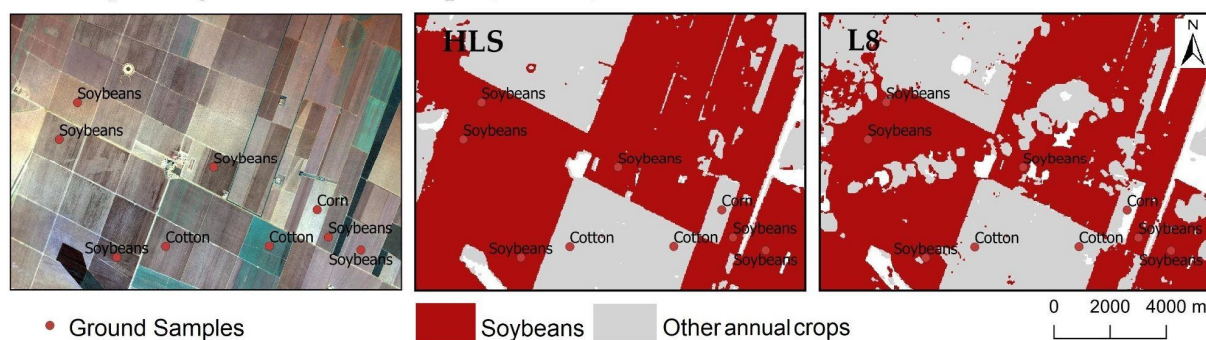
## Example: Soybeans x other crops (Level 3)



**Figure 8.** Level 3 classification example illustrates the potential of the HLS time series, denser than L8, in distinguishing soybean from other annual crops even with a very high occurrence of clouds during the key periods of the season. In Western Bahia, the lack of optical remote sensing data due to cloud cover still is one of the main challenges of crop monitoring and it became evident, especially in the L8 map.

In addition to causing a greater accumulation of clouds, the uncommon rainfall volume in this harvest period may have approximated sowing and phenological calendars for different crops, as reported in previous studies [78,79]. In this scenario, the HLS' higher frequency of valid observations may have improved the detection of relevant dissimilarities between these crops.

The increase in temporal resolution provided by the HLS is a significant differential in the monitoring of surface dynamics in a spatial resolution that is more suitable for management on a regional scale, contributing to the improvement of the characterization of surface phenology under different stimuli [32,34,80,81] and the detection of phenological metrics that are key to agricultural management such as emergence [29,82], which in turn can improve the separability between crop types under different environmental conditions [26,58,83–86], as was the case in this study.

The HLS datasets from NASA or independent harmonization methods were also tested and improved results in many tasks that were related to agricultural monitoring and management such as tracking agricultural extension and intensification [27,87,88], provided denser VIs time series to support yield estimates [23], were included as input data in crop growth models [28], and can be used to assist in managing the use of water resources in agriculture, recovering and modeling evapotranspiration [89] and surface temperature [90], and detecting the distribution of irrigation areas [30].

Despite presenting a high potential for agricultural management and monitoring on different fronts, the balance between temporal and spatial resolution of HLS data also proved to be important in monitoring diverse surface events, such as the detection of pasture cutting [31,91] and the occurrence of fires [92], flooding [93], and disturbances [94,95], among others. However, this is a new approach that remains underexplored, with few studies employing such data. Therefore, we expect growth in its exploration since many issues involving the combined use of Landsat and Sentinel-2 data [96] still need to be addressed to enhance the capacity to obtain dense time series useful for land surface monitoring purposes.

## 5. Conclusions

Our results show that the time series of HLS images acquired over the crop growing season can map rainfed soybean plantations more effectively than HLSS30, HLSL30, and L8. The Landsat images (HLSL30 and L8) were not capable of accurately distinguishing soybean from other annual crops at Level 3. The RF classifier was able to deal with inherent limitations related to the reduced number of cloud-free images over the crop growing season in the region as well as with spectral similarities between some annual crops (e.g., soybean and bean) and varying crop planting dates. This study demonstrated that the RF classifier applied to HLS images can address the challenges of identifying different crop types, which can be useful for public policies of monitoring and forecasting agricultural commodities over tropical regions.

There are still other important issues to be addressed in future work. For example, the use of different classification algorithms and variables such as phenometrics, the inclusion of all 13 spectral bands of the Sentinel-2 images, and the comparison between Sentinel-2 and HLS30 images need to be evaluated. In the Cerrado region, the crop–livestock integration management system has been increasingly adopted by farmers, especially the integration between soybeans and maize, both with different Brachiaria varieties used for pasture. Ground truth involving crop–livestock integration from representative sites, which is the case of Cerrado/Amazon ecotone in the Mato Grosso State, should be also considered.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs14153736/s1, twelve tables containing error matrices, users' and producers' accuracies of models from all datasets and levels. One table containing the results of the hypothesis test (Students' *t*-test) indicating the level of significance of the HLS complete time-series performances (superiority or inferiority) against the other datasets, and three figures providing the top 10 most important variables for the HLSS30, HLSL30 and L8 datasets at each level.

**Author Contributions:** Conceptualization, É.L.B., E.E.S., I.D.S. and T.C.P.; methodology, T.C.P., M.E.D.C., I.D.S., É.L.B and E.E.S.; software, T.C.P. and M.E.D.C.; validation, É.L.B., E.E.S., I.D.S., T.C.P. and G.M.B.; formal analysis, T.C.P., M.E.D.C., É.L.B., E.E.S., I.D.S. and D.d.C.V.; investigation, É.L.B., E.E.S., I.D.S., T.C.P., M.E.D.C. and L.E.V.; resources, É.L.B.; data curation, T.C.P., M.E.D.C. and D.d.C.V.; writing—original draft preparation, É.L.B., T.C.P., M.E.D.C., I.D.S. and E.E.S.; writing—review and editing, É.L.B. and E.E.S.; visualization, T.C.P. and É.L.B.; project administration, É.L.B.; funding acquisition, É.L.B., E.E.S. and I.D.S. All authors have read and agreed to the published version of the manuscript.

**Data Availability Statement:** Publicly available satellite datasets were used in this study. The HLS data were downloaded at no cost via https://search.earthdata.nasa.gov/, accessed on 15 April 2022. The Landsat 8 and 9 data were also downloaded free via https://earthexplorer.usgs.gov/, accessed on 1 May 2022.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Spera, S. Agricultural intensification can preserve the Brazilian Cerrado: Applying lessons from Mato Grosso and Goiás to Brazil's last agricultural frontier. *Trop. Conserv. Sci.* **2017**, *10*, 1–7. [CrossRef]
2. FAO—Food and Agriculture Organization of the United Nations FAOSTAT—Crops and Livestock Products. Available online: https://https://www.fao.org/faostat/en/#data/QCL (accessed on 10 June 2022).
3. Bolfe, É.L.; Jorge, L.A.D.C.; Sanches, I.D.; Luchiari Júnior, A.; da Costa, C.C.; Victoria, D.D.C.; Inamasu, R.Y.; Grego, C.R.; Ferreira, V.R.; Ramirez, A.R. Precision and Digital Agriculture: Adoption of technologies and perception of Brazilian farmers. *Agriculture* **2020**, *10*, 653. [CrossRef]
4. Rada, N. Assessing Brazil's Cerrado agricultural miracle. *Food Policy* **2013**, *38*, 146–155. [CrossRef]

5. Santana, C.; Souza, G.; Campos, S.; Sanches, I.; Gomes, E.; Sano, E. Dinâmicas agropecuárias e socioeconômicas no Cerrado, de 1975 a 2015. In *Dinâmica agrícola no Cerrado: Análises e Projeções*; Bolfe, É., Sano, E., Campos, S., Eds.; Embrapa: Brasília, Brazil, 2020; Volume 1, pp. 1–308.

6. Sano, E.E.; Rodrigues, A.A.; Martins, E.S.; Bettiol, G.M.; Bustamante, M.M.C.; Bezerra, A.S.; Couto, A.F.; Vasconcelos, V.; Schüler, J.; Bolfe, E.L. Cerrado ecoregions: A spatial framework to assess and prioritize Brazilian Savanna environmental diversity for conservation. *J. Environ. Manag.* **2019**, *232*, 818–828. [CrossRef] [PubMed]

7. Pimenta, F.M.; Speroto, A.T.; Costa, M.H.; Dionizio, E.A. Historical changes in land use and suitability for future agriculture expansion in Western Bahia, Brazil. *Remote Sens.* **2021**, *13*, 1088. [CrossRef]

8. Chaves, M.E.D.; Soares, A.R.; Sanches, I.D.; Fronza, J.G. CBERS data cubes for land use and land cover mapping in the Brazilian Cerrado Agricultural Belt. *Int. J. Remote Sens.* **2021**, *42*, 8398–8432. [CrossRef]

9. Rattis, L.; Brando, P.M.; Macedo, M.N.; Spera, S.A.; Castanho, A.D.A.; Marques, E.Q.; Costa, N.Q.; Silverio, D.V.; Coe, M.T. Climatic limit for agriculture in Brazil. *Nat. Clim. Chang.* **2021**, *11*, 1098–1104. [CrossRef]

10. Araújo, M.L.S.; Sano, E.E.; Bolfe, É.L.; Santos, J.R.N.; dos Santos, J.S.; Silva, F.B. Spatiotemporal dynamics of soybean crop in the Matopiba region, Brazil (1990–2015). *Land Use Policy.* **2019**, *80*, 57–67. [CrossRef]

11. Portman, F.T.; Siebert, S.; Döll, P. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Glob. Biogeochem. Cycles* **2010**, *24*, GB1011. [CrossRef]

12. Sano, E.E.; Rosa, R.; Brito, J.L.S.; Ferreira, L.G. Land cover mapping of the tropical Savanna region in Brazil. *Environ. Monit. Assess.* **2010**, *166*, 113–124. [CrossRef]

13. Prudente, V.H.R.; Martins, V.S.; Vieira, D.C.; e Silva, N.R.D.F.; Adami, M.; Sanches, I.D. Limitations of cloud cover for optical remote sensing of agricultural areas across South America. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100414. [CrossRef]

14. Luo, Y.; Guan, K.; Peng, J. A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product. *Remote Sens. Environ.* **2018**, *214*, 87–99. [CrossRef]

15. Gao, F.; Masek, F.; Schwaller, M.; Hall, F. STAIR: On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218. [CrossRef]

16. Gao, F.; Anderson, M.C.; Zhang, X.; Yang, Z.; Alfieri, J.G.; Kustas, W.P.; Mueller, R.; Johnson, D.M.; Prueger, J.H. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.* **2017**, *188*, 9–25. [CrossRef]

17. Sun, R.; Chen, S.; Su, H.; Mi, C.; Jin, N. The Effect of NDVI time series density derived from spatiotemporal fusion of multisource remote sensing data on crop classification accuracy. *ISPRS Int. J. Geo. Inf.* **2019**, *8*, 502. [CrossRef]

18. Dhillon, M.S.; Dahms, T.; Kübert-Flock, C.; Steffan-Dewenter, I.; Zhang, J.; Ullmann, T. Spatiotemporal fusion modelling using STARFM: Examples of Landsat 8 and Sentinel-2 NDVI in Bavaria. *Remote Sens.* **2022**, *14*, 677. [CrossRef]

19. Zhu, X.; Cai, F.; Tian, J.; Williams, T.K.-A. Spatiotemporal fusion of multisource remote sensing data: Literature survey, Taxonomy, principles, applications, and future directions. *Remote Sens.* **2018**, *10*, 527. [CrossRef]

20. Zhu, X.; Chen, J.; Gao, F.; Chen, X.; Masek, J.G. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623. [CrossRef]

21. Hilker, T.; Wulder, M.A.; Coops, N.C.; Linke, J.; McDermid, G.; Masek, J.G.; Gao, F.; White, J.C. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sens. Environ.* **2009**, *113*, 1613–1627. [CrossRef]

22. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.-C.; Skakun, S.; Justice, C. The Harmonized Landsat and Sentinel-2 Surface reflectance data set. *Remote Sens. Environ.* **2018**, *219*, 145–161. [CrossRef]

23. Skakun, S.; Vermote, E.; Franch, B.; Roger, J.-C.; Kussul, N.; Ju, J.; Masek, J. Winter wheat yield assessment from Landsat 8 and Sentinel-2 Data: Incorporating surface reflectance, through phenological fitting, into regression yield models. *Remote Sens.* **2019**, *11*, 1768. [CrossRef]

24. Zhang, H.K.; Roy, D.P.; Yan, L.; Li, Z.; Huang, H.; Vermote, E.; Skakun, S.; Roger, J.-C. Characterization of Sentinel-2A and Landsat-8 Top of Atmosphere, surface, and Nadir BRDF adjusted reflectance and NDVI differences. *Remote Sens. Environ.* **2018**, *215*, 482–494. [CrossRef]

25. Franch, B.; Vermote, E.; Skakun, S.; Roger, J.-C.; Masek, J.; Ju, J.; Villaescusa-Nadal, J.; Santamaria-Artigas, A. A method for Landsat and Sentinel 2 (HLS) BRDF normalization. *Remote Sens.* **2019**, *11*, 632. [CrossRef]

26. Torbick, N.; Huang, X.; Ziniti, B.; Johnson, D.; Masek, J.; Reba, M. Fusion of moderate resolution earth observations for operational crop type mapping. *Remote Sens.* **2018**, *10*, 1058. [CrossRef]

27. Hao, P.; Tang, H.; Chen, Z.; Yu, L.; Wu, M. High resolution crop intensity mapping using Harmonized Landsat-8 and Sentinel-2 data. *J. Integr. Agric.* **2019**, *18*, 2883–2897. [CrossRef]

28. Dong, T.; Liu, J.; Qian, B.; He, L.; Liu, J.; Wang, R.; Jing, Q.; Champagne, C.; McNairn, H.; Powers, J.; et al. Estimating crop biomass using leaf area index derived from Landsat 8 and Sentinel-2 data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 236–250. [CrossRef]

29. Gao, F.; Anderson, M.; Daughtry, C.; Karnieli, A.; Hively, D.; Kustas, W. A within-season approach for detecting early growth stages in corn and soybean using high temporal and spatial resolution imagery. *Remote Sens. Environ.* **2020**, *242*, 111752. [CrossRef]

30. Bolognesi, S.F.; Pasolli, E.; Belfiore, O.; de Michele, C.; D'Urso, G. Harmonized Landsat 8 and Sentinel-2 time Sseries data to detect irrigated areas: An application in southern Italy. *Remote Sens.* **2020**, *12*, 1275. [CrossRef]

31. Griffiths, P.; Nendel, C.; Pickert, J.; Hostert, P. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. *Remote Sens. Environ.* **2020**, *238*, 111124. [CrossRef]

32. Zhou, Q.; Rover, J.; Brown, J.; Worstell, B.; Howard, D.; Wu, Z.; Gallant, A.; Rundquist, B.; Burke, M. Monitoring landscape dynamics in Central U.S. grasslands with Harmonized Landsat-8 and Sentinel-2 time series data. *Remote Sens.* **2019**, *11*, 328. [CrossRef]

33. Pastick, N.J.; Dahal, D.; Wylie, B.K.; Parajuli, S.; Boyte, S.P.; Wu, Z. Characterizing land surface phenology and exotic annual grasses in dryland ecosystems using Landsat and Sentinel-2 Data in harmony. *Remote Sens.* **2020**, *12*, 725. [CrossRef]

34. Bolton, D.K.; Gray, J.M.; Melaas, E.K.; Moon, M.; Eklundh, L.; Friedl, M.A. Continental-scale land surface phenology from Harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sens. Environ.* **2020**, *240*, 111685. [CrossRef]

35. IBGE. Produção Agrícola Municipal. Available online: https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9117-producao-agricola-municipal-culturas-temporarias-e-permanentes.html?=&t=destaques (accessed on 15 February 2022).

36. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; Gonçalves, J.L.M.; Sparovek, G. Köppen's climate classification map for Brazil. *Meteorol. Z.* **2013**, *22*, 711–728. [CrossRef]

37. INMET. Tabelas de Dados das Estações. Available online: https://tempo.inmet.gov.br/TabelaEstacoes/A404 (accessed on 15 February 2022).

38. IBGE. Pedologia 1:250,000. Available online: https://www.ibge.gov.br/geociencias/informacoes-ambientais/pedologia/10871-pedologia.html?=&t=acesso-ao-produto (accessed on 20 June 2022).

39. USGS. Landsat 8–9 Collection 2 (C2) Level 2 Science Product (L2SP) Guide. Available online: https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/media/files/LSDS-1619_Landsat-8-9-C2-L2-ScienceProductGuide-v4.pdf (accessed on 15 February 2022).

40. Masek, J.; Ju, J.; Claverie, M.; Skakun, S.; Roger, J.-C.; Vermote, E.; Franch, B.; Yin, Z.; Dungan, J. Harmonized Landsat Sentinel-2 (HLS) Product User Guide—Product Version 2.0. Available online: https://lpdaac.usgs.gov/documents/1326/HLS_User_Guide_V2.pdf (accessed on 15 February 2022).

41. Rouse, J.W.; Haas, R.W.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Greatplains with ERTS. In Proceedings of the Third ERTS—1 Symposium, NASA Goddard Space Flight Center, Washington, DC, USA, 10–14 December 1974; Volume 1, pp. 309–317.

42. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]

43. Gao, B. NDWI—A Normalized Difference Water Index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **1996**, *58*, 257–266. [CrossRef]

44. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]

45. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]

46. Souza, 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.; et al. Reconstructing three decades of land use and land cover changes in Brazilian biomes with Landsat archive and Earth Engine. *Remote Sens.* **2020**, *12*, 2735. [CrossRef]

47. IBGE. Áreas Urbanizadas do Brasil: 2015. Available online: https://www.ibge.gov.br/geociencias/cartas-e-mapas/redes-geograficas/15789-areas-urbanizadas.html?=&t=acesso-ao-produto (accessed on 15 February 2022).

48. Spinelli-Araújo, L.; Vicente, L.E.; Manzatto, C.V.; Skorupa, L.A.; Victoria, D.D.C.; Soares, A.R. AgroTag: Um sistema de coleta, análise e compartilhamento de dados de campo para qualificação do uso e cobertura das Terras no Brasil. In Proceedings of the XIX Simpósio Brasileiro de Sensoriamento Remoto; INPE—Instituto Nacional de Pesquisas Espaciais, Santos, Brazil, 14–17 April 2019; pp. 451–454.

49. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2019.

50. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

51. Biau, G.; Scornet, E. A Random Forest guided tour. *Test* **2016**, *25*, 197–227. [CrossRef]

52. Fang, P.; Zhang, X.; Wei, P.; Wang, Y.; Zhang, H.; Liu, F.; Zhao, J. The Classification performance and mechanism of machine learning algorithms in winter wheat mapping using Sentinel-2 10 m resolution imagery. *Appl. Sci.* **2020**, *10*, 5075. [CrossRef]

53. Belgiu, M.; Drăguţ, L. Random Forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]

54. Arnholt, A.T.; Evans, B. Basic Statistics and Data Analysis (BSDA). Available online: https://cran.r-project.org/web/packages/BSDA/BSDA.pdf (accessed on 25 March 2022).

55. MAPA. Zoneamento Agrícola de Risco Climático—Safra 2021/22. Available online: https://indicadores.agricultura.gov.br/zarc/index.htm (accessed on 1 December 2021).

56. Dahal, D.; Pastick, N.J.; Boyte, S.P.; Parajuli, S.; Oimoen, M.J.; Megard, L.J. Multi-species inference of exotic annual and native perennial grasses in rangelands of the western United States using Harmonized Landsat and Sentinel-2 data. *Remote Sens.* **2022**, *14*, 807. [CrossRef]

57. Scaramuzza, C.A.M.; Sano, E.E.; Adami, M.; Bolfe, E.L.; Coutinho, A.C.; Esquerdo, J.C.D.M.; Maurano, L.E.P.; Narvaes, I.S.; Oliveira Filho, F.J.B.; Rosa, R.; et al. Land-use and land-cover mapping of the Brazilian Cerrado based mainly on Landsat-8 Satellite images. *RBC* **2017**, *69*, 1041–1051.

58. Nguyen, M.; Baez-Villanueva, O.; Bui, D.; Nguyen, P.; Ribbe, L. Harmonization of Landsat and Sentinel 2 for crop monitoring in drought prone areas: Case studies of Ninh Thuan (Vietnam) and Bekaa (Lebanon). *Remote Sens.* **2020**, *12*, 281. [CrossRef]

59. Ajadi, O.A.; Barr, J.; Liang, S.-Z.; Ferreira, R.; Kumpatla, S.P.; Patel, R.; Swatantran, A. Large-scale crop type and crop area mapping across Brazil using Synthetic Aperture Radar and optical imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *97*, 102294. [CrossRef]

60. Wang, S.; Azzari, G.; Lobell, D.B. Crop type mapping without field-level labels: Random Forest transfer and unsupervised clustering techniques. *Remote Sens. Environ.* **2019**, *222*, 303–317. [CrossRef]

61. Bolfe, E.L.; Barbedo, J.G.A.; Massruhá, S.M.F.S.; de Souza, K.X.S.; Assad, E.D. Desafios, Tendências e Oportunidades em Agricultura Digital no Brasil. In *Agricultura Digital: Pesquisa, Desenvolvimento e Inovação nas Cadeias Produtivas*; EMBRAPA: Brasília, Brazil, 2020; Volume 1, pp. 1–406.

62. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]

63. Zhu, Z.; Wulder, M.A.; Roy, D.P.; Woodcock, C.E.; Hansen, M.C.; Radeloff, V.C.; Healey, S.P.; Schaaf, C.; Hostert, P.; Strobl, P.; et al. Benefits of the free and open Landsat data policy. *Remote Sens. Environ.* **2019**, *224*, 382–385. [CrossRef]

64. Picoli, M.C.A.; Camara, G.; Sanches, I.; Simões, R.; Carvalho, A.; Maciel, A.; Coutinho, A.; Esquerdo, J.; Antunes, J.; Begotti, R.A.; et al. Big Earth Observation time series analysis for monitoring brazilian agriculture. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 328–339. [CrossRef]

65. Oldoni, L.V.; Sanches, I.D.; Picoli, M.C.A.; Covre, R.M.; Fronza, J.G. LEM+ dataset: For agricultural remote sensing applications. *Data Br.* **2020**, *33*, 106553. [CrossRef] [PubMed]

66. Esquerdo, J.C.D.M.; Antunes, J.F.G.; Coutinho, A.C.; Speranza, E.A.; Kondo, A.A.; dos Santos, J.L. SATVeg: A web-based tool for visualization of MODIS vegetation indices in South America. *Comput. Electron. Agric.* **2020**, *175*, 105516. [CrossRef]

67. Wardlow, B.; Egbert, S.; 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.* **2007**, *108*, 290–310. [CrossRef]

68. Martinez, J.A.C.; Rosa, L.E.C.; Feitosa, R.Q.; Sanches, I.D.; Happ, P.N. Fully convolutional recurrent networks for multidate crop recognition from multitemporal image sequences. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 188–201. [CrossRef]

69. Bendini, H.N.; Fonseca, L.M.G.; Schwieder, M.; Körting, T.S.; Rufin, P.; Sanches, I.D.; Leitão, P.J.; Hostert, P. Detailed agricultural land classification in the Brazilian Cerrado based on phenological information from dense satellite image time series. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101872. [CrossRef]

70. Montibeller, B.; Sanches, I.D.; Luiz, A.J.B.; Gonçalves, F.; Aguiar, D.A. Spectral-temporal profile analysis of maize, soybean and sugarcane based on OLI/Landsat-8 data. *Braz. J. Agric.* **2019**, *94*, 242–258. [CrossRef]

71. Epiphanio, R.D.V.; Formaggio, A.R.; Rudorff, B.F.T.; Maeda, E.E.; Luiz, A.J.B. Estimating soybean crop areas using spectral-temporal surfaces derived from MODIS images in Mato Grosso, Brazil. *Pesqui. Agropecu. Bras.* **2010**, *45*, 72–80. [CrossRef]

72. Arvor, D.; Meirelles, M.; Dubreuil, V.; Bégué, A.; Shimabukuro, Y.E. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Appl. Geogr.* **2012**, *32*, 702–713. [CrossRef]

73. Chen, Y.; Lu, D.; Moran, E.; Batistella, M.; Dutra, L.V.; Sanches, I.D.; da Silva, R.F.B.; Huang, J.; Luiz, A.J.B.; de Oliveira, M.A.F. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *69*, 133–147. [CrossRef]

74. Wang, N.; Zhai, Y.; Zhang, L. Automatic cotton mapping using time series of Sentinel-2 images. *Remote Sens.* **2021**, *13*, 1355. [CrossRef]

75. Xun, L.; Zhang, J.; Cao, D.; Yang, S.; Yao, F. A novel cotton mapping index combining Sentinel-1 SAR and Sentinel-2 multispectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 148–166. [CrossRef]

76. Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* **2017**, *2017*, 1–17. [CrossRef]

77. Vilela, L.; Manjabosco, E.A.; Marchão, R.L.; Guimarães Júnior, R. *Integrated Crop-Livestock in Western Bahia State: The Off-Season Cattle Model*; (Circular Técnica 37); Embrapa Cerrados: Planaltina, Brazil, 2018. [CrossRef]

78. Beuchle, R.; Grecchi, R.C.; Shimabukuro, Y.E.; Seliger, R.; Eva, H.D.; Sano, E.; Achard, F. Land cover changes in the brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. *Appl. Geogr.* **2015**, *58*, 116–127. [CrossRef]

79. Müller, H.; Rufin, P.; Griffiths, P.; Barros Siqueira, A.J.; Hostert, P. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian Savanna landscape. *Remote Sens. Environ.* **2015**, *156*, 490–499. [CrossRef]

80. Small, C.; Sousa, D. Spatiotemporal characterization of mangrove phenology and disturbance response: The angladesh Sundarban. *Remote Sens.* **2019**, *11*, 2063. [CrossRef]

81. Shen, Y.; Zhang, X.; Yang, Z. Mapping corn and soybean phenometrics at field scales over the United States Corn Belt by fusing time series of Landsat 8 and Sentinel-2 data with VIIRS data. *ISPRS J. Photogramm. Remote Sens.* **2022**, *186*, 55–69. [CrossRef]

82. Gao, F.; Anderson, M.C.; Johnson, D.M.; Seffrin, R.; Wardlow, B.; Suyker, A.; Diao, C.; Browning, D.M. Towards routine mapping of crop emergence within the season using the Harmonized Landsat and Sentinel-2 dataset. *Remote Sens.* **2021**, *13*, 5074. [CrossRef]

83. Nguyen, L.H.; Henebry, G.M. Characterizing land use/land cover using multi-sensor time series from the perspective of land surface phenology. *Remote Sens.* **2019**, *11*, 1677. [CrossRef]

84. Wang, J.; Xiao, X.; Liu, L.; Wu, X.; Qin, Y.; Steiner, J.L.; Dong, J. Mapping sugarcane plantation dynamics in Guangxi, China, by time series Sentinel-1, Sentinel-2 and Landsat images. *Remote Sens. Environ.* **2020**, *247*, 111951. [CrossRef]

85. Xu, F.; Li, Z.; Zhang, S.; Huang, N.; Quan, Z.; Zhang, W.; Liu, X.; Jiang, X.; Pan, J.; Prishchepov, A.V. Mapping winter wheat with combinations of temporally aggregated Sentinel-2 and Landsat-8 data in Shandong Province, China. *Remote Sens.* **2020**, *12*, 2065. [CrossRef]

86. Parreiras, T.C.; Bolfe, E.B.; Sano, S.E.; Victoria, D.C.; Sanches, I.D.; Vicente, L.E. Exploring the Harmonized Landsat and Sentinel-2 (HLS) datacube to map an agricultural landscape in the Brazilian Savanna. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2022**, *43*, 967–973. [CrossRef]

87. Blickensdörfer, L.; Schwieder, M.; Pflugmacher, D.; Nendel, C.; Erasmi, S.; Hostert, P. Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sens. Environ.* **2022**, *269*, 112831. [CrossRef]

88. Liu, L.; Xiao, X.; Qin, Y.; Wang, J.; Xu, X.; Hu, Y.; Qiao, Z. Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine. *Remote Sens. Environ.* **2020**, *239*, 111624. [CrossRef]

89. Xue, J.; Anderson, M.C.; Gao, F.; Hain, C.; Yang, Y.; Knipper, K.R.; Kustas, W.P.; Yang, Y. Mapping daily evapotranspiration at field scale using the Harmonized Landsat and Sentinel-2 dataset, with sharpened VIIRS as a Sentinel-2 thermal proxy. *Remote Sens.* **2021**, *13*, 3420. [CrossRef]

90. Xue, J.; Anderson, M.C.; Gao, F.; Hain, C.; Sun, L.; Yang, Y.; Knipper, K.R.; Kustas, W.P.; Torres-Rua, A.; Schull, M. Sharpening ECOSTRESS and VIIRS Land Surface Temperature using Harmonized Landsat-Sentinel surface reflectances. *Remote Sens. Environ.* **2020**, *251*, 112055. [CrossRef]

91. Schwieder, M.; Wesemeyer, M.; Frantz, D.; Pfoch, K.; Erasmi, S.; Pickert, J.; Nendel, C.; Hostert, P. Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. *Remote Sens. Environ.* **2022**, *269*, 112795. [CrossRef]

92. Roy, D.P.; Huang, H.; Boschetti, L.; Giglio, L.; Yan, L.; Zhang, H.H.; Li, Z. Landsat-8 and Sentinel-2 burned area mapping—A combined sensor multi-temporal change detection approach. *Remote Sens. Environ.* **2019**, *231*, 111254. [CrossRef]

93. Tulbure, M.G.; Broich, M.; Perin, V.; Gaines, M.; Ju, J.; Stehman, S.; Pavelsky, T.; Masek, J.G.; Yin, S.; Mai, J.; et al. Can we detect more ephemeral floods with higher density Harmonized Landsat Sentinel 2 data compared to Landsat 8 alone? *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 232–246. [CrossRef]

94. Lechler, S.; Picoli, M.C.A.; Soares, A.R.; Sanchez, A.; Chaves, M.E.D.; Verstegen, J. Exploring Nasa's Harmonized Landsat and Sentinel-2 (HLS) dataset to monitor deforestation in the Amazon Rainforest. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *43*, 705–711. [CrossRef]

95. Zhang, Y.; Ling, F.; Wang, X.; Foody, G.M.; Boyd, D.S.; Li, X.; Du, Y.; Atkinson, P.M. Tracking small-scale tropical forest disturbances: Fusing the Landsat and Sentinel-2 data record. *Remote Sens. Environ.* **2021**, *261*, 112470. [CrossRef]

96. Chaves, M.E.D.; Picoli, M.C.A.; Sanches, I.D. Recent Applications of Landsat 8/OLI and Sentinel-2/MSI for Land Use and Land Cover Mapping: A Systematic Review. *Remote Sens.* **2020**, *12*, 3062. [CrossRef]