

Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil



Luan Pierre Pott^{a,b,*}, Telmo Jorge Carneiro Amado^c, Raí Augusto Schwalbert^{b,d}, Geomar Mateus Corassa^d, Ignacio Antonio Ciampitti^{b,*}

^a Agricultural Engineering Department, Federal University of Santa Maria, Rural Science Centre, Santa Maria, Brazil

^b Department of Agronomy, Kansas State University, Manhattan, KS, USA

^c Soil Department, Federal University of Santa Maria, Rural Science Centre, Santa Maria, Brazil

^d Cooperativa Central Gaúcha Ltd. – CCGL, Cruz Alta, Rio Grande do Sul, Brazil

ARTICLE INFO

Keywords:

Agricultural monitoring
Remote sensing
Machine learning
Sentinel-1
Sentinel-2
SRTM Digital Elevation

ABSTRACT

Field-scale crop monitoring is essential for agricultural management and policy making for food security and sustainability. Automating crop classification process while elaborating a workflow is a key step for reliable and precise crop mapping. This study aims to develop an approach for crop classification in the state of Rio Grande do Sul, Brazil, following the specific goals of i) evaluating spatial satellite-based features to guide crop data collection; ii) testing transfer learning model with subsequent growing season data; iii) examining accuracy in early-season prediction model; and lastly, iv) developing a crop classification model for estimating large scale crop area. As main data inputs, Sentinel-2, Sentinel-1, and Shuttle Radar Topographic Mission (SRTM) Digital Elevation data were used to extract features to input in the Random Forest classifier. Spatial variability of satellite features was evaluated using Moran's I Index and cluster k-means. Crop area prediction data were obtained at municipality level to compare with census data (standard method). A crop summer map layer was generated for three major crops: soybeans (*Glycine max* L.), corn (*Zea mays* L.), and rice (*Oryza sativa* L.) in the state of Rio Grande do Sul, Brazil. The crop classification model achieved an overall accuracy of 0.95. Model performance was influenced by sample size and spatial variability of the samples. The random forest model was transferred to the next growing season with 0.89 and 0.91 overall accuracy for 250 and 750 samples, respectively. However, overall accuracy increased from 0.93 to 0.95 when 50 to 250 samples of same-year data was aggregated to the model. Similar accuracy was obtained for predictions done with data until March relative to when the entire season was considered, until May. When data for more growing seasons were aggregated, the model produced more accurate early season predictions (January and February). Soybean prediction area obtained the highest performance ($R^2 = 0.94$), relative to rice ($R^2 = 0.90$) and corn ($R^2 = 0.37$). The rice prediction area presented a high precision, but the crop area was overestimated due to errors with wetland target relative to other class. Lastly, this study presents the first crop map layer of the three major field crops for the state of Rio Grande do Sul, Brazil, serving as a foundation for the creation of crop type maps for other states in the country and around the globe.

1. Introduction

Crop monitoring of agricultural systems can provide information for policymakers related to food production and sustainability, identify research priorities, and evaluate commodity markets (Van Ittersum et al., 2013; Carletto et al., 2015a). The first step to monitor agricultural crops is obtaining crop mapping information at the field-scale with high

accuracy of land crop regions to improve agricultural policies.

Field surveys and census are the standard method for obtaining data about crop area for most agricultural regions (Wang et al., 2019). In Brazil, data collection is achieved by both the Brazilian Institute of Geography and Statistics (IBGE) and the National Supply Company (CONAB). While survey and census provide valuable crop data, limitations related to high cost, spatial inference level, frequency of updates,

* Corresponding authors at: Agricultural Engineering Department, Federal University of Santa Maria, Rural Science Centre, Santa Maria, Brazil. Department of Agronomy, Kansas State University, Manhattan, KS, USA.

E-mail addresses: luanpierrepott@hotmail.com (L.P. Pott), ciampitti@ksu.edu (I.A. Ciampitti).

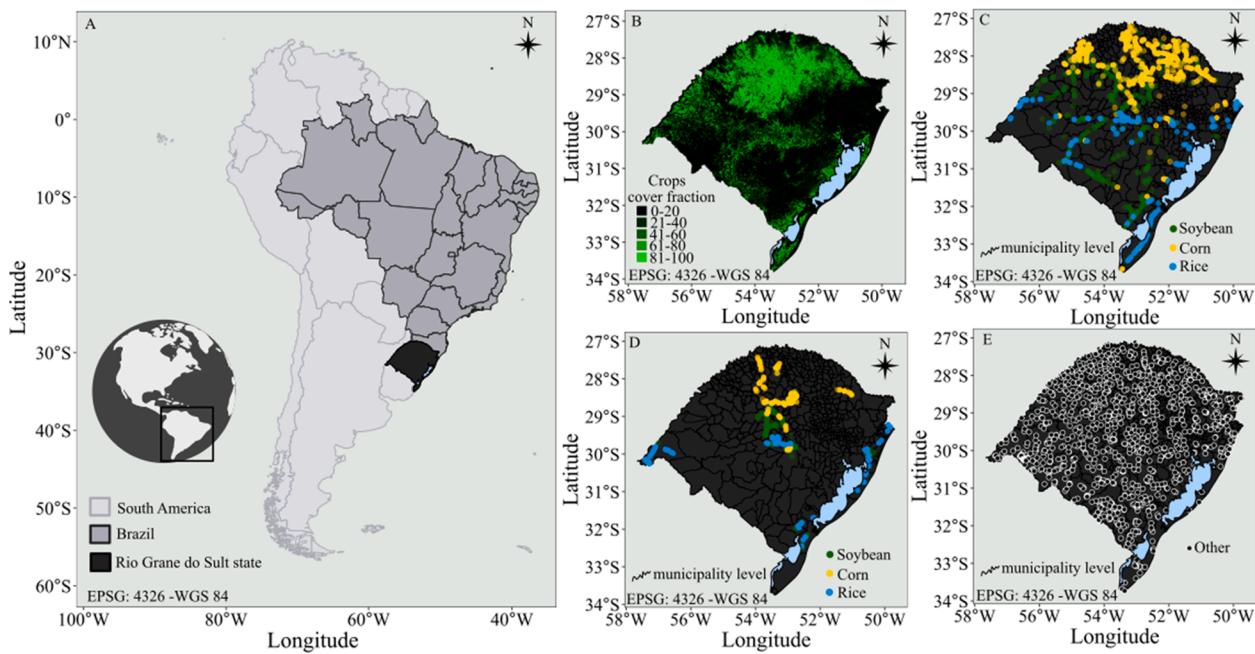


Fig. 1. (A) South America map, Brazil (shaded in grey), and Rio Grande do Sul state (shaded in black). (B) Copernicus Global Land Cover Layers – crops cover fraction 2019 (Buchhorn et al., 2020). (C) Crop type data points accessed by Google Street View for 2018–2019 growing season. (D) Crop type data points collected manually utilizing GNSS system for 2019–2020 growing season. (E) Other classes (pasture/native vegetation, forest, urban, water, and other crops data points) retrieved from Google Street View and Google Earth Engine.

and subjectivity to bias are main constraints linked to these methods (Carletto et al., 2015b; Gourlay et al., 2017).

In Brazil, crop type classification has been studied using satellite data MODIS with maximum likelihood estimation (Rudorff et al., 2007; Arvor et al., 2011); unsupervised classification, multi-spectral and -temporal visual interpretation (Lemos, 2016); and decision tree classifier (Zhong et al., 2016). Mengue and Fontana (2015) proposed a parallelepiped classifier from MODIS and Shuttle Radar Topographic Mission (SRTM) data. Schultz et al. (2015) utilized random forests with Landsat 8 images. Silva Junior et al. (2020) proposed Perpendicular Crop Enhancement Index (PCEI) as a threshold decision from Sentinel-2, Landsat and MODIS data. Furthermore, the CONAB has been proposing satellite-based crop monitoring for major crops such as soybean (*Glycine max* (L.) Merr.), corn (*Zea mays* L.) and rice (*Oryza sativa* L.) in order to overcome challenges due to the high spatial variability, such as large range of latitude, terrain, multiple crops and high variability of management of Brazilian production (CONAB, 2020a).

The advent of more accessible satellite data and new statistical methods in machine learning will assist a fast implementation of crop classification approaches to enable agricultural monitoring at high spatio-temporal resolutions (Balzter et al., 2015; Bargiel, 2017; Cai et al., 2018; Jin et al., 2019; Wang et al., 2020; Weiss et al., 2020). The Sentinel-2 and Sentinel-1 satellites offer near-real time images with high spatial (10–60 m) and temporal (1–5 days) resolution. Recent studies have reported high performance of optical data for crop mapping over different cropping systems and climates (Ozdogan and Gutman, 2008; Belgii and Csillik, 2018; Wang et al., 2019). Additionally, recent investigations have portrayed benefits of combining data from optical (e.g., Sentinel-2 or Landsat archives) and radar (e.g., SAR as the Sentinel-1) for improving model performance for crop mapping (Abubakar et al., 2020; Inglada et al., 2016; Sonobe et al., 2017; Jin et al., 2019; Orynbaykzy et al., 2020). Furthermore, features from SRTM aggregated with different input sources have significant relevancy for improving land cover (Balzter et al., 2015; Li et al., 2020) and for in-season crop mapping (Demarez et al., 2019).

Crop classification models need proper classification algorithms and

quality ground truth data to train and test models. Data beyond the years included in the model are also required to test spatio-temporal transferability and to foster better in-season predictions. From the many classification techniques, machine learning algorithms and supervised classifications have been globally used. Among those methods, random forest models are highlighted (Demarez et al., 2019; Jin et al., 2019; Wang et al., 2020; Deines et al., 2020). Random forests are an ensemble machine learning method comprised of many decision trees in aggregate (Breiman, 2001), which usually outperform other learning methods (Kayad et al., 2019). From the ground truthing perspective, spatial distribution of crop data should be collected to fairly represent gradients of environment and management for each crop class in the region of study (Waldner et al., 2019; Fowler et al., 2020).

Brazil is the largest country in South America and the fifth largest in the world, presenting a high range of latitude and variability in terrain, soils, crops, and agricultural management. The state of Rio Grande do Sul is a key agricultural producer for the country, but still presents a high spatial variability creating challenges for large-scale crop monitoring. Thus, high quality survey data is essential for developing a stable satellite-based model transferable to other environments (regions) and growing seasons for crop classification. Identification and proper characterization of crop phenology patterns for various crops, regions, and seasons can lead to successful transfer learning performance in classification models (Cai et al., 2018; Wang et al., 2019; Hao et al., 2020a). Improving near-real time crop mapping prediction will have large impacts in agricultural and socioeconomic decision-making procedures such as crop insurance, supply-chain logistics, and financial market forecasting. For in-season crop classification, instead of using satellite data from the entire season, many studies have evaluated the effect of time series length on crop classification performance (Cai et al., 2018; Hao et al., 2020a; Hao et al., 2020b). Demarez et al. (2019) reported an improvement in random forest classifier for mapping irrigated crops with SRTM data mainly for early season classification.

Thus, the aims of this study were to i) evaluate spatial satellite-based features to guide crop data collection; ii) testing transfer learning model with subsequent growing season data; iii) examine accuracy in early-

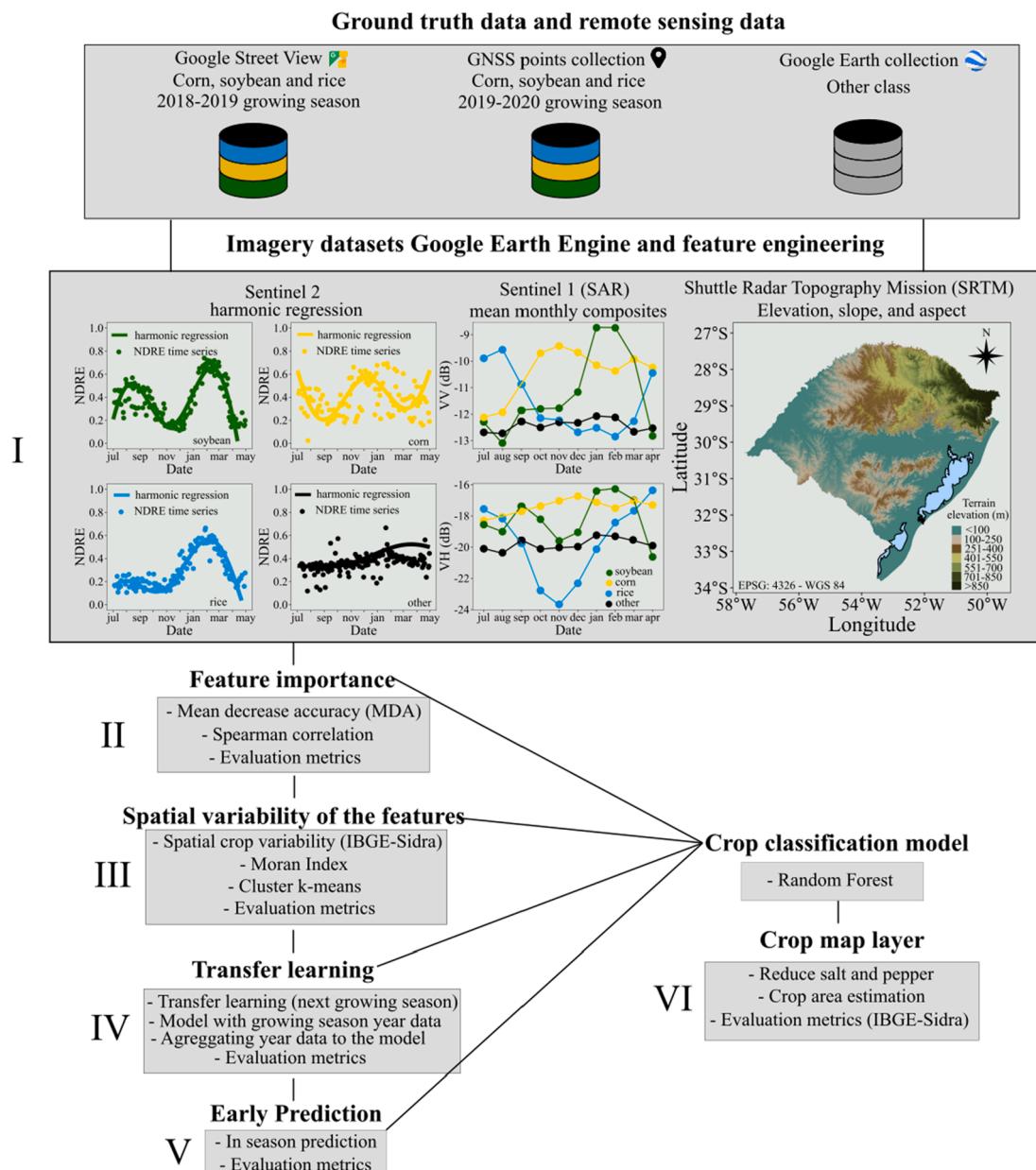


Fig. 2. Framework for model development. Data acquisition – crop type points based on Google Street View for 2018–2019 growing season; crop type points collected utilizing GNSS manually for 2019–2020 growing season; other class collected by visually assessing high-resolution imagery from Google Earth collection. I- Imagery datasets and feature engineering – Google Earth Engine - Sentinel-2 (harmonic regression), Sentinel-1 SAR (mean monthly composition), and SRTM Digital Elevation (elevation, slope, aspect). II- Feature importance of the model. III- Spatial variability of the features. IV- Transfer learning. V- Early prediction. VI- Crop map layer and crop area estimation.

season prediction model; and lastly, iv) develop a crop classification model for mapping and estimating the crop area for soybean, corn, and rice field crops in Rio Grande do Sul state, Brazil.

2. Material and methods

2.1. Study area

The study area consisted of Rio Grande do Sul state (27° to 34° S and 50° to 58° W) in Southern Brazil (Fig. 1A). Rio Grande do Sul state has approximately 281.707 km^2 with 497 municipalities (IBGE, 2019). According to Alvares et al. (2013), the dominant climate types are Humid subtropical without dry season with -hot summer (Cfa), and with -temperate summer (Cfb). Rio Grande do Sul has a wide variety of soil

types with predominance of Oxisols, Alfisols, Mollisols, Ultisols, Entisols and Inceptisols (Santos et al., 2018; Soil Survey Staff, 2014). The state is ranked first nationally in irrigated rice production, third in soybean, and sixth in corn. The projection for soybean cropping area is 6.06 Mha with a production of 19.8 Mt, irrigated rice area is 0.969 Mha with a production of 7.6 Mt, the corn area is 0.808 Mha with a production of 3.6 Mt in the 2020–21 growing season (CONAB, 2021). These three grain crops comprise more than 81% of the crop area in the state (IBGE, 2020). Soybean is planted from October to December and harvested from February to May (CONAB, 2020b). Corn is planted in a wide window from August to January, and harvested from January to July (CONAB, 2020b). Rice, concentrated in the southern and central regions of the state, is planted from September to January and harvested from February to May (CONAB, 2020b).

2.2. Data acquisition

The ground truth crop data were collected following two approaches. The first, via Google Street View, presents a compendium of panoramic images around the globe gathering geographic coordinates and time of the collected data. Google Street View has been recently used in several scientific studies (Ringland et al., 2019; Yan and Ryu, 2021). Yan and Ryu (2021) developed an automatic ground truthing via Google Street View images utilizing convolutional neural network model for ground referencing and crop mapping. Data from Google Street View has been manually searched specifying latitude and longitude coordinates at intervals of around 500–1000 m considering crop cover fractions (Fig. 1B) along roads in the state of Rio Grande do Sul. The Google Street View database in Rio Grande do Sul contains images from October 2018 to April 2019. Imagery including one of the three crops (soybean, corn, or rice), geographic coordinates, and time were retrieved from this database (Fig. 1C). The second collection method utilized a Global Navigation Satellite System (GNSS) portable receiver, GPSmap 62®, to gather crop information along roads throughout the state of Rio Grande do Sul for 2019–2020 growing season (Fig. 1D). The dataset for the 2018–2019 season contained 1,000 data points (Google Street View) for each crop, while the dataset for 2019–2020 season (field survey-based) generated 500 data points for each crop.

Other class were collected by visually assessing high-resolution imagery from Google Earth, including water, urban, forest, pasture, and native vegetation (Fig. 1E). Species of native vegetation for the Southern Brazilian grasslands are commonly *Andropogon* spp., *Axonopus* spp., *Aristida* spp., and *Paspalum* spp. (Andrade et al., 2019). Other summer crops such as horticulture crops, tobacco (*Nicotiana tabacum* L.), potato (*Solanum tuberosum* L.), yerba mate (*Ilex paraguariensis* A. St. -Hil.), grape (*Vitis* spp.), apple (*Malus domestica* Borkh.), walnut (*Carya illinoiensis* (Wangengh) K. Koch), olive (*Olea europaea* L.) and orange (*Citrus* spp.) were covered assessing Google Street View and Google Earth and belong to other class.

2.3. Remote sensing data and feature engineering

Optical, radar satellite, and digital elevation data sources were combined to build a crop classification model using Google Earth Engine (GEE) (Fig. 2 I).

2.3.1. Sentinel-2

Sentinel-2 satellites carry an optical sensor Multi-Spectral Instrument (MSI), which acquires data with a wide-swath and high-resolution (10 m). Images are gathered every 5 days supporting Copernicus Land Monitoring studies, including vegetation, soil, and water cover. The Sentinel-2 product Level-1C consists of top of atmosphere (TOA) reflectance observations in cartographic geometry. The higher-level product, Level-2A, consists of a process including scene classification and atmospheric corrections applied to Level-1C. Jin et al. (2019) reported Sentinel-2 Level-2A as non-essential for crop classification but important to crop yield estimates. We have selected Sentinel-2 Level-1C due the availability of this data for Rio Grande do Sul state in GEE. Sentinel-2 Level-2A presented a collection of imagery from December 2018 onwards in the region of study, while Sentinel-2 Level-1C has imagery data for the interest period (July 1, 2018 to May 1, 2019, and July 1, 2019 to May 1, 2020). We used the following reflectance bands: blue, green, red, red edge 1, near infrared (NIR), shortwave infrared 1 (SWIR1), and shortwave infrared 2 (SWIR2). In addition, we have derived four vegetation indices (VIs) commonly used for agricultural purposes: NDVI (Tucker, 1979), NDRE (Barnes et al., 2000), GCVI (Gitelson et al., 2005), and EVI2 (Jiang et al., 2008). The time frame for image collection ranged from July 1, 2018 to May 1, 2019, and July 1, 2019 to May 1, 2020 in order to perform harmonic regressions. The imagery data were filtered for less than 20% pixel percentage of clouds, and masked with a Quality Assessment (QA) band provided by European

Space Agency (ESA) that flags clouds and cirrus pixels. Furthermore, we used a decision tree as proposed by Hollstein et al. (2016) to detect contaminated pixels, including clouds, cirrus, shadows, and snow. Decision tree classifiers, such as Hollstein's trees, were completely serialized allowing us to generate scalable engineering through GEE (Hollstein et al. 2016; Jin et al., 2019). The harmonic regression was selected as the method to extract features from the image time series to apply machine learning models. The harmonic regression of optical data has been used as a potential feature for crop classification (Jakubauskas et al., 2002; Ghazaryan et al., 2018), with the advantage of being directly deployable in GEE using the built-in linear regression function (Wang et al., 2019).

Sentinel-2 bands and VIs can be represented by the following harmonic regression function:

$$f_i(t) = c_i + \sum_{k=1}^n [a_{ik}\cos(2\pi\omega_k t) + b_{ik}\sin(2\pi\omega_k t)]$$

where a_{ik} are cosine coefficients, b_{ik} are sine coefficients, and c_i is the intercept term. The independent variable t represents the time of the images taken within a year expressed as a fraction between 0 (July 1) and 1 (May 1).

Increasing the number of harmonic terms (n) results in a closer fit to the function, requiring a balance of fit and overfitting in the function. For the ω , it was necessary find the value which best fit the data. Coefficient analyses were performed using the mean-squared error for the three crop classes (soybean, corn and rice) and other classes (Fig. 2 II), resulting in the best fit utilizing the same coefficients of Wang et al. (2019). Likewise, the coefficients were utilized in Wang et al. (2020), Deines et al. (2020) and Dado et al. (2020) studies for crop classification in different states of the United Sates. The final regression at each Sentinel-2 point for band or VIs is accordingly

$$f_i(t) = c_i + a_{i1}\cos(3\pi t) + b_{i1}\sin(3\pi t) + a_{i2}\cos(6\pi t) + b_{i2}\sin(6\pi t)$$

We have extracted the coefficients c_{i1} , a_{i1} , b_{i1} , a_{i2} and b_{i2} for 7 bands and to 4 VIs, resulting in 55 features for the Sentinel-2 data.

2.3.2. Sentinel-1

The Sentinel-1 mission provides data from a dual-polarization C-band Synthetic Aperture Radar (SAR) instrument. The Sentinel-1 SAR GRD is processed using the Sentinel-1 Toolbox to generate a calibrated, *ortho*-corrected product with 10 m and 6-day revisit-time resolution available in GEE. We have used the dual polarization (VV and VH) and the ratio (VV/VH), resulting in three backscatter bands. Finally, we have computed mean monthly composition from July to April in both years resulting in 30 features for the SAR data.

2.3.3. Shuttle Radar topography Mission (SRTM)

The SRTM digital elevation data is an international research effort for digital elevation models on a near-global scale (Farr et al. 2007). For the purpose of this study, we have used the SRTM V3 product provided by NASA JPL at a resolution of 30-m available in GEE. From this dataset, we have derived elevation, slope, and aspect, resulting in three main features from SRTM digital elevation data.

2.4. Feature selection

Due to the high number of features, we have performed feature selection based on Spearman correlation and mean decrease accuracy (MDA) of the random forest model to infer the complexity of the model and the tradeoff between accuracy and computational cost (Fig. 2 II). Feature selection based on Spearman correlation and MDA retains the properties of the feature engineering, as opposed to the dimensionality reduction using a transformation algorithm such as the principal component analysis (PCA) (Wang et al., 2019).

Table 1

Summary of sample sizes utilized to train the model and transfer learning. All the models evaluated were validated with 250 samples.

Model	Trained samples
Trained model, 2018–2019	50; 150; 250; 350; 450; 550; 650; 750
Transfer year learning	250; 750
Year model, 2019–2020	50; 100; 150; 200; 250
Aggregating year data to trained model	700_50; 650_100; 600_150; 550_200; 500_250

2.5. Random forest model

The features were fused to train the random forest models. Random forest is a machine learning method encompassing of many decision trees in aggregate (Breiman, 2001), and offering a high performance for crop classification, with the algorithm available to perform training in GEE platform with a large possibility of executing transfer learning for an entire area or other years as a transfer year learning (Jin et al., 2019; Wang et al., 2019; Wang et al., 2020). To train the random forest model, we used the *randomForest* (Breiman, 2001) package in the R environment. We tuned the parameter to reduce model variance, and the final parameters resulted in 100 trees and variables per split equal to 8, with all other parameter as default. We used 75% of the data to train and 25% to test the model with overall accuracy, out-of-bag (OOB) error and balanced accuracy using *caret* (Kuhn, 2008) package in R. To create the final Rio Grande do Sul state map, we used Google Earth Engine's *ee.Classifier.smileRandomForest* algorithm.

2.6. Data analyses

We have analyzed the features for spatial crop variability, transfer learning, early prediction, and generated the crop map and statistics. Statistical analyses were completed in R language (R Core Team, 2020) in Google Colaboratory (Colab) (Bisong, 2019). The proposed model was implemented in GEE platform (Gorelick et al., 2017).

2.6.1. Spatial variability of crops features

Spatial variability of the features for each crop throughout the state of Rio Grande do Sul state were evaluated due to the high variability in soils and climate, in addition to the large range of sowing dates and farming management (Fig. 2 III).

First, we have assessed the selected features derived from Sentinel-2,

Sentinel-1, and SRTM with geographic coordinates to perform Moran's I index analyses to verify the spatial dependence of each crop. A good spatial balance when collecting training data is important for classification of remotely-sensed crops, especially when no prior information is available about spatial pattern in crop growth (Fowler et al., 2020). In our study, we have analyzed the features separately for each crop, using the Moran's I index.

Finally, after the Moran's I index analyses we have performed the cluster k-means algorithm. The k-means clustering algorithm has been used in unsupervised crop classification (Wang et al. 2019; Xu et al., 2018). In our study, we have utilized k-means to associate satellite features in cluster regions within the state. As the features presented different magnitude values, we have normalized them on same scale. The number of clusters was defined according to the gap statistic. Gap statistic compares the total within intra-cluster variation for different values of k with their expected values under null reference distribution of the data (Tibshirani et al., 2001).

2.6.2. Sample size and transfer learning

In this study, we have evaluated various sample allocations in the 2018–2019 growing season, testing 50–750 samples to train the model while utilizing 250 samples to test the model. Additionally, we have tested the transfer learning approach to evaluate the model for prediction of the subsequent growing season (2019–2020) with 250 and 750 samples from 2018 to 2019 growing season for training and 250 samples from 2019 to 2020 growing season for validation. Furthermore, we aggregated data from the two growing seasons to train the model, validating with 2019–2020 growing season (Fig. 2 IV). More details about sample size and transfer learning are presented in Table 1.

2.6.3. Early prediction

In-season crop classification was evaluated by modifying the date for Sentinel-2 and Sentinel-1 (Fig. 2 V). The complete model has been set utilizing dates ranging from July 1 to May 1. For early predictions, we have trained and tested the model for the period ranging from July 1 until i) April 1 (herein termed as 'April'), ii) March 1 (termed as 'March'), iii) February 1 (termed as 'February'), and iv) January 1 (termed as 'January').

2.6.4. Crop map and statistics

Before producing the final crop map, we filtered a majority vote of size 5×5 pixels to reduce the salt and pepper effect. Pixels outside the mask were left unchanged, and using a modal filter results in an

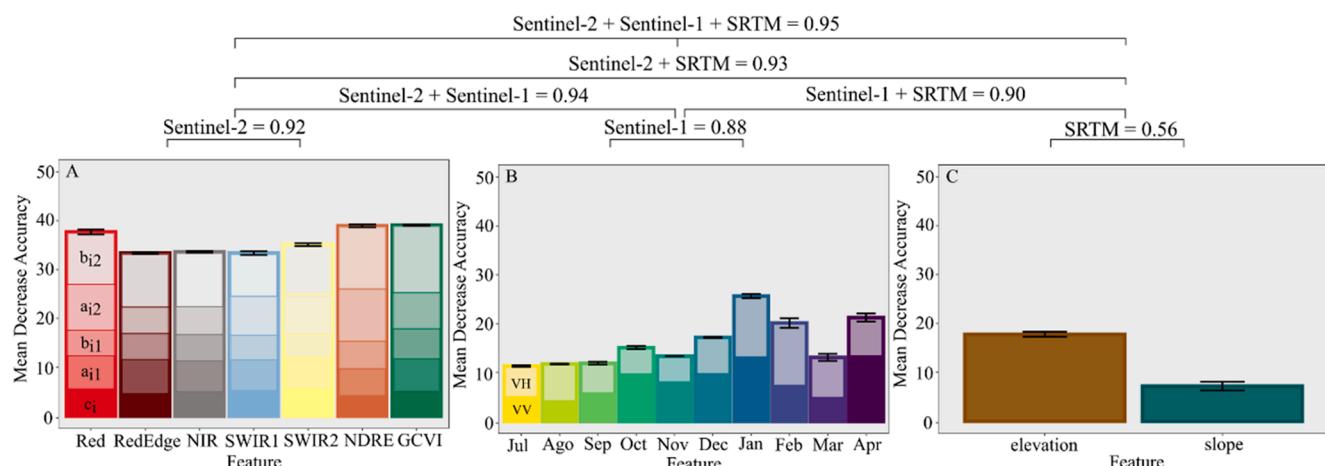


Fig. 3. Mean Decrease Accuracy of features from Sentinel-2 (A), Sentinel-1 (B), and SRTM Digital Elevation (C). c_i, a_{i1}, b_{i1}, a_{i2}, b_{i2} are the harmonic coefficients extracted from the Sentinel-2 time-series. VV and VH are the monthly mean features from Sentinel-1. Elevation and slope from SRTM Digital Elevation. Random forests overall accuracy of the feature's sources (Sentinel-2, Sentinel-1 and SRTM Digital Elevation) and data fusion were presented in the upper with the horizontal lines.

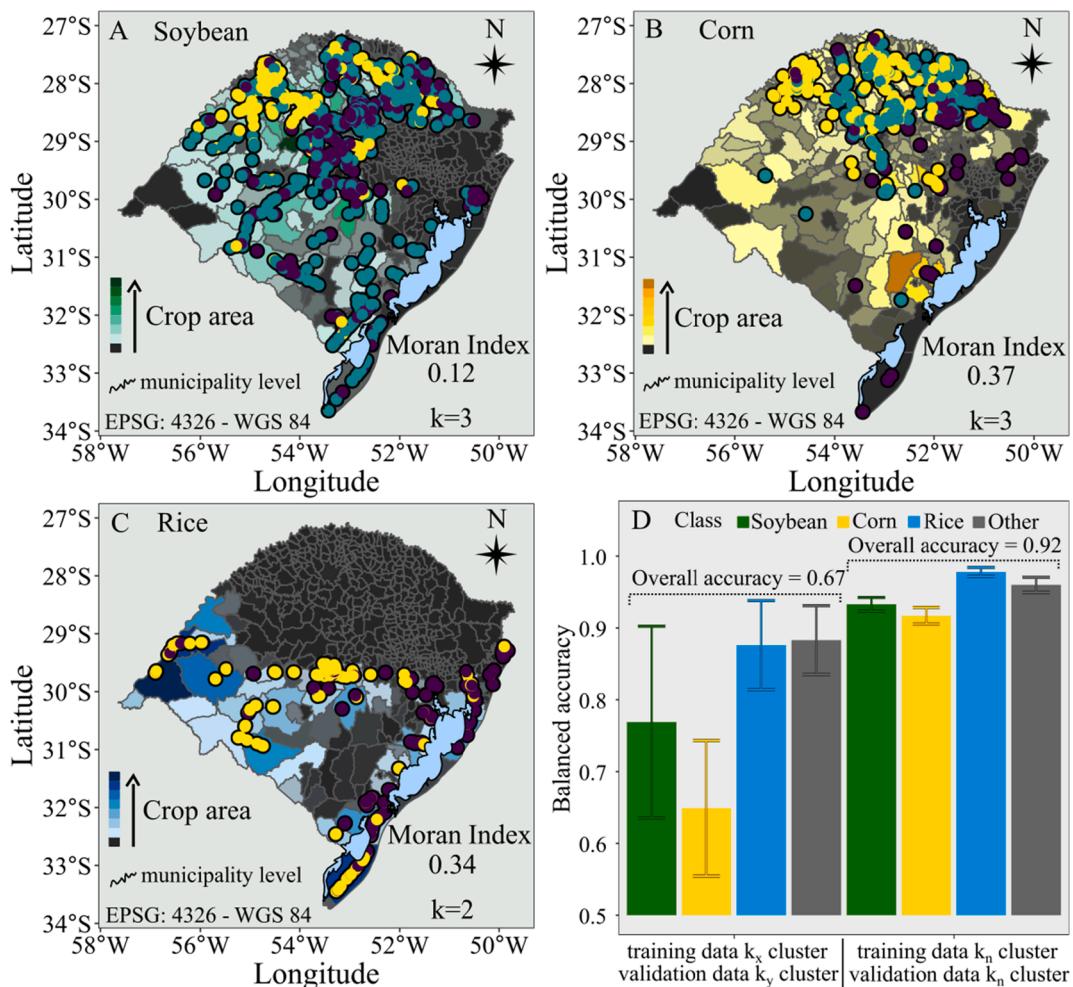


Fig. 4. (A) Spatial variabilities of the features of the model (Moran's Index and Cluster k-means analyses for soybean; (B) corn; (C) rice. Circles with different colors in the maps refer to the clusters of the features from Sentinel-2, Sentinel-1, and SRTM data. Background colors in the maps refer to cultivated area for each crop, mean from 2000–2001 to 2019–2020 growing seasons, IBGE statistics. (D) Balanced accuracy for each class regarding different clusters sampling data, model utilizing k_x training data and k_y validation data, model utilizing k_n training data and k_n validation data. Both training and validation with 100 samples for each cluster.

improvement of overall accuracy (Booth and Oldfield, 1989; Waldner et al., 2015; Belgiu and Csillik, 2018).

Rio Grande do Sul crop maps for 2017–2018, 2018–2019, and 2019–2020 growing seasons were created, and the area composed by each crop was extracted at a municipality-level. To evaluate the area prediction performance, we have compared the predicted crop area for 497 municipalities from IBGE estimates (IBGE, 2020) for each season (Fig. 2 VI).

The following statistical parameters were used to evaluate the prediction crop area model: coefficient of determination (R^2), mean absolute percentage error (MAPE), and mean absolute error (MAE). The R^2 examines how well differences in one variable can be explained by the difference in a second variable. MAPE measures the accuracy as a percentage, and can be calculated as the average absolute percent error minus actual values, divided by actual values. The mean absolute error (MAE) represents the average magnitude of the errors. As MAPE has the significant disadvantage that it produces undefined values for zero or close-to-zero actual values, we considered municipalities with more than 0.02% of Rio Grande do Sul crop area in the census to avoid affecting the statistic evaluation.

3. Results

3.1. Feature selection

Based on the combination of Spearman correlation and MDA of the variables, we have selected and discarded features when comparing the model performance (Figs. A1 and A2). The discarded features included the blue and green bands, EVI2 and NDVI VIs; mean monthly composition of the ratio (VV/VH) from July to April; and aspect. The selected features were the red, NIR, SWIR1, and SWIR2 bands, NDRE and GCVI VIs; mean monthly composition of VV and VH from July to April; elevation, and slope. The random forest model with all features obtained an OOB error of 6.61%, and overall accuracy of 0.93, while the model for selected features presented an OOB error of 5.63% and overall accuracy of 0.95.

The models utilizing only selected features from Sentinel-2, Sentinel-1, and SRTM obtained overall accuracy of 0.92, 0.88, and 0.56, respectively. The models utilizing two data sources with Sentinel-2 + Sentinel-1, Sentinel-2 + SRTM, and Sentinel-1 + SRTM presented overall accuracy of 0.94, 0.93, 0.90, respectively; while the data fusion merging the three data sources achieved an overall accuracy of 0.95 (Fig. 3).

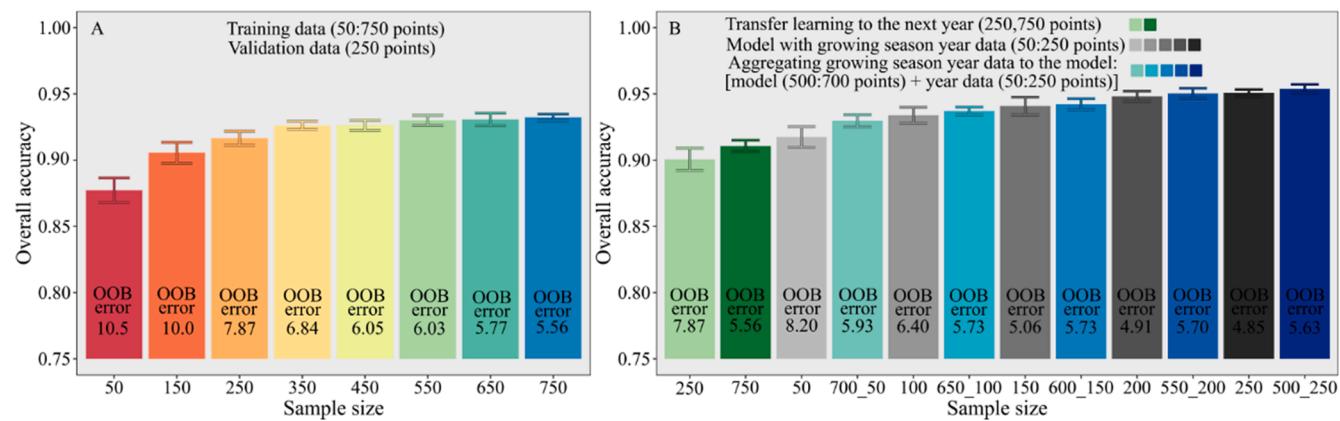


Fig. 5. (A) Comparison of overall accuracy of sample sizes for training the model (50:250 data samples). (B) Comparison of overall accuracy for: subsequent year growing season transfer learning (250 and 750 data samples of the 2018–2019 growing season model - greens), model with same-year growing season data (50:250 data samples - grays), and aggregating same-year growing season data to the model, training with 750 data samples, [model (500:700 data samples) + same-year data (50:250) - blues]. All the models evaluated were validated with 250 samples.

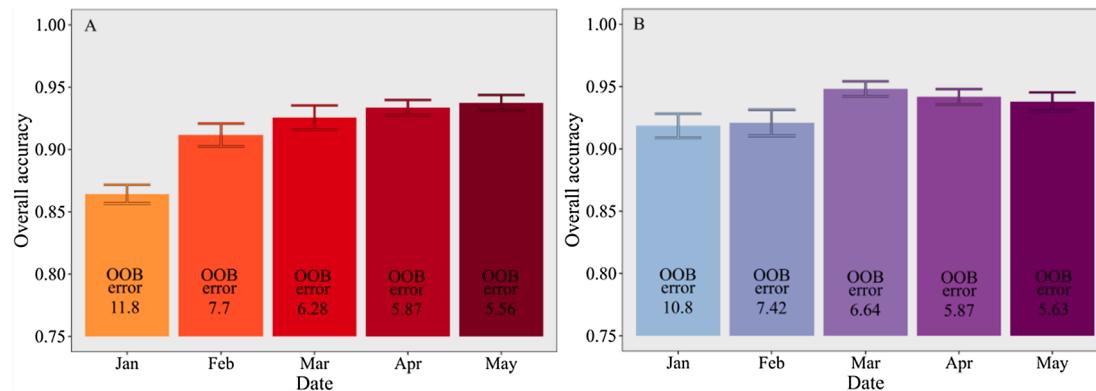


Fig. 6. (A) Comparison of overall accuracy for early crop classification to 2018–2019 growing season (2018–2019 data samples); and (B) 2019–2020 growing season (aggregating same-year data to the 2018–2019 model).

3.2. Spatial variability of crops features

The collected data points k-means cluster reflected the spatial variability of each cultivated crop area (Fig. 4 A, B, C). The selected features from Sentinel-2, Sentinel-1, and SRTM digital elevation displayed spatial dependence, mainly for corn and rice, with minimal spatial dependence for soybeans, with Moran's I index equal to 0.37, 0.34, and 0.12, respectively.

The cluster k-means analysis had $k = 3$ for soybeans and corn, but $k = 2$ for rice (Fig. 4 A, B, C). The clusters denoted the variability of the features from satellite data, representing an indirect spatial variability of soils, climate, and crop management.

A sensitivity analysis of the cluster selection, exploring all possible combinations, reported an overall decrease in accuracy from 0.92 to 0.67 due to the lack of consideration of the spatial variability in the ground truth data from Rio Grande do Sul state. The most affected crop in the classification was corn followed by soybean, with balanced accuracy equal to 0.64 and 0.74, respectively. Rice crop and other classes did not present a significant reduction in accuracy (Fig. 4D).

3.3. Sample size and transfer learning

Increases in sample size to train the model improved OOB error from 10.5 to 5.56%, and overall accuracy from 0.87 to 0.94, for 50 samples and 750 samples, respectively. Training the model with 750 samples achieved an accuracy of 0.94, no different from the model with 250

samples for the 2018–2019 growing season (Fig. 5A).

In addition, when the model was transferred to the next growing season with 250 and 750 data samples, respectively, the greater number of samples increased overall accuracy from 0.89 to 0.91. The model with same-year growing season data, even with low samples (50 data samples), presented greater overall accuracy relative to transfer year learning, with a change of 0.92 to 0.94 as the sample year data increased. The OOB error decreased as the sample size increased in the model with same-year growing season data (Fig. 5B).

Although the transfer learning with single year data obtained lower accuracy relative to fewer same-year samples, when more years of data were aggregated to the model with greater variation in climate, soil, and crop management, the overall accuracy increased. The model accuracy aggregating all data for the growing season to the model ranged from 0.93 (with 50 samples for 2019–2020, + 700 samples for 2018–2019 season) to 0.95 (with 250 samples for 2019–2020, + 500 samples for 2018–2019 season) (Fig. 5B).

3.4. In-season prediction

Relative to the entire season, early season predictions in crop classification with single growing season data decreased overall accuracy from 0.94 to 0.86, and increased OOB error from 5.56 to 11.8% for May and January, respectively (Fig. 6A). However, predictions with data through March obtained an overall accuracy similar to the May predictions (Fig. 6). Moreover, when the model was built with two growing

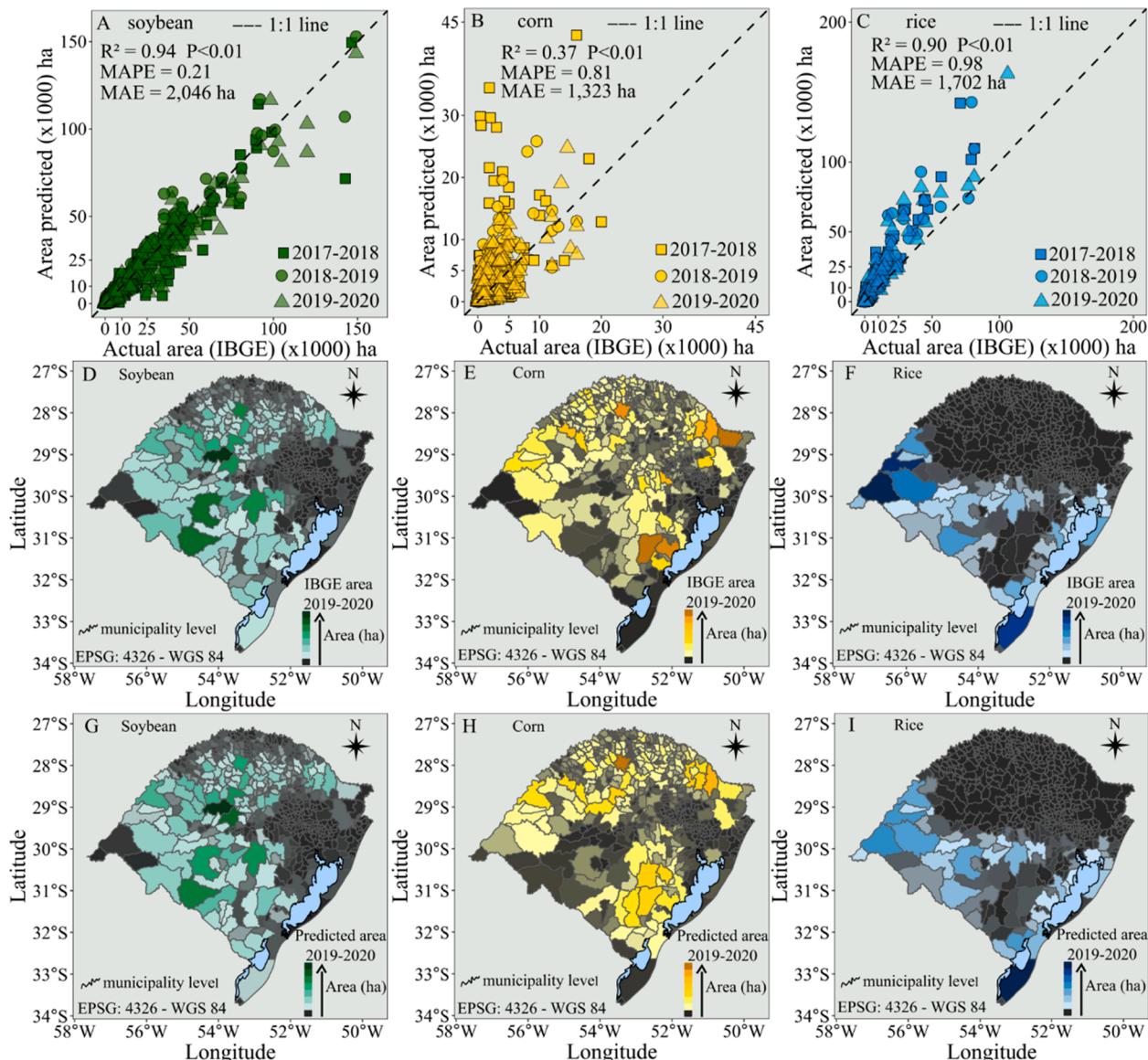


Fig. 7. (A) Relation 1:1 area predicted by the model and actual area IBGE statistics for soybean, (B) corn, and (C) rice. Data from 2017–2018, 2018–2019, 2019–2020 growing seasons. (D) Crop municipality area map of IBGE statistics from 2019 to 2020 growing season for soybean, (E) corn, and (F) rice. (G) Crop municipality area map predicted from 2019 to 2020 growing season for soybean, (H) corn, and (I) rice.

seasons, 2018–2019 and 2019–2020, March predictions presented the highest overall accuracy relative to the early and late season predictions (Fig. 6B). Balanced accuracy of the classes with single year data had a small penalty for soybean and corn until March, with ~0.95 balanced accuracy from April onwards. However, with two growing seasons of data, corn was able to obtain a high balanced accuracy as early as January, and a balanced accuracy >0.95 was reached for soybean from March onwards (Table A1).

3.5. Crop map product and statistics

The statistics of the crop classification model reveal a high performance of soybean crop area predictions ($R^2 = 0.94$ and MAPE = 0.21) related to IBGE area estimation. The soybean area MAE for the three growing seasons was ~2,046 ha at the municipality level (Fig. 7A). For corn, the model performance achieved was lower relative to soybeans ($R^2 = 0.37$ and MAPE = 0.81) with a MAE for the three growing seasons of ~1,323 ha (Fig. 7B). While the rice prediction presented high performance ($R^2 = 0.90$), with an over estimation due to the confounding

effect of other class (targets of wetland with rice targets throughout the state), with MAPE = 0.98 and MAE of ~1,702 ha (Fig. 7C).

The crop classification model at municipality level presented a normal distribution error, with no specific spatial associated error in the crop area (Fig. 7D, E, F, G, H, and I).

Lastly, a final crop map layer for 2019–2020 season (Fig. 8A), and one example of a municipality with two crop classes (Fig. 8B), and a municipality detailed with three crop classes (Fig. 8C) were mapped. In regions mainly cultivated with two crops (soybean and corn), such as in Northern Rio Grande do Sul, the crop classification map layer produced more defined field boundaries in the model with less pixel noises compared to municipalities with three crop classes like Southern Rio Grande do Sul, where soybean, corn, and rice are the major three crops during the summer. In Table A2, examples are presented of crop area prediction based on satellite-based data fusion and IBGE census (standard method) for main crop cultivating municipalities in Rio Grande do Sul, Brazil.

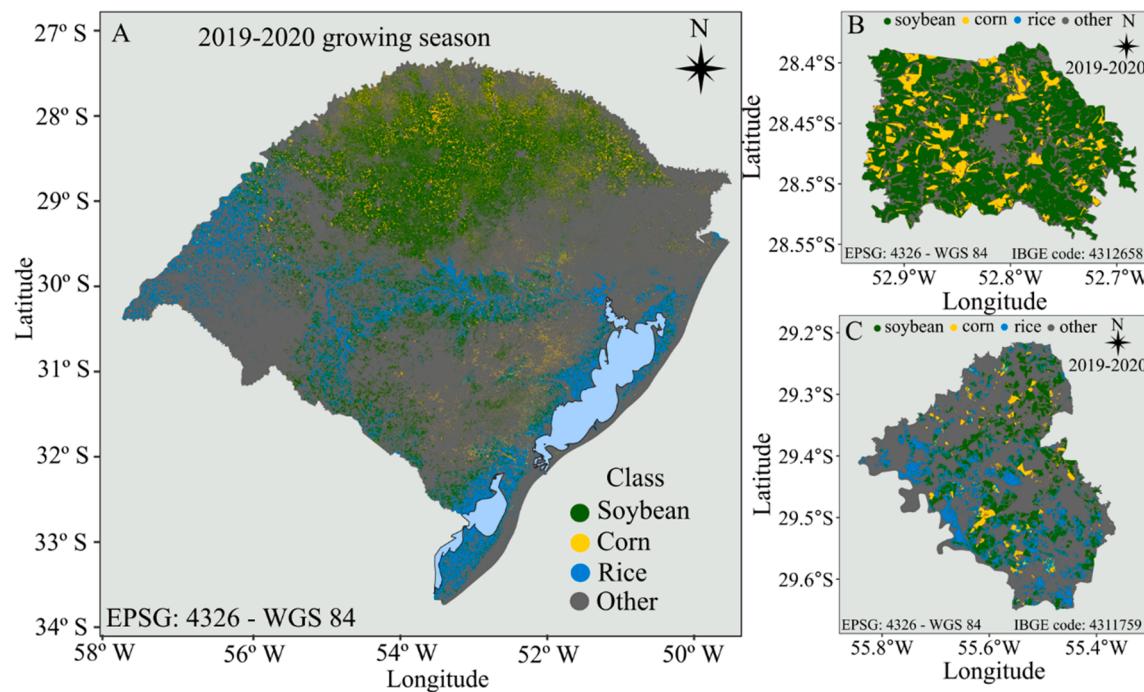


Fig. 8. (A) Rio Grande do Sul state, Brazil, crop map of 2019–2020 growing season. (B) Municipality detailed with two crop classes. (C) Municipality detailed with three crop classes.

4. Discussion

The data fusion merging Sentinel-2, Sentinel-1, and Digital Elevation (SRTM) was relevant for developing this complex regional classification with an overall accuracy of 0.95. Similar to past studies, the combination of both optical and SAR data have improved crop classification (Abubakar et al., 2020; Inglada et al., 2016; Jin et al., 2019; Orynbai-kyzy et al., 2020), with overall accuracy improving when SRTM data was included in the current study.

For crop classification in Brazil, Rudorff et al. (2007) presented a kappa of ~ 0.50 for soybean crop land in the north region of Rio Grande do Sul state with MODIS. In the state of Mato Grosso (Brazil), a study of crop classification presented an overall accuracy of 0.74 with EVI using MODIS (Arvor et al., 2011). Lemos (2016) developed a soybean mapping classification utilizing EVI from MODIS in Southern Rio Grande do Sul with accuracy ~ 0.95, while Mengue and Fontana (2015) utilized MODIS and SRTM data for mapping soybean and rice crops in Rio Grande do Sul obtaining a kappa of 0.61 and 0.66, respectively for each crop. Zhong et al. (2016) obtained 87% overall accuracy utilizing MODIS for mapping soybean and corn cropland in Paraná (Brazil). Schultz et al. (2015) achieved 80% overall accuracy for classification of soybean, sugarcane (*Saccharum officinarum* L.), cassava (*Manihot esculenta* Crantz), peanut (*Arachis hypogaea* L.), and other crops. Furthermore, Silva Junior et al. (2020) used a new index developed by Silva Junior et al. (2017) to map planted soybean area in the state of Mato Grosso comparing MODIS, Landsat 8, and Sentinel 2 sensors with greater accuracy for MODIS ($R^2 = 0.90$).

One of the main problems associated with optical sensors in satellite platforms such as Sentinel-2 is the issue of missing pixels due to cloud, cirrus, and shadow. For this reason, sensors having more temporal with high to medium spatial resolutions have the advantage to reduce time series data losses (Jin et al., 2019; Silva Junior et al., 2020). To handle missing pixels, feature engineering techniques such as harmonic regression coefficient extraction have been broadly used for crop classification (Ghazaryan et al., 2018; Wang et al. 2019, 2020; Deines et al., 2020; Dado et al., 2020). Multitemporal Sentinel-1 data have been used as a powerful tool for crop classification and agricultural monitoring in

several studies (Skriver et al., 2011; Bargiel 2017; Bazzi et al., 2019). Furthermore, SRTM Digital Elevation data play an important role in land cover classification (Balzter et al., 2015; Sadeghi et al., 2018). The state of Rio Grande do Sul has a wide variation in terrain features such as elevation and slope, and the input of these features in the model improved crop classification performance in our study.

The spatial distribution of ground truth crop data exerted a strong impact on model performance, as reflected by the Moran's Index and k-mean cluster analyses. The quality of the data collected for the supervised classification model also influences model performance (Campbell and Wynne, 2011; Fowler et al., 2020). A representative subset of the population for each crop class with varying environmental and management conditions is important in the data collection process (Waldner et al., 2017; Waldner et al., 2019). The Moran's index and cluster k-mean provided more efficient guidance for crop type collection. Surveying data points/polygons in all the cluster regions of the Rio Grande do Sul state substantially improved the crop classification model.

Transfer learning for the subsequent growing season presented high accuracy when the model was built with more data input (large sample size). Random forests trained on regions and seasons with similar growing degree days (GDD) presented accuracies greater than 80% when transferred to target region and season with a comparable crop season (Wang et al., 2019). The similarity of crop phenology patterns, along with location and years lead to higher performance of transfer learning in random forest classification models (Hao et al., 2016; Wang et al., 2019; Hao et al., 2020a). Although transfer model learning showed good potential in crop classification for the subsequent growing season, when some samples data from same-year growing season were aggregated to the model, the model performance presented higher accuracy. A random forest classification utilizing local training samples increased overall accuracy compared to transfer learning for another region in United States (US) (Hao et al., 2020a). Improving model performance of transfer learning by merging data from a variety of growing conditions enlarges the potential that the training samples represent the region or growing season.

In-season crop classification is valuable information for decision-

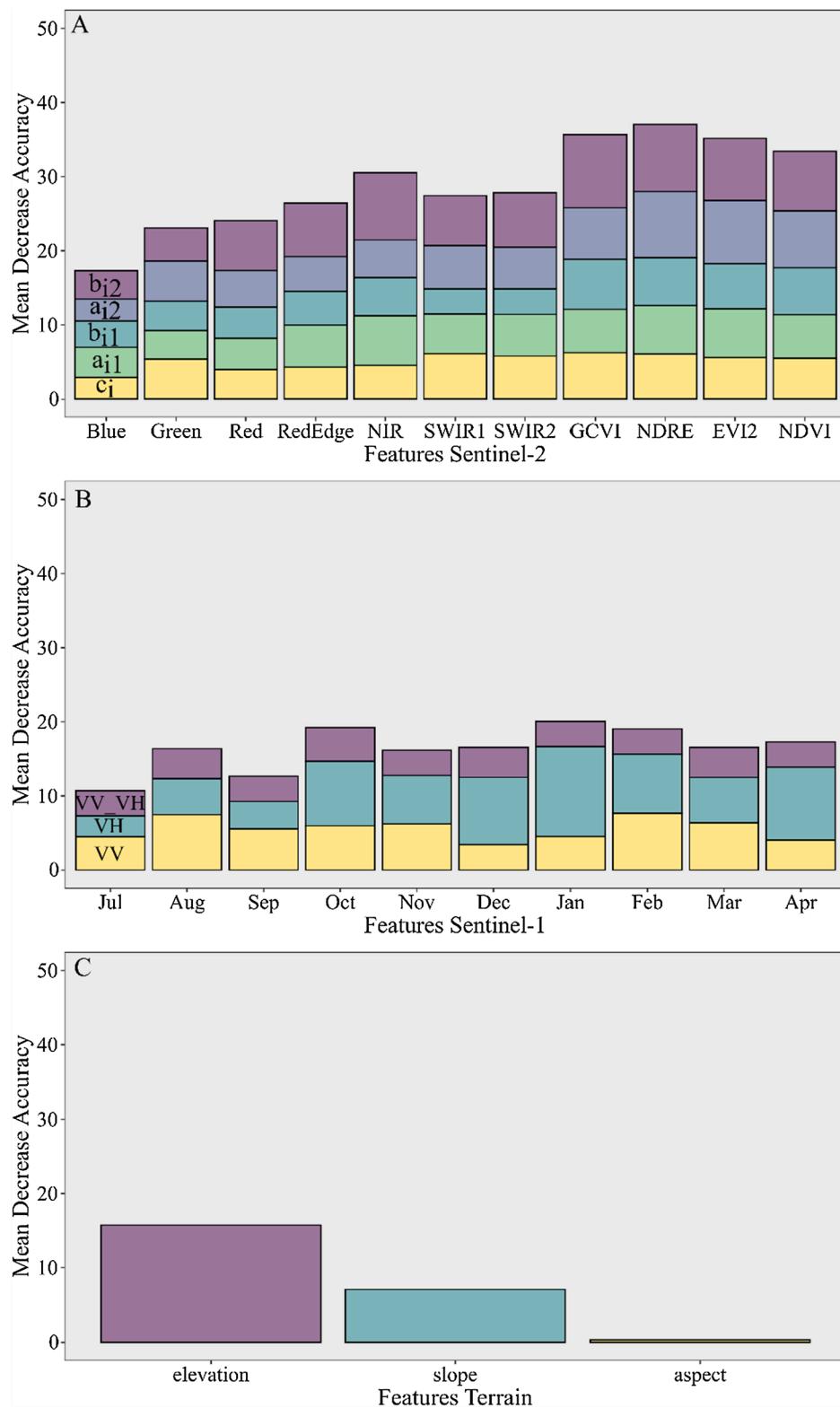


Fig. A1. Mean Decrease Accuracy of all the features from Sentinel-2 (A), Sentinel-1 (B), and SRTM Digital Elevation (C). c_i , a_{i1} , b_{i1} , a_{i2} , b_{i2} are the harmonic coefficients extracted from the Sentinel-2 time-series. VV, VH, and VV_VH are the monthly mean features from Sentinel-1. Elevation, slope, and aspect from SRTM Digital Elevation.

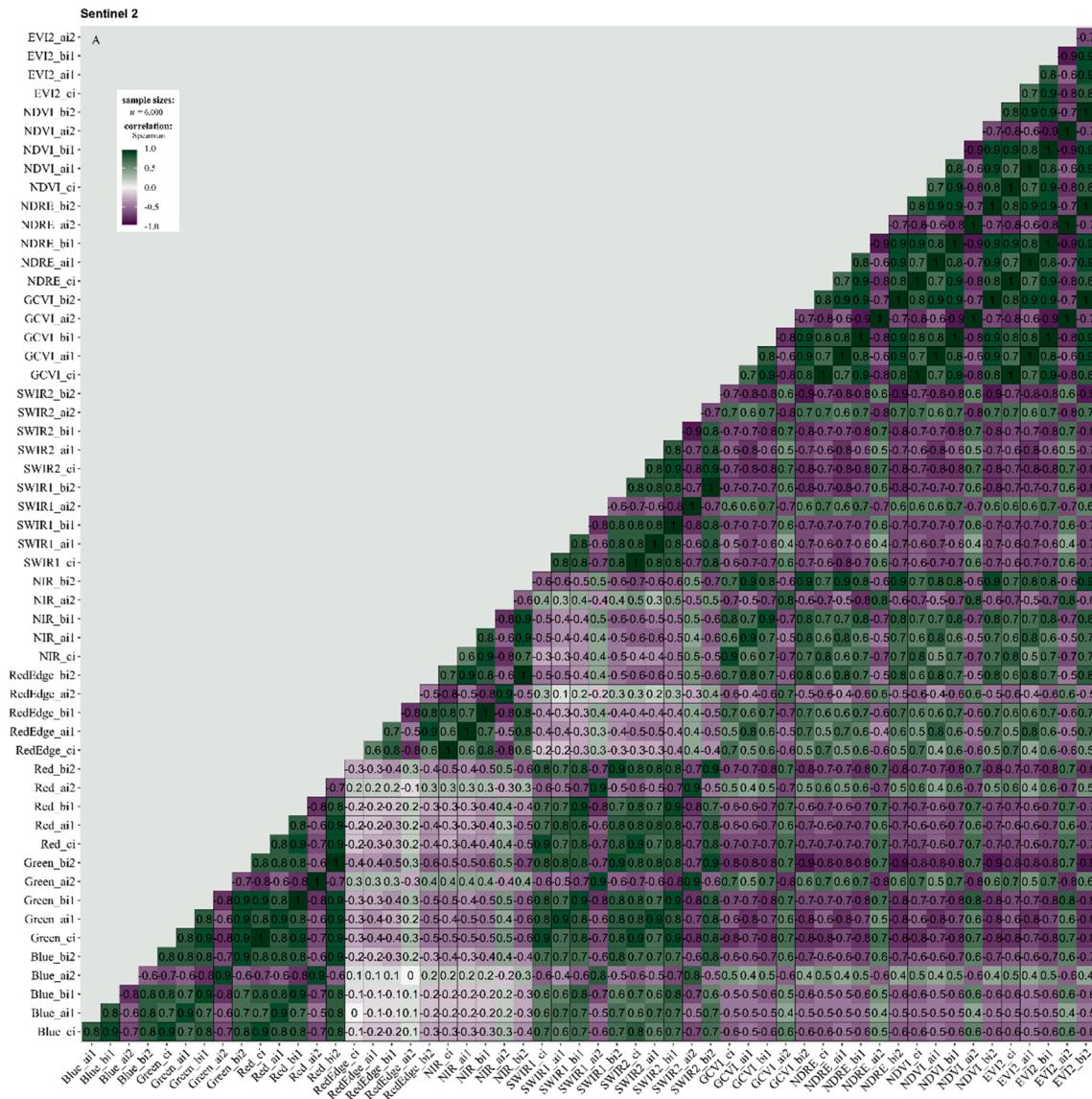


Fig. A2. Spearman correlation for all the features from Sentinel-2 (A), Sentinel-1 (B), and SRTM Digital Elevation (C).

makers. We obtained similar prediction accuracy for early March (DOY 60) relative to the entire time-series (with all season data until May), with the potential use of January and February predictions when aggregating same-season data to the model. For the present study, aggregating data across years increased accuracy for in-season crop classification. Likewise, Hao et al. (2020a) reported an increase in early season predictions when local observations were utilized. Cai et al. (2018) reported high accuracy for late-June (DOY 180) in-season crop classification for both soybean and corn in the US, which represents approximately late December in the Rio Grande do Sul growing season.

Future studies should focus on high-resolution satellite sources, integrating crop modeling, and including more soils and climate features with detected agricultural field boundaries in order to improve model performance across regions and growing seasons. Moreover, further study should focus on crop classification including other minor crops, additional land features, and specific crop rotation patterns.

5. Conclusion

Satellite-based crop classification and mapping for soybean, corn, and rice crops was developed for the state of Rio Grande do Sul, Brazil.

Features derived from Sentinel-2, Sentinel-1, and SRTM Digital Elevation provided high accuracy for crop classification.

From the ground truthing perspective, sample size and spatial distribution are key aspects, with more than 250 samples resulting in high overall accuracy. Random forest presented high accuracy in transferring the model to the next season, with the performance of the model increasing when full-season data was included.

Predictions of crop classification for March obtained similar results to when the entire season was taken into account (until May). Moreover, when data from additional growing seasons were aggregated, earlier predictions (January and February) reached higher overall accuracy.

Lastly, crop area prediction was most accurate for soybeans and least accurate for corn, with rice area overestimated due to errors with wetland target for other class.

In summary, to the extent of our knowledge, this is the first crop map layer for soybean, corn and rice crops for Rio Grande do Sul state (Brazil), establishing a foundation for creating maps for other crops, seasons, and regions around the globe.

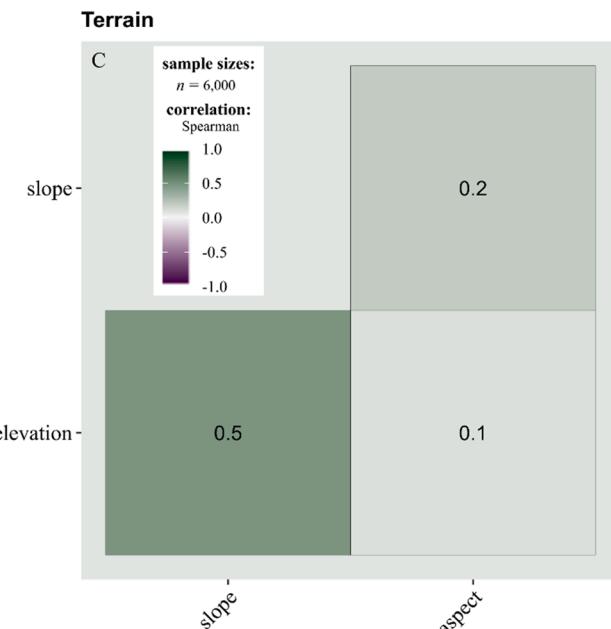
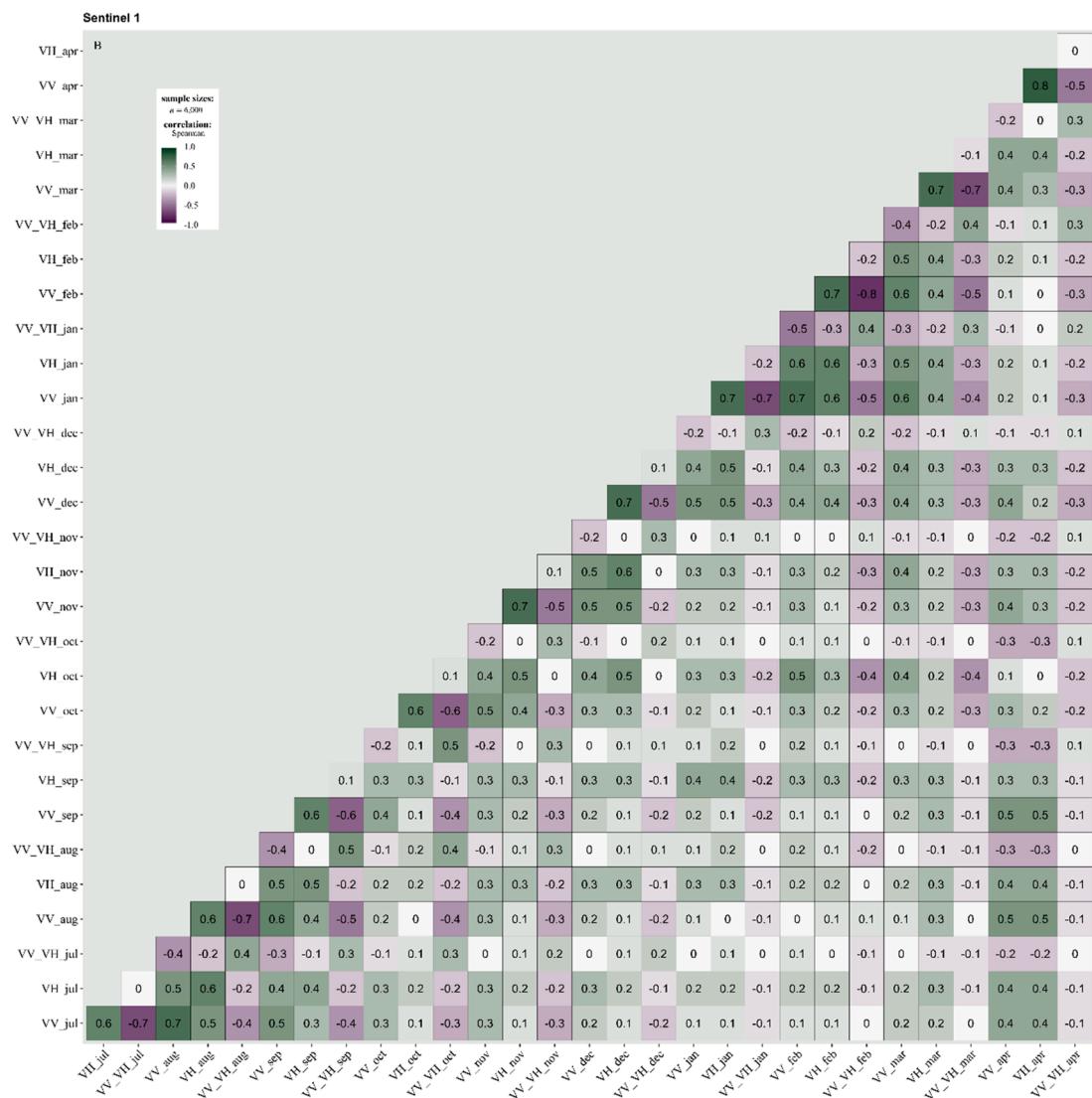


Fig. A2. (continued).

Table A1

Balanced accuracy of the classes in early predictions with one (2018–2019 data samples) and two (2018–2019 + 2019–2020 data samples) growing seasons.

Class	Balanced accuracy									
	2018–2019 data samples					2018–2019 + 2019–2020 data samples				
	Jan	Feb	Mar	Apr	May	Jan	Feb	Mar	Apr	May
Corn	0.91	0.92	0.91	0.95	0.95	0.97	0.97	0.96	0.96	0.96
Soybean	0.90	0.94	0.94	0.95	0.96	0.94	0.93	0.96	0.96	0.96
Rice	0.97	0.98	0.99	0.97	0.98	0.99	0.98	0.99	0.98	0.98
Other	0.91	0.93	0.92	0.96	0.97	0.97	0.97	0.98	0.98	0.97

Table A2

Municipality crop area of select cities in Rio Grande do Sul. Census from IBGE and crop area predicted by model.

City	IBGE			Predicted		
	Corn	Rice	Soybean	Corn	Rice	Soybean
Agudo	3500	9100	950	825	10,112	1642
Alegrete	3800	49,641	38,000	3700	77,923	38,985
Cachoeira do Sul	3500	25,522	103,000	6489	61,037	92,775
Carazinho	2300	0	40,420	6564	1	42,006
Cruz Alta	5100	0	92,000	12,469	2	90,716
Dom Pedrito	550	38,923	120,000	882	56,685	102,975
Erechim	1320	0	9700	6189	1	13,345
Ijuí	1900	0	40,000	2114	2	42,122
Jóia	3700	0	81,000	5575	17	71,941
Lagoa Vermelha	4500	0	46,000	13,039	4	53,331
Não-Me-Toque	3000	0	22,500	4220	0	22,843
Palmares do Sul	150	12,845	4090	639	19,728	4566
Palmeira das Missões	14,500	0	90,400	24,767	1	81,077
Passo Fundo	1400	0	41,000	3924	2	43,070
Pelotas	7000	7177	22,000	6658	22,356	15,165
Santa Bárbara do Sul	3200	0	76,000	8345	0	68,648
Santa Maria	190	6250	49,345	231	17,874	41,701
Santa Rosa	2500	0	19,500	5705	1	19,200
Santo Ângelo	3400	0	36,300	4685	7	33,155
São Borja	5000	38,800	65,000	9571	72,483	68,469
Soledade	800	0	42,000	2206	3	52,754
Tupanciretã	2530	0	149,100	9469	18	143,320
Uruguaiana	0	76,319	2980	2443	89,168	2564
Vacaria	13,500	0	55,000	24,767	1	81,077

CRediT authorship contribution statement

Luan Pierre Pott: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Validation, Writing - original draft. **Telmo Jorge Carneiro Amado:** Validation, Supervision, Writing - review & editing. **Raí Augusto Schwalbert:** Investigation, Software, Formal analysis, Writing - review & editing. **Geomar Mateus Corassa:** Validation, Writing - review & editing. **Ignacio Antonio Ciampitti:** Conceptualization, Investigation, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank A.K.B., C.R., E.M.G., E.R., G.H., and R.D. for helping the data collection process. The first author thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The second author thanks to CNPq for research scholarship n 305622/2017-0. Contribution no. 21-244-J from the Kansas Agricultural Experiment Station.

Appendix A

See Figs. A1 and A2.

See Tables A1 and A2.

References

- Abubakar, G.A., Wang, K., Shahtahanssebi, A., Xue, X., Belete, M., Gudo, A.J.A., Gan, M., 2020. Mapping maize fields by using multi-temporal Sentinel-1A and Sentinel-2A images in Makarfi, Northern Nigeria. *Africa. Sustainability* 12 (6), 2539. <https://doi.org/10.3390/su12062539>.
- Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. *Meteorol. Z.* 22 (6), 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>.
- Andrade, B.O., Bonilha, C.L., Overbeck, G.E., Vélez-Martin, E., Rolim, R.G., Bordignon, S. A.L., Boldrini, I.I., 2019. Classification of South Brazilian grasslands: Implications for conservation. *Appl. Veg. Sci.* 22 (1), 168–184. <https://doi.org/10.1111/avsc.12413>.
- Arvor, D., Jonathan, M., Meirelles, M.S.P., Dubreuil, V., Durieux, L., 2011. Classification of MODIS EVI temporal series for crop mapping in the state of Mato Grosso, Brazil. *Int. J. Remote Sens.* 32 (22), 7847–7871. <https://doi.org/10.1080/01431161.2010.531783>.
- Balzter, H., Cole, B., Thiel, C., Schmullius, C., 2015. Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM Digital Elevation Model Data using Random Forests. *Remote Sens.* 7 (11), 14876–14898. <https://doi.org/10.3390/rs71114876>.
- Bargiel, D., 2017. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sens. Environ.* 198, 369–383. <https://doi.org/10.1016/j.rse.2017.06.022>.
- Barnes, E.M., Clarke, T.R., Richards, S.E., Colaizzi, P.D., Haberland, J., Kostrzewski, M., Moran, M.S., 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground based multispectral data. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), 'Proceedings of the Fifth International Conference on Precision Agriculture'. American Society of Agronomy: Madison, WI (CD-ROM).

- Bazzi, H., Baghdadi, N., Ienco, D., El Hajj, M., Zribi, M., Belhouchette, H., Demarez, V., 2019. Mapping irrigated areas using Sentinel-1 time series in Catalonia, Spain. *Remote Sens.* 11 (15), 1836. <https://doi.org/10.3390/rs11151836>.
- Belgiu, M., Csillik, O., 2018. 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>.
- Bisong, E., 2019. Google Colaboratory. Build. Mach. Learn. Deep Learn. Models Google Cloud Platform 59–64. https://doi.org/10.1007/978-1-4842-4470-8_7.
- Booth, D.J., Oldfield, R.B., 1989. A comparison of classification algorithms in terms of speed and accuracy after the application of a post-classification modal filter. *Int. J. Remote Sens.* 10, 1271–1276. <https://doi.org/10.1080/01431168908903965>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/a:1010933404324>.
- Buchhorn, Marcel, Lesiv, Myroslava, Tsendsazar, Nandin-Erdene, Herold, Martin, Bertels, Luc, Smets, Bruno, 2020. Copernicus global land cover layers—collection 2. *Remote Sens.* 12 (6) <https://doi.org/10.3390/rs12061044>.
- Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. 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. <https://doi.org/10.1016/j.rse.2018.02.045>.
- Campbell, J.B., Wynne, R.H., 2011. *Introduction to Remote Sensing*. Guilford Press.
- Carletto, C., Jolliffe, D., Banerjee, R., 2015a. From tragedy to renaissance: improving agricultural data for better policies. *J. Develop. Stud.* 51, 133–148. <https://doi.org/10.1080/00220388.2014.968140>.
- Carletto, C., Gourlay, S., Winters, P., 2015b. From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. *J. African Econ.* 24 (5), 593–628. <https://doi.org/10.1093/jae/evj011>.
- CONAB, Companhia Nacional de Abastecimento (2020a). - Mapeamentos Agrícolas. Retrieved November 19, 2020, from Conab.gov.br website: <https://www.conab.gov.br/info-agro/safras/mapeamentos-agricolas>.
- CONAB, Companhia Nacional de Abastecimento. (2020b). Acompanhamento da safra brasileira de grãos, v. 7 - Safra 2019/20 - Décimo segundo levantamento, Brasília, p. 1-68, setembro 2020. Retrieved November 4, 2020, from Conab.gov.br website: <http://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>.
- CONAB, Companhia Nacional de Abastecimento. (2020b). Boletim da Safra de Grãos. Estimativas - Produção e Balanço de Oferta e Demanda. Quarto Levantamento - Safra 2020/21. Retrieved January 15, 2021, from Conab.gov.br website: <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>.
- Dado, W.T., Deines, J.M., Patel, R., Liang, S.-Z., Lobell, D.B., 2020. High-resolution Soybean yield mapping across the US midwest using subfield harvester data. *Remote Sens.* 12 (21), 3471. <https://doi.org/10.3390/rs12213471>.
- Deines, J.M., Patel, R., Liang, S.-Z., Dado, W., Lobell, D.B., 2020. A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sens. Environ.* 112174 <https://doi.org/10.1016/j.rse.2020.112174>.
- Demarez, V., Helen, F., Marais-Sicre, C., Baup, F., 2019. In-Season Mapping of Irrigated Crops Using Landsat 8 and Sentinel-1 Time Series. *Remote Sens.* 11 (2), 118. <https://doi.org/10.3390/rs11020118>.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. *Rev. Geophys.* 45, RG2004. <https://doi.org/10.1029/2005RG000183>.
- Fowler, J., Waldner, F., Hochman, Z., 2020. All pixels are useful, but some are more useful: Efficient in situ data collection for crop-type mapping using sequential exploration methods. *Int. J. Appl. Earth Obs. Geoinf.* 91, 102114 <https://doi.org/10.1016/j.jag.2020.102114>.
- Ghazaryan, G., Dubovik, O., Löw, F., Lavreniuk, M., Kolotii, A., Schellberg, J., Kussul, N., 2018. A rule-based approach for crop identification using multi-temporal and multi-sensor phenological metrics. *Eur. J. Remote Sens.* 51 (1), 511–524. <https://doi.org/10.1080/22797254.2018.1455540>.
- Gitelson, A.A., Vina, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* 32 (8) <https://doi.org/10.1029/2005gl022688>.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>.
- Gourlay, S., Kilic, T., Lobell, D., 2017. Could the Debate be Over? Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda. The World Bank.
- Hao, P., Di, L., Zhang, C., Guo, L., 2020a. Transfer Learning for Crop classification with Cropland Data Layered data (CDL) as training samples. *Sci. Total Environ.* 733, 138869 <https://doi.org/10.1016/j.scitotenv.2020.138869>.
- Hao, P., Tang, H., Chen, Z., Meng, Q., Kang, Y., 2020b. Early-season crop type mapping using 30-m reference time series. *J. Integrative Agric.* 19 (7), 1897–1911. [https://doi.org/10.1016/s2095-3119\(19\)62812-1](https://doi.org/10.1016/s2095-3119(19)62812-1).
- Hao, P., Wang, L., Zhan, Y., Niu, Z., 2016. Using moderate-resolution temporal NDVI profiles for high-resolution crop mapping in years of absent ground reference data: A Case Study of Bole and Manas Counties in Xinjiang, China. *ISPRS Int. J. Geo-Inform.* 5 (5), 67. <https://doi.org/10.3390/ijgi5050067>.
- Hollstein, A., Segl, K., Guanter, L., Brell, M., Enesco, M., 2016. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI Images. *Remote Sens.* 8 (8), 666. <https://doi.org/10.3390/rs8080666>.
- IBGE, Instituto Brasileiro de Geografia e Estatística. (2019). Retrieved November 4, 2020, from Ibge.gov.br website: <https://www.ibge.gov.br/cidades-e-estados/rs/>.
- IBGE, Instituto Brasileiro de Geografia e Estatística. Área plantada, área colhida, quantidade produzida, rendimento médio e valor da produção das lavouras temporárias, 2020. Retrieved October 5, 2020, from Ibge.gov.br website: <https://sidra.ibge.gov.br/tabela/1612#>.
- IBGE, Instituto Brasileiro de Geografia e Estatística. Série histórica da estimativa anual da área plantada, área colhida, produção e rendimento médio dos produtos das lavouras, 2020. Retrieved November 4, 2020, from Ibge.gov.br website: <https://sidra.ibge.gov.br/tabela/6588>.
- Ingla, J., Vincent, A., Arias, M., Marais-Sicre, C., 2016. Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. *Remote Sens.* 8 (5), 362. <https://doi.org/10.3390/rs8050362>.
- Jakubauskas, M.E., Legates, D.R., Kastens, J.H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Comput. Electron. Agric.* 37 (1–3), 127–139. [https://doi.org/10.1016/s0168-1699\(02\)00116-3](https://doi.org/10.1016/s0168-1699(02)00116-3).
- Jiang, Z., Huete, A., Didan, K., Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* 112 (10), 3833–3845. <https://doi.org/10.1016/j.rse.2008.06.006>.
- Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* 228, 115–128. <https://doi.org/10.1016/j.rse.2019.04.016>.
- Kayad, A., Sozzi, M., Gatto, S., Marinello, F., Pirotti, F., 2019. Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques. *Remote Sens.* 11 (23), 2873. <https://doi.org/10.3390/rs11232873>.
- Kuhn, M., 2008. Building predictive models in R using the caret package. *J. Stat. Softw.* 28 (5), 1–26. <https://doi.org/10.18637/jss.v028.i05>.
- Lemos, G.S., 2016. Mapeamento de áreas de soja em municípios da metade sul do estado do Rio Grande Do Sul a partir de imagens de satélite. (Mestrado em Ciências Programa de Pós-graduação em Manejo e Conservação do solo e da Água, Universidade Federal de Pelotas, Pelotas, pp. 82).
- Li, Q., Qiu, C., Ma, L., Schmitt, M., Zhu, X., 2020. Mapping the Land Cover of Africa at 10 m Resolution from Multi-Source Remote Sensing Data with Google Earth Engine. *Remote Sens.* 12 (4), 602. <https://doi.org/10.3390/rs12040602>.
- Mengue, V.P., Fontana, D.C., 2015. Avaliação da dinâmica espectro-temporal visando o mapeamento dos principais cultivos de verão no Rio Grande do Sul. *Bragantia* 74 (3), 331–340. <https://doi.org/10.1590/1678-4499.0452>.
- Orynbaikyzy, A., Gessner, U., Mack, B., Conrad, C., 2020. Crop type classification using fusion of Sentinel-1 and Sentinel-2 data: assessing the impact of feature selection, optical data availability, and parcel sizes on the accuracies. *Remote Sens.* 12 (17), 2779. <https://doi.org/10.3390/rs12172779>.
- Ozdogan, M., Gutman, G., 2008. A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: An application example in the continental US. *Remote Sens. Environ.* 112 (9), 3520–3537. <https://doi.org/10.1016/j.rse.2008.04.010>.
- R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ringland, J., Bohm, M., Baek, S.-R., 2019. Characterization of food cultivation along roadside transects with Google Street View imagery and deep learning. *Comput. Electron. Agric.* 158, 36–50. <https://doi.org/10.1016/j.compag.2019.01.014>.
- Rudorff, C. de M., Rizzi, R., Rudorff, B.F.T., Sugawara, L.M., Vieira, C.A.O., 2007. Superfícies de resposta espectro-temporal de imagens do sensor MODIS para classificação de área de soja no Estado do Rio Grande do Sul. *Ciência Rural* 37(1), 118–125. <https://doi.org/10.1590/s0103-84782007000100019>.
- Sadeghi, Y., St-Onge, B., Leblon, B., Prieur, J.-F., Simard, M., 2018. Mapping boreal forest biomass from a SRTM and TanDEM-X based on canopy height model and Landsat spectral indices. *Int. J. Appl. Earth Obs. Geoinf.* 68, 202–213. <https://doi.org/10.1016/j.jag.2017.12.004>.
- Santos, H.G., Jacomine, P.K.T., Dos Anjos, L.H.C., De Oliveira, V.A., Lumbreiras, J.F., Coelho, M.R., Cunha, T.J.F., 2018. Sistema brasileiro de classificação de solos. Brasília, DF: Embrapa, 2018.
- Schultz, B., Immitzer, M., Formaggio, A., Sanches, I., Luiz, A., Atzberger, C., 2015. Self-guided segmentation and classification of multi-temporal landsat 8 images for crop type mapping in Southeastern Brazil. *Remote Sens.* 7 (11), 14482–14508. <https://doi.org/10.3390/rs71114482>.
- Silva Junior, C.A., Leonel-Junior, A.H.S., Rossi, F.S., Correia Filho, W.L.F., Santiago, D. de B., Oliveira-Júnior, J.F. de Capristo-Silva, G.F., 2020. Mapping soybean planting area in midwest Brazil with remotely sensed images and phenology-based algorithm using the Google Earth Engine platform. *Comput. Electron. Agric.* 169, 105194. <https://doi.org/10.1016/j.compag.2019.105194>.
- Silva Junior, C.A., Nanni, M.R., Teodoro, P.E., Silva, G.F.C., 2017. Vegetation indices for discrimination of soybean areas: A new approach. *Agron. J.* 109 (4), 1331–1343. <https://doi.org/10.2134/agronj2017.01.0003>.
- Skriver, H., Mattia, F., Satalino, G., Balenzano, A., Pauwels, V.R.N., Verhoest, N.E.C., Davidson, M., 2011. Crop Classification Using Short-Revisit Multitemporal SAR Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 4 (2), 423–431. <https://doi.org/10.1109/jstars.2011.2106198>.
- Soil Survey Staff, 2014. Keys to soil taxonomy, 12th ed. USDA-Natural Resources Conservation Service, Washington, DC.
- Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., Mochizuki, K.-I., 2017. Assessing the suitability of data from Sentinel-1A and 2A for crop classification. *GIScience & Remote Sens.* 54, 918–938. <https://doi.org/10.1080/15481603.2017.1351149>.
- Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *J. Roy. Statist. Soc.: Series B (Statistical Methodology)* 63 (2), 411–423. <https://doi.org/10.1111/1467-9868.00293>.

- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8 (2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0).
- Van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance—A review. *Field Crops Res.* 143, 4–17. <https://doi.org/10.1016/j.fcr.2012.09.009>.
- Waldner, F., Bellemans, N., Hochman, Z., Newby, T., de Abelleira, D., Verón, S.R., Defourny, P., 2019. Roadside collection of training data for cropland mapping is viable when environmental and management gradients are surveyed. *Int. J. Appl. Earth Obs. Geoinf.* 80, 82–93. <https://doi.org/10.1016/j.jag.2019.01.002>.
- Waldner, F., Canto, G.S., Defourny, P., 2015. Automated annual cropland mapping using knowledge-based temporal features. *ISPRS J. Photogramm. Remote Sens.* 110, 1–13. <https://doi.org/10.1016/j.isprsjprs.2015.09.013>.
- Waldner, F., Jacques, D.C., Löw, F., 2017. The impact of training class proportions on binary cropland classification. *Remote Sens. Lett.* 8 (12), 1122–1131. <https://doi.org/10.1080/2150704X.2017.1362124>.
- Wang, S., Azzari, G., Lobell, D.B., 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sens. Environ.* 222, 303–317. <https://doi.org/10.1016/j.rse.2018.12.026>.
- Wang, S., Di Tommaso, S., Deines, J.M., Lobell, D.B., 2020. Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. *Sci. Data* 7 (1). <https://doi.org/10.1038/s41597-020-00646-4>.
- Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236, 111402 <https://doi.org/10.1016/j.rse.2019.111402>.
- Xu, L., Zhang, H., Wang, C., Zhang, B., Liu, M., 2018. Crop Classification Based on Temporal Information Using Sentinel-1 SAR Time-Series Data. *Remote Sensing* 11 (1), 53. <https://doi.org/10.3390/rs11010053>.
- Yan, Y., Ryu, Y., 2021. Exploring Google Street View with deep learning for crop type mapping. *ISPRS J. Photogramm. Remote Sens.* 171, 278–296. <https://doi.org/10.1016/j.isprsjprs.2020.11.022>.
- Zhong, L., Hu, L., Yu, L., Gong, P., Biging, G.S., 2016. Automated mapping of soybean and corn using phenology. *ISPRS J. Photogramm. Remote Sens.* 119, 151–164. <https://doi.org/10.1016/j.isprsjprs.2016.05.014>.