Revista Internacional de Ciencia y Tecnología de la Información Geográfica
International Review of Geographical Information Science and Technology

Alemán-Montes, B., Zabala, A., Serra, P. (2025). Classification of sugarcane varieties with harmonized Sentinel-2 and Landsat-8/9 data using parametric and non-parametric methods. *GeoFocus*, *Revista Internacional de Ciencia y Tecnología de la Información Geográfica* (Articles), 35, 25-46. http://dx.doi.org/10.21138/GF.906

CLASSIFICATION OF SUGARCANE VARIETIES WITH HARMONIZED SENTINEL-2 AND LANDSAT-8/9 DATA USING PARAMETRIC AND NON-PARAMETRIC METHODS

[1,2a]Bryan Alemán-Montes, [1b]Alaitz Zabala, [1c]Pere Serra
[1]Grumets Research Group. Departament de Geografia, Edifici B. Universitat Autònoma de Barcelona. 08193 Bellaterra, Catalonia, Spain
[2]Laboratorio de Suelos y Foliares, Centro de Investigaciones Agronómicas, Universidad de Costa Rica, 11501-2060 San Pedro Montes de Oca, San José, Costa Rica
[a]bryan.aleman@ucr.ac.cr, [b]alaitz.zabala@uab.cat, [c]pere.serra@uab.cat

ABSTRACT

Remote sensing data has been successfully used to enhance sugarcane monitoring and management, in topics such as yield estimation, health anomaly detection, or variety classification. Specifically, variety classification is an essential objective for optimizing crop management, as it can guide strategies such as plant renovation, pest control, or yield estimation. A literature review allowed identifying that the integration of diverse satellite platforms to enhance time series for sugarcane variety classification has not been explored. This strategy can improve the temporal density of available imagery in our study area, Costa Rica, with frequent cloud cover. Therefore, our research proposed to classify six sugarcane varieties using an additive approach (aggregating them in four variety groups) and employing parametric and non-parametric algorithms on harmonized data from Sentinel-2 and Landsat-8/9. Validation was done at both pixel and plot scales. The best classifications were achieved using green and near infrared bands, along with the Enhanced Bloom Index and Normalized Difference Infrared Index vegetation indices. Regarding temporal dynamics, the most relevant months were September, November, and December, corresponding to advanced growth cycle stages. Support Vector Machine and Random Forest provided the best classification accuracies. At the pixel scale, the overall accuracy of all groups exceeded 0.86, with a slight decrease as the number of varieties increased. When validation was done at plot scale, the overall accuracy remained above 0.89 in all the groups. These achievements were suitable and valuable for sugarcane sustainable planning and decision-making.

Keywords: variety classification; support vector machine; random forest; discriminant analysis; spectral bands; vegetations indices.

Received: 04/04/2025 Accepted: 17/05/2025
Academic Editor: Dr. Miquel Ninyerola www.geofocus.org

CLASIFICACIÓN DE VARIEDADES DE CAÑA DE AZÚCAR CON DATOS ARMONIZADOS DE SENTINEL-2 Y LANDSAT-8/9 USANDO MÉTODOS PARAMÉTRICOS Y NO-PARAMÉTRICOS

RESUMEN

Los datos de teledetección han mejorado el monitoreo y la gestión de la caña de azúcar en aspectos como la estimación de rendimientos, la detección de problemas sanitarios o la clasificación de variedades. Específicamente, la clasificación de variedades permite optimizar la gestión del cultivo, ya que puede orientar estrategias de renovación, control de plagas o la estimación de rendimiento. La revisión de literatura permitió identificar que la integración de plataformas satelitales para mejorar las series temporales en la clasificación de variedades de caña de azúcar no ha sido explorada. Esta estrategia aumenta la disponibilidad de imágenes en zonas con alta nubosidad, como en nuestra área de estudio en Costa Rica. Esta investigación propuso clasificar variedades de caña de azúcar mediante un enfoque aditivo (agregadas en cuatro grupos de variedades) y el uso de algoritmos paramétricos y no paramétricos sobre datos armonizados de Sentinel-2 y Landsat-8/9. La validación se realizó a escala de píxel y lote. Las mejores clasificaciones utilizaron las bandas verde e infrarrojo cercano, junto con los índices de vegetación *Enhanced Bloom Index* y *Normalized Difference Infrared Index*. En cuanto a la dinámica temporal, los meses más relevantes fueron septiembre, noviembre y diciembre, correspondientes a etapas avanzadas del ciclo de crecimiento. Los algoritmos máquina de vectores de soporte y bosque aleatorio proporcionaron las mejores exactitudes de clasificación. A escala de píxel, la exactitud general en todos los grupos superó 0.86, con una ligera disminución cuando aumentó el número de variedades. A escala de lote, la exactitud general fue superior a 0.89. Estos resultados son valiosos para la toma de decisiones en la producción de caña de azúcar.

Palabras clave: clasificación de variedades; máquina de vectores de soporte; bosque aleatorio; análisis discriminante; bandas espectrales; índices de vegetación.

1. Introduction

Sugarcane production in Costa Rica is relatively low compared to the largest sugarcane-producing countries (FAO 2024). Nonetheless, this production makes significant economic and social contributions to the regional and national economy, generating more than 16 million USD annually in taxes and social contributions (LAICA 2025). The country is divided in six producing regions according to La Liga Agrícola Industrial de la Caña de Azúcar (LAICA), a public institution that organizes Costa Rican sugarcane producers.

The sugarcane (*Saccharum officinarum*) crop is well-known for its high capacity to concentrate sucrose within the stalks. This perennial grass is grown in tropical and subtropical environments and has four phenological stages characterized by changes in biomass production throughout the growth cycle. For a better development, it requires optimal weather conditions, management activities and nutriments (Manzoor *et al.* 2023). The first stage is the seed and germination, which can occur after new planting or due to ratooning, causing partial soil cover with vegetation (Allison *et al.* 2007). The second stage is tillering and canopy development, which allows a rapid growth of sprouted plant, resulting in a complete soil cover. The third stage is great growth, during which high foliar development allows the interception of a high percentage of solar radiation, and the stalks elongate rapidly (Inman-Bamber 1994). The last stage is the maturation, characterized by an increase in sucrose concentration in the stalks, leading to a reduction in leaf area due to the yellowish and shedding of leaves (Cock 2003). Each phenological stage lasting approximately ninety days.

The global importance of the sugarcane as a source of food and energy, along with its ideal conditions for monitoring using Earth Observation (EO) data, has facilitated the development of decision-making tools based on remote sensing (RS) technologies. The main applications of RS in sugarcane cultivation include crop mapping, variety classification, crop growth anomaly, health monitoring, and yield estimation (Abdel-Rahman & Ahmed 2008, Som-Ard *et al.* 2021). The common goals of these applications are to increase the productivity, enhance understanding of sugarcane growth cycle, improve the decision-making through spatial information, and promote crop sustainability.

Specifically, the challenge of improving sugarcane productivity has driven the selection and development of sugarcane varieties adapted to local conditions. In some cases, the varieties planted are correctly identified but in others not, which become a limitation for guiding strategies such as renovation, pest control, or yield estimation (Apan *et al.* 2004, Galvão *et al.* 2005).

In cases where the sugarcane varieties are unknown, some studies have developed algorithms that use RS data to identify them. From a literature review, most research aimed to discriminate sugarcane varieties using RS data obtained from original bands and/or vegetation indices (VIs) (Appendix 1). One of the earliest studies on sugarcane variety classification was conducted by Apan *et al.* (2004) who evaluated the separability of eight sugarcane varieties using discriminant analysis (DA) through a Hyperion image. The results showed an overall classification accuracy (OA) of 97.0 % for five varieties, which decreased to 74 % when eight varieties were included. The best predictor variables were VIs generated from bands located in the visible and Near-Infrared (NIR) region. Galvão *et al.* (2005), also using a Hyperion image, assessed the discrimination of five key sugarcane varieties also using DA. Their findings indicated a classification accuracy of 87.5 % for four varieties using visible, NIR and Shortwave Infrared (SWIR) bands. Fortes and Demattê (2006), using Landsat-7 images, evaluated the discrimination of four sugarcane varieties using individual bands and discriminant equations, obtaining an OA of 93.6 %. Part of their results demonstrated that the NIR band was the most effective for classification of varieties. Similarly, Everingham *et al.* (2007) evaluated the classification of nine sugarcane varieties using diverse discriminant analysis and machine learning techniques, obtaining the best results (around an OA of 85.0 %) with Support Vector Machine (SVM) and Random Forest (RF) algorithms. Murillo-Sandoval *et al.* (2011) evaluated two Landsat-7 images to discriminate between two sugarcane varieties using individual bands and VIs and achieving an OA of 80.8 %. The most relevant variables were the Green, NIR and SWIR bands when the crop had between four and five months of crop development. Duft *et al.* (2019) assessed the classification of 25 sugarcane varieties through original bands and VIs derived from three Sentinel-2; using the RF classification algorithm they achieved an OA of 86.0 % being the SWIR bands the most important. Finally, Kai *et al.* (2022) compared different non-parametric methods to classify four sugarcane varieties using reflectance bands and VIs from Sentinel-2 images. The highest classification accuracy (99.5 %) was obtained using the SVM algorithm, while the most appropriate variables were Red-Edge and SWIR bands.

Although these studies have achieved important advances in the classification of varieties, there is still a need to evaluate the combined use of different sensors, MSI (Multispectral Imager onboard Sentinel-2A/B) and OLI (Operational Land Imager onboard Landsat-8/9) in our work, to construct complete time series throughout the sugarcane growth cycle. This strategy could enhance the temporal density of available imagery in regions with frequent cloud cover. Moreover, this study explores the following hypothesis: (1) agronomic characteristics of varieties influence classification results because they exhibit different spectral patterns along grown cycle; (2) classification accuracy decreases as the number of sugarcane varieties increases, owing to spectral similarities among them; (3) the most effective classification models require multi-temporal data covering different stages of the crop growth cycle; and (4) non-parametric methods are expected to outperform parametric approaches, as they better differentiate canopy variations among varieties at the same phenological stage.

Therefore, to address the research hypotheses, the three specific objectives of this work were: (1) to classify sugarcane varieties in Costa Rica from RS data combining MSI and OLI images and crop-related information, using an additive variety approach (to the initial first group, composed by three varieties, a fourth, a fifth and a sixth variety were added, respectively); (2) to compare the performance of two non-parametric classifiers (RF and SVM) and a parametric method (DA) for sugarcane variety classification, and (3) to validate classification results at the corresponding crop management scale, the plot, a less addressed research purpose.

2. Materials, data and methods

2.1. Study area

This classification of sugarcane varieties was developed in sugarcane plantations owned by the CoopeAgri R.L. cooperative. These plantations are located in the General Valley of Costa Rica, a region with an average altitude of ~550 m above sea level, an annual rainfall of around 3000 mm, and a mean temperature of 25 °C (Solano and Villalobos, 2001). According to Mata *et al.* (2020) the main soil orders are ultisols, oxisols, and their associations. The analysis was conducted using plot information provided in a shapefile delivered by the cooperative. The selected plots had the following characteristics: larger than 0.2 ha, contained variety information, were between the first and the fifth ratoon, and started the phenological cycle in February, March, or April of 2023 (Figure 1). In total, this study included 430 plots, covering 462.5 ha, with an average plot size of 1.07 ha.

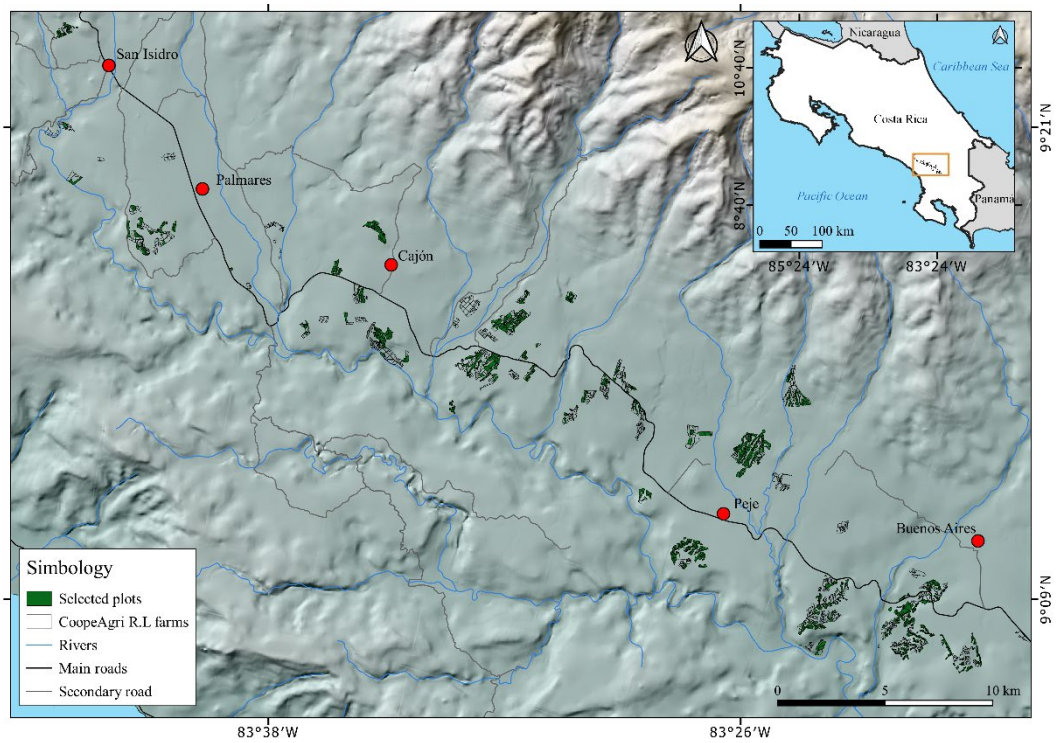

Figure 1. Location of the study area and the sugarcane production areas selected, owned by the CoopeAgri R.L.

2.2. Sugarcane varieties

The research was conducted from six main varieties cultivated in the CoopeAgri sugarcane plantation during 2023-2024 harvest season. The selected varieties were Laica 04-809, Laica 05-805, Laica 07-801, RB 86-7515, RB 98-710, and RB 99-381. All of them have a phenological cycle of 12 months, being the start not determined by the variety itself but by the logistical planning of the mill. The most relevant agronomic characteristics, which had implications for the spectral response of sugarcane through satellite remote RS, are defined in Table 1.

Table 1. The main characteristics of selected sugarcane varieties.

Variety	Characteristics
Laica 04-809	Semi-erect growth habit; large-sized leaves of medium fineness; regular leaf shedding. This variety exhibits regular ratooning, fast growth, and mid-to-late maturation. The blooming percentage is approximately 20 %.

Laica 05-805	Erect growth habit; large-sized and tapered leaves; easy leaf shedding. This variety exhibits regular ratooning, fast growth, and later maturation. The blooming percentage is around 10 %.
Laica 07-801	Erect growth habit; medium-sized leaves of short fineness; easy leaf shedding. This variety exhibits regular ratooning, fast growth, and mid maturation. The blooming percentage is around 25 %. It can be affected by orange rust.
RB 86-7515	Erect to semi-erect growth habit; large-sized leaves of medium fineness; regular leaf shedding. This variety exhibits fast ratooning and growth, and mid maturation. In this region, it has a high blooming percentage, about 100 %.
RB 98-710	Erect growth habit; large-sized leaves of short fineness; low leaf shedding. This variety exhibits fast ratooning, slow growth, and late maturation. The blooming percentage is around 30 %.
RB 99-381	Erect growth habit; medium-sized leaves of short fineness; high leaf shedding. This variety exhibits fast ratooning, fast growth, and mid-to-late maturation. The blooming percentage is low, around 10 %.

Source: own elaboration based on information from Vignola *et al.* (2018), LAICA (2023), and personal communication from Andrey Chinchilla, agronomic advisor from CoopeAgri.

2.3. Geospatial data

2.3.1 Source of geospatial data

The vector file for the sugarcane varieties was provided by CoopeAgri and consisted of a polygon file delineating the boundaries of the plots for the 2023-2024 harvest season. Each plot contained the following information: name, area, zone, variety, planting year, and age. Due to the intense presence of cloud cover and the need to improve the classification of sugarcane varieties, this work utilized raster data from Sentinel-2A/B and Landsat-8/9 platforms, covering the period from January 1, 2023, to January 31, 2024 (Figure 2). A total of 43 images were available for Sentinel-2, while 29 images were used from Landsat-8/9. Sentinel-2A/B imagery at level 2, processed with the Sen2Cor algorithm (Mueller-Wilm *et al.* 2019), were obtained from the European Space Agency through the Data Space Copernicus portal (https://dataspace.copernicus.eu/). These rasters were used at 20×20 m^2 of pixel resolution. The Landsat-8/9 images at level 2, processed using Landsat Surface Reflectance Code (USGS, 2022) were downloaded from Collection 2 of the United States Geological Survey, using the Earth Explorer portal (https://earthexplorer.usgs.gov/). These rasters were used at 30×30 m^2 of pixel resolution.

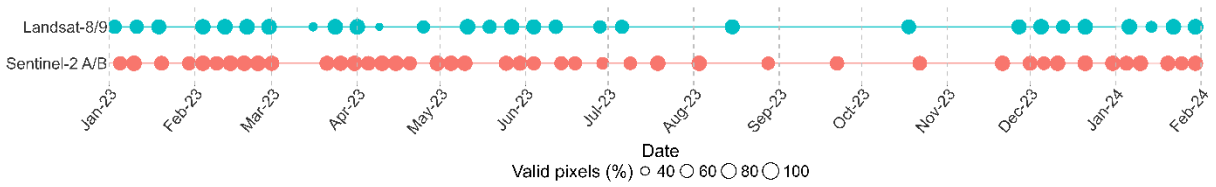

Figure 2. Available images with more than 40 % of valid pixels within the study area from January 2023 to January 2024.

2.3.2 Harmonized Sentinel and Landsat imagery

Although the level 2 radiometric processing for the MSI and OLI sensors can be considered largely equivalent due to the relatively flat terrain, a linear regression model was applied to harmonize the spectral bands between both sensors, as done in previous works (Rahman and Robson 2020, Alemán-Montes *et al.* 2023, Berra *et al.* 2024). MSI bands served as dependent variables, while OLI bands were used as independent variables, aiming to minimize possible sources of error. The $SWIR_1$ from Sentinel-

2 and Landsat-8/9 is centred at 1.61 µm and $SWIR_2$ at 2.19 µm. The harmonization process involved randomly selecting only pixels from both sensors that spatially coincided within the images path and unaffected by clouds, shadows, and other atmospheric noises, resulting in a total of 548 selected pixels. Reflectance values for each band were extracted, and with linear regression models were fitted accordingly. To ensure, robust calibration, four images from each sensor were acquired on the same dates (February 4, 2023; June 4, 2023; December 21, 2023; and January 30, 2024). In total, reflectance data from 2192 pixels were used to calibrate the harmonization model for each band (Figure 3).

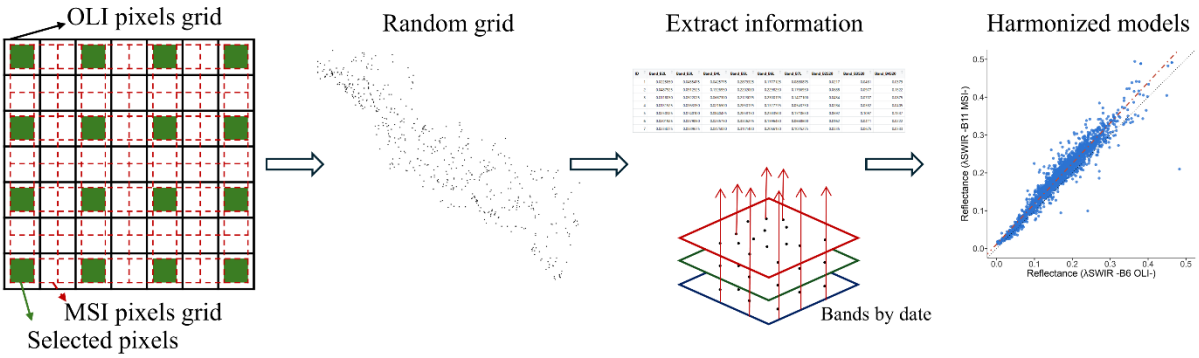

Figure 3. Workflow for harmonizing OLI and MSI reflectance based on common grids pixels.

2.3.3 Dataset construction for classification

The harmonization equations obtained were applied to each corresponding band of the entire selected Landsat imagery archive to reduce radiometric differences between the sensors. Following the harmonization process, six VIs were calculated using the harmonized Landsat-8/9 bands and the original Sentinel-2A/B bands. The selected VIs, the most used in sugarcane research, were: Enhanced Bloom Index (EBI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Vegetation Index (NDVI), Simple Ratio (SR), Normalized Difference Infrared Index 1 ($NDII_1$) and Normalized Difference Infrared Index 2 ($NDII_2$) (Table 2).

Table 2. Vegetation indexes (VIs) obtained from the satellite images.

Vegetation Index	Equation	Source
EBI	$\frac{Red + Green + Blue}{\frac{Green}{Blue} \times (Red - Blue + \varepsilon)}$	(Chen *et al.* 2019)
GNDVI	$\frac{Green - Red}{Green + Red}$	(Gitelson *et al.* 1996)
NDVI	$\frac{NIR - Red}{NIR + Red}$	(Rouse *et al.* 1973)
SR*	$\frac{NIR}{Red}$	(Jordan, 1969)
$NDII_1$	$\frac{NIR - SWIR_1}{NIR - SWIR_1}$	(Hardisky *et al.* 1983)
$NDII_2$	$\frac{NIR - SWIR_2}{NIR - SWIR_2}$	(Hardisky *et al.* 1983)

*This has also been defined in the literature as VIN (Vegetation Index Number) or RVI (Ratio Vegetation Index). ε is an adjusting constant to ensure the non-negative denominator; in this work was set to 1 because the reflectance data was ranging from 0 to 1.

Although a cloud mask was applied to each band using the Scene Classification Layer for Sentinel-2 and the Pixel Quality Assessment for Landsat-8/9, some problematic pixels were detected yet (due to haze, undetected clouds, or unnoticed shadows). To address this issue, a visual inspection was performed on all plots to identify data affected by interferences, which were subsequently removed. Additionally, to mitigate the boundary effects (presence of trees and their shadows, and roads very near to the plots), an erosion tool of 10-meters for all polygons was applied using RStudio, ensuring that remained data stayed unaffected by external anomalies (Serra *et al.* 2003). The 10-meters applied erosion preserved the shape of the polygons while removing external noise considering that satellite data had pixel sizes of 20×20 m^2 and 30×30 m^2. This choice balances the need to ensure result quality with minimal loss of usable crop area.

Subsequently, a dataset was built joining variety information, and values of all the reflectance bands and VIs for all the available dates. Afterwards, a new point vector file was created from each Sentinel-2 pixel (20×20 m^2), linking each point to its corresponding information. Later, using the *raster*, *terra*, and *sf* packages in RStudio, spectral bands and VI values were extracted and integrated with variety information to generate the final dataset for classification. Exceptionally, spectral bands and VIs were available for 72 dates per point because information was missing on some of them due to atmospheric conditions. To address this problem and create a dataset with the largest number of values, spectral and VI data were aggregated by month, calculating the median for all available dates within each month. This process was applied to all points, being removed those with missing data. Finally, a dataset of 6505 points from a total of 10983 was obtained. Due to differences in sample sizes among varieties, 500 random points per variety were selected for classifying. The final dataset was then randomly split, with 80 % used for training and 20 % for validation, considering these percentages as used in other remote sensing studies (Kai *et al.* 2022, Schulthess *et al.* 2023). The 3505 points not used were included in a second validation process at plot scale, which is explained in the following section.

2.4. Classifications methods and validation

This study evaluated the three most common sugarcane classifiers —Random Forest (RF), Support Vector Machine (SVM), and Discriminant Analysis (DA)— using an additive approach for varieties discrimination using RS data (Fortes and Demattê 2006, Everingham *et al.* 2007, Duft *et al.* 2019, Kai *et al.* 2022). The classification process was conducted across four variety classification groups, being each group of varieties selected according to the most successful combination: (1) initially, three sugarcane varieties (Laica 05-805, RB 86-7515, and RB 98-170) were analysed because they gave the best results; (2) in the second stage, the variety RB 99-381 was then added to the classification because it yielded the best results; (3) in the third stage, the variety Laica 07-801 was incorporated for the same reason; and (4) finally, the variety Laica 04-809 was included, completing the full classification group. The main characteristics of each method are defined as follow.

Random forest is a non-parametric method of classification and regression that ensembles estimates' trees from explanatory variables with the aim to finding a prediction function *f(x)* for forecasting *Y*. The result obtained combines the solutions of multiples decision trees to produce a unique result (Breiman 2001). Each tree is trained using a random sample of data using a random subset of predictor variables (Cutler *et al.* 2012). The RF classification algorithm was applied in RStudio using the "randomForest" package (Breiman *et al.* 2024). The hyperparameters that control the structure of each tree (and the forests) were adjusted using the optimal *mtry* (number of candidate variables considered at each split) and *ntree* (number of trees in the forest), and the best threshold for the *MeanDecreaseGini* (importance of variable) was defined. The last adjustment considered the minimum number of variables that allowed better classification with low correlation.

Support vector machine is a non-parametric method used for classification and regression (Sheykhmousa *et al.* 2020). The algorithm objective is to maximize the space between classes from the construction of an optimal hyperplane (limit of decision among categories) that separates the space in a discrete number or classes (Suykens and Vandewalle 1999, Mountrakis *et al.* 2011). The best hyperparameters for this method were obtained through a *grid search* to find the optimal values for *cost* (C) and *gamma* (γ), using the combinations (0.1, 1, 10) for *cost* and (0.01, 0.1, 1) for *gamma* (γ), the *grid search* method has been used in other study with RS data (Saini and Ghosh 2018, Wang *et al.*

2020). The optimization was performed using the *tune()* function from the *e1071* package in RStudio, which evaluated the performance of the hyperparameter combinations through cross-validation. Finally, the optimal values for *cost* and *gamma* were used to train the model using the Radial Basis Function kernel. This method was selected due to its ability to model complex nonlinear relationships in RS data (Razaque *et al.* 2021).

Discriminant analysis is a statistical method that generates discriminant functions based on linear contributions from predictor variables, allowing for the identification of which variables play a more significant role in classification. These functions maximize the statistical differences between predefined groups while also reducing the statistical dimensionality. One advantage is that the coefficients provide significant information applicable to other available data. However, the method assumes that the data follow a normal distribution to produce reliable results, and its performance decreases when dealing with a high number of predictor variables (Apan *et al.* 2004, Galvão *et al.* 2005, Ramayah *et al.* 2010, Moré *et al.* 2011). The selection of explanatory variables was initially done using the Bayesian Model Averaging, which allows for the identification of the most suitable variables to classify different groups using a probabilistic assessment of each variable contribution. This method was also applied from RStudio using the "BMA" package (Raftery *et al.* 2025).

To validate the results for each classification method and variety classification group, the following metrics were calculated: Kappa index (KI), Overall accuracy (OA), Producer accuracy (PA), and User accuracy (UA), which have been widely used to evaluate classification in RS approaches (Liu *et al.* 2007, Rwanga and Ndambuki 2017). Once the best classification model by variety group was obtained, it was applied to the raster data (at 20×20 m^2) that included all the study area, 6505 pixels (3000+3505), allowing for an assessment of spatial variability at the plot scale. Using the modal value from raster data inside each plot, the predominant variety was determined with QGIS. These results were then compared with the reference variety, and accuracy metrics were calculated at plot scale with error matrices.

3. Results

3.1. Harmonization of images

The harmonization process between the bands of the MSI and OLI sensors showed enough adjustment. The coefficients of determination (R^2 with p-value < 0.05) ranged from 0.91 (Red and NIR bands) to 0.78 (Blue band), while the root mean square error (RMSE) varied between 0.02 (Green band) and 0.03 (Blue band) (Figure 4). These results confirm the suitability of the approach. After developing this local harmonization, it was possible to adjust the bands of OLI sensor, reducing the amount of missing data and enabling a more robust time series.

3.2. Understanding the temporal dynamics of sugarcane varieties

The temporal dynamics of bands and selected VIs facilitates a deeper understanding of the growth cycle patterns for each variety. Certain varieties showed distinct differences when analysed through the patterns of bands and VIs. For the varieties' classification, the bands that exhibited the most notable differences among varieties were Blue, Green, NIR, and $SWIR_1$, becoming apparent after May (Figure 5A-B-D-E). A noteworthy pattern was observed in the varieties Laica 04-809, Laica 07-801, and RB 99-381, which displayed similar patterns across almost all bands. The variety RB 98-710 had the highest reflectance after May in the Blue and NIR bands (Figure 5B-D). The Red and $SWIR_1$ bands showed the grouping of the varieties into two groups after October: one group consisting of Laica 05-805, RB 86-7515, and RB 98-710, and the other group consisting of Laica 04-809, Laica 07-801, and RB 99-381 (Figure 5C-E-F).

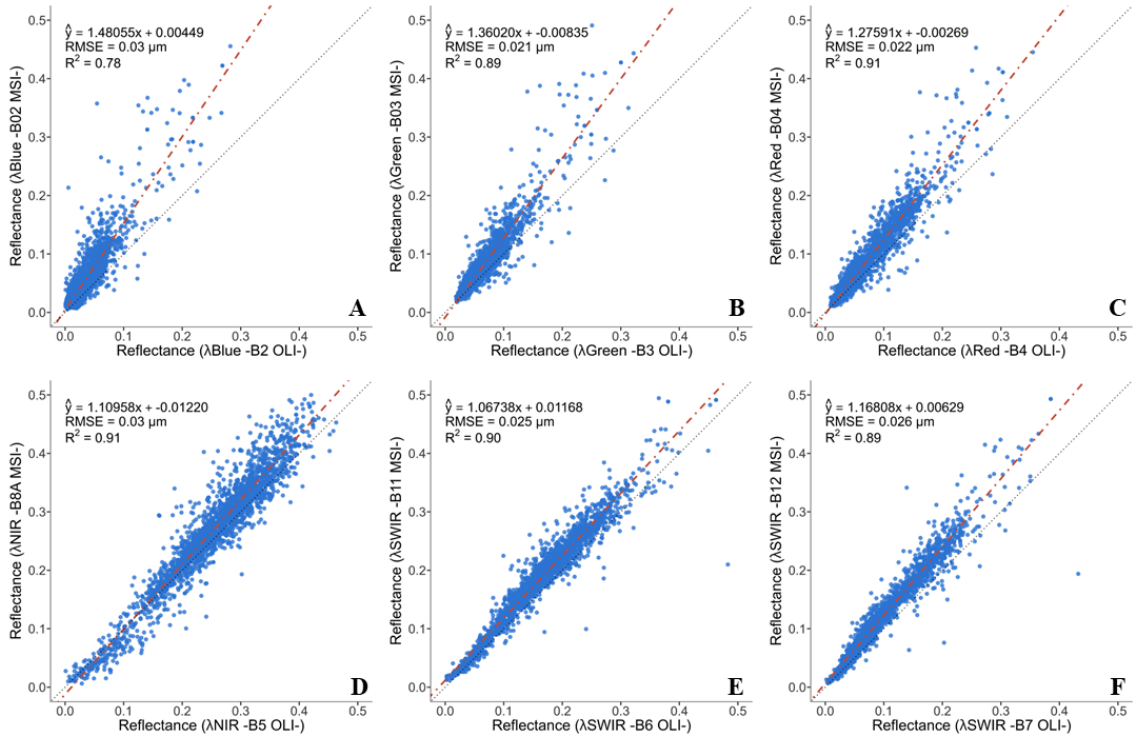

Figure 4. Regressions to harmonize OLI bands with MSI bands from coincident pixels selection. Blue (A), Green (B), Red (C), NIR (D), $SWIR_1$ (E), and $SWIR_2$ (F).

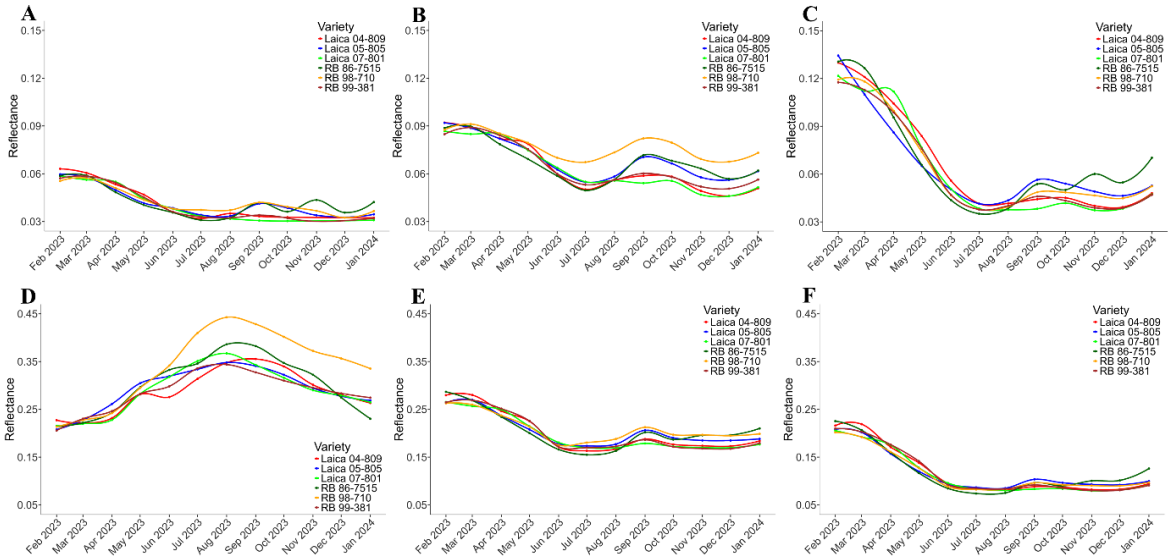

Figure 5. Temporal dynamics of six sugarcane varieties using the reflectance in different spectral bands. Blue (A), Green (B), Red (C), NIR (D), $SWIR_1$ (E), and $SWIR_2$ (F).

The patterns of the sugarcane growth cycle by variety also showed differences when they were analysed by VIs. For instance, EBI initially displayed a decreasing trend until approximately July-August. After that, the varieties Laica 05-805, RB 86-7515, and RB 98-710 showed the most variability compared to the other three varieties (Figure 6A). The other VIs began with an increasing trend, reaching a maximum in July-August. The GNDVI showed the greatest difference between varieties after August, with Laica 05-805 having the lowest values (Figure 6B). This behaviour was also observed in NDVI, SR, $NDII_1$, and $NDII_2$ (Figure 6C-F). The Laica 04-809 experienced the smallest increase in June compared to the other varieties, which was reflected in the GNDVI and NDVI (Figure 6B-C). The SR

showed significant differences between the varieties, with RB 98-710 having the highest values almost every month, while Laica 05-805 had the lowest (Figure 6D). This pattern was also observed in $NDII_1$ and $NDII_2$, but with smaller differences between varieties (Figure 6E-F).

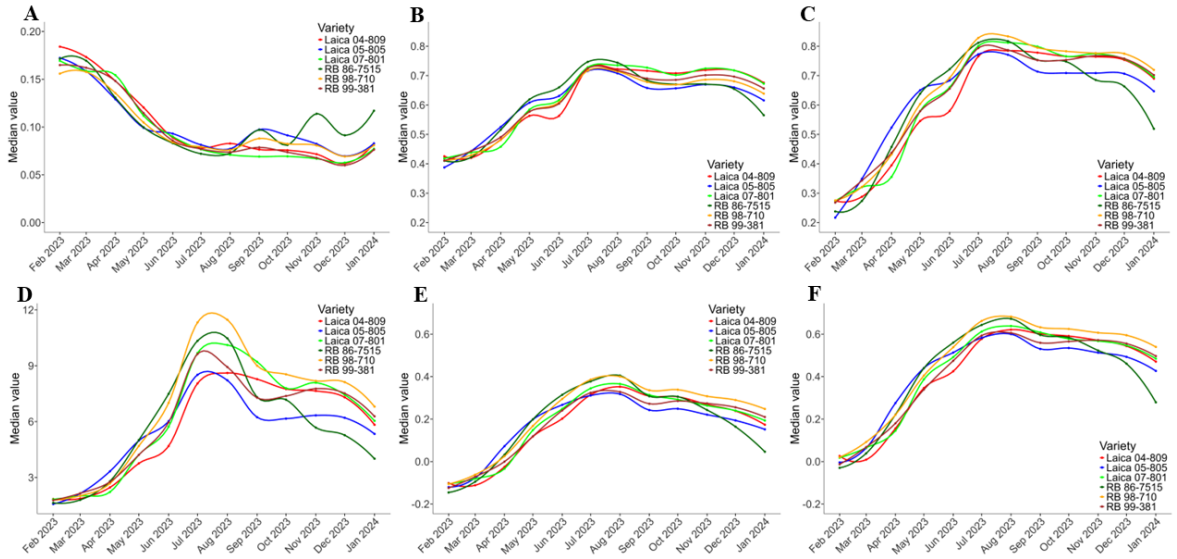

Figure 6. Temporal evolution of six sugarcane varieties using different vegetation indices. EBI (A), GNDVI (B), NDVI (C), SR (D), $NDII_1$ (E), and $NDII_2$ (F).

3.3. Sugarcane variety classification and validation results

The first group included Laica 05-805, RB 86-7515, and RB 98-710 varieties. The predictor variables selected from the optimal threshold of the *MeanDecreaseGini* of RF, using the best *mtry* (8) and *ntree* (900), were the Green and NIR bands of July, $NDII_1$ of September, $NDII_2$ of December, and EBI of January. KI and OA were 0.87 and 0.91, respectively. For the SVM, using the same explanatory variables and adjusting the best hyperparameters for *cost* (1) and *gamma* (1), the KI and OA were 0.90 and 0.94, respectively. The final method used was DA, which showed a KI of 0.79 and an OA of 0.86 (Table 3). The selection of the most relevant variables for DA was conducted using both Bayesian Model Averaging and the *MeanDecreaseGini* method. In almost all cases, including the following three variety classification groups, the selection using *MeanDecreaseGini* showed the best results (Appendix 2).

The second group involved the three previous varieties plus RB 99-381. The predictor variables selected, using the optimal *MeanDecreaseGini* threshold from the best *mtry* (7) and *ntree* (700), were the Green band of July, NIR band of September, $SWIR_1$ band of December, SR of January, $NDII_1$ of October, and EBI of November. The KI and OA were 0.81 and 0.86, respectively. The SVM, also using the same explanatory variables and adjusting the hyperparameters *cost* (1) and *gamma* (1), had a KI of 0.83 and an OA of 0.87. Finally, DA showed the lowest values for KI (0.70) and OA (0.77) (Table 3).

The third group of varieties included the addition of Laica 07-801. The best predictor variables selected using the same criteria were Age, the Green bands of July and November, the NIR band of January, and the VIs $NDII_2$ of September and NDVI of December. The optimal values for *mtry* and *ntree* were 3 and 900, respectively. RF showed the best results in KI and OA, with values of 0.85 and 0.88, respectively. With SVM, adjusting the hyperparameters *cost* (10) and *gamma* (1), the KI and OA were 0.83 and 0.87, respectively. Finally, DA showed the lowest values for all indicators, with KI and OA values of 0.56 and 0.64, respectively (Table 3).

The fourth group included all the evaluated varieties. The best predictor variables selected were Age, the SR of June, NIR band of August, Blue band of September, $SWIR_1$ and EBI of November, and Green band of December, using an *mtry* of 8 and *ntree* of 1000. For RF, the KI and OA were 0.81 and 0.84, respectively. The SVM, using the hyperparameters *cost* (10) and *gamma* (1), showed the highest values for KI and OA, 0.86 and 0.88, respectively. In DA, the KI and OA were the lowest values, 0.51

and 0.59, respectively (Table 3). Finally, table 4 shows the confusion matrix of SVM selected as a best model in the fourth group with six varieties.

Table 3. Validation results of sugarcane varieties classification using Random Forest (RF), Support Vector Machine (SVM), and Discriminant Analysis (DA).

	Kappa index			Overall accuracy			User accuracy			Producer accuracy		
Varieties	RF	SVM	DA	RF	SVM	DA	RF	SVM	DA	RF	SVM	DA
Laica 05-805							0.90	0.92	0.81	0.90	0.93	0.80
RB 86-7515	0.87	0.90	0.79	0.91	0.94	0.86	0.95	0.98	0.88	0.90	0.94	0.98
RB 98-710							0.90	0.91	0.89	0.94	0.94	0.82
Laica 05-805							0.79	0.82	0.68	0.87	0.87	0.68
RB 86-7515	0.81	0.83	0.70	0.86	0.87	0.77	0.94	0.93	0.72	0.90	0.89	0.91
RB 98-710							0.89	0.89	0.84	0.79	0.82	0.76
RB 99-381							0.84	0.85	0.84	0.88	0.90	0.74
Laica 05-805							0.82	0.83	0.47	0.88	0.87	0.48
RB 86-7515							0.94	0.91	0.70	0.88	0.87	0.81
RB 98-710	0.85	0.83	0.56	0.88	0.87	0.64	0.88	0.86	0.84	0.93	0.92	0.76
RB 99-381							0.84	0.85	0.37	0.83	0.83	0.40
Laica 07-801							0.92	0.90	0.84	0.87	0.85	0.74
Laica 05-805							0.82	0.89	0.40	0.88	0.92	0.58
RB 86-7515							0.88	0.92	0.77	0.86	0.92	0.71
RB 98-710	0.81	0.86	0.51	0.84	0.88	0.59	0.85	0.85	0.78	0.81	0.88	0.71
RB 99-381							0.80	0.90	0.71	0.84	0.82	0.59
Laica 07-801							0.88	0.84	0.65	0.80	0.89	0.54
Laica 04-809							0.85	0.90	0.25	0.88	0.86	0.33

Table 4. Confusion matrix of the best model (SVM) for six varieties by pixels.

		Reference data							User accuracy
		Laica 05-805	RB 86-7515	RB 98-710	RB 99-381	Laica 07-801	Laica 04-809	Total pixels	
Classified data	Laica 05-805	**92**	5	5	0	1	0	103	0.89
	RB 86-7515	3	**92**	2	0	1	2	100	0.92
	RB 98-710	1	1	**88**	10	4	0	104	0.85
	RB 99-381	1	2	3	**82**	2	1	91	0.90
	Laica 07-801	1	0	0	5	**89**	11	106	0.84
	Laica 04-809	2	0	2	3	3	**86**	96	0.90
	Total pixels	100	100	100	100	100	100	**600**	
	Producer accuracy	0.92	0.92	0.88	0.82	0.89	0.86		Overall Accuracy **0.88**

3.4. Spatial validation by plots

Once determining the predominant variety (modal value) inside each plot, obtained from the best classification method for each variety group, the results were compared with the reference variety, and the accuracy metrics were calculated. The variety classification group with three varieties had a KI of 0.86 and an OA of 0.92. In the second variety classification group, the KI was also 0.86, and the OA was 0.90 (Table 5). The variety classification group for five varieties also showed satisfactory results, being the KI of 0.93 and the OA of 0.94, while the classification with six varieties also presented good adjustments: KI was 0.90 and OA was 0.92.

Table 5. Validation results of sugarcane varieties classification at plot scale using the best method for each variety classification groups.

Varieties	Kappa index	Overall accuracy	Plots	User accuracy	Producer accuracy
Laica 05-805	0.86	0.92	32	0.94	0.79
RB 86-7515			39	0.90	0.92
RB 98-710			108	0.93	0.97
Laica 05-805	0.86	0.90	32	0.94	0.79
RB 86-7515			39	0.95	0.84
RB 98-710			108	0.87	0.98
RB 99-381			61	0.90	0.89
Laica 05-805	0.93	0.94	32	1.00	0.89
RB 86-7515			39	0.97	0.88
RB 98-710			108	0.94	0.97
RB 99-381			61	0.93	0.92
Laica 07-801			135	0.93	0.97
Laica 05-805	0.90	0.92	32	0.97	0.91
RB 86-7515			39	1.00	0.90
RB 98-710			108	0.92	0.97
RB 99-381			61	0.90	0.86
Laica 07-801			135	0.91	0.94
Laica 04-809			48	0.92	0.90

These results are also supported by the spatial variability of each variety. Figure 7 shows examples of the spatial variability of classified varieties in four different farms, using the best classification model for the six-variety classification group. As previously defined, each farm is divided into plots (polygons with black boundaries), and these are cultivated with the same sugarcane variety. Most of the plots have not pixels with different varieties (variety noise), which confirms the satisfactory model adjustment. In most cases, it was observed that variety-induced noise was primarily located at plot edges and was subsequently minimized through the previously described erosion procedure.

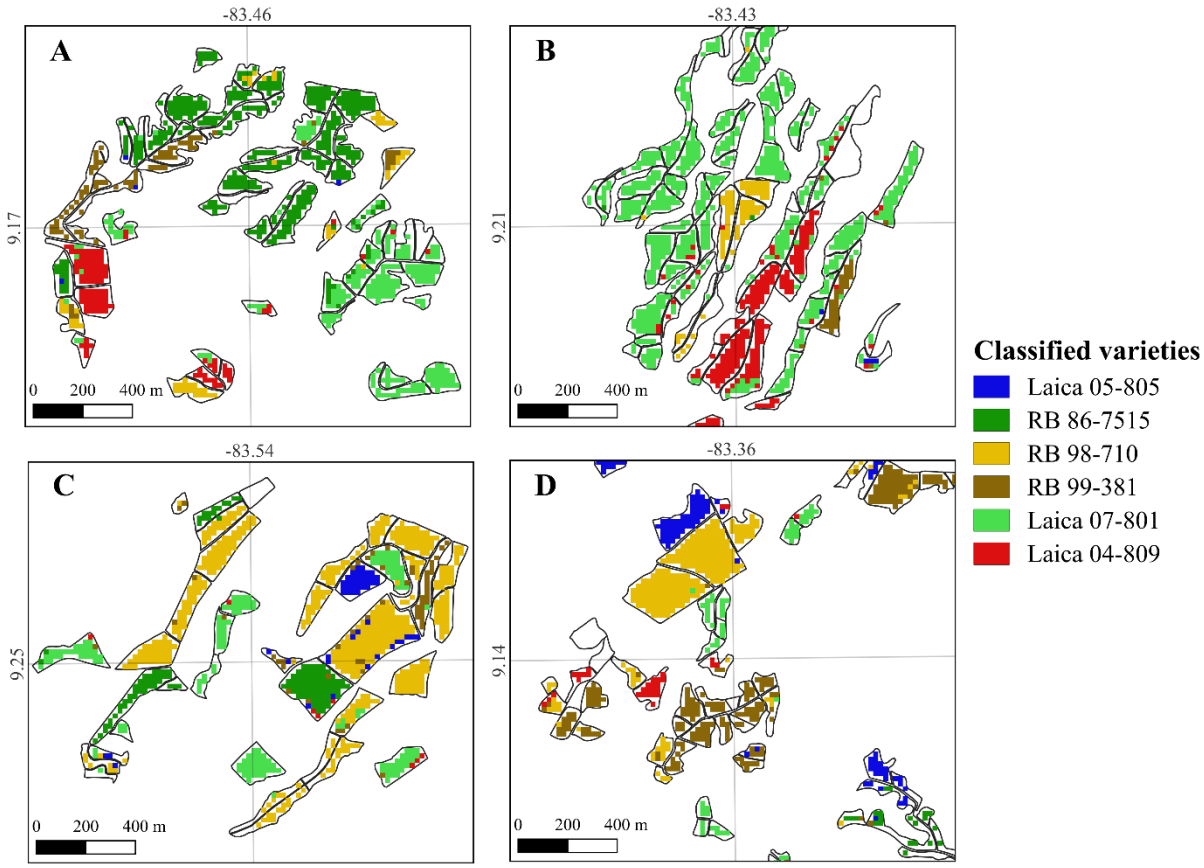

Figure 7. Spatial variability of sugarcane varieties classified in CoopeAgri, around the following farms: Peje, Peje Weber, and Peje Nuevo (A), Volcán (B), Porvenir (C), and Guayacan (D).

4. Discussion

Harmonization of different satellite imagery represents an effective strategy for improving RS time series, as it enables gap-filling and enhances available information, particularly in regions with high cloudiness. In our study area, the period from August to November exhibited the highest cloud-related noise (Figure 2). MSI and OLI have advantages for harmonization due to their similarities in spectral bands and angular characteristics (Mandanici & Bitelli 2016, Claverie *et al.* 2018). Although our harmonization results were, in general, better than those obtained by Berra *et al.* (2024), in both cases the blue band showed the lowest R^2, which can be attributed to the greater sensitivity of blue wavelengths to atmospheric effects as aerosol absorption. Besides, the approach of aggregating the spectral information and VIs by month allows for a robust time series for each sugarcane variety.

The reflectance of each band and VIs throughout the growth cycle allowed us to understand and confirm that the temporal and spectral patterns for each variety can differ due to physiological and morphological characteristics (Fortes and Demattê 2006). This fact is clearly visible in the case of RB 98-710 variety, which showed the highest differences with the other varieties in the Green and NIR bands, as well as NDVI, SR and $NDII_2$ (Figures 5 and 6). The reason of this clear difference can be attributed to its agronomic characteristics, because this variety has erected growth with large-size leaves and fast ratooning (Table 1). The RB 86-7515 variety showed the second highest differences in the NIR band and in the same VIs before August-September, after which it displayed a clear decrease until January. This pattern is related to the high blooming on this variety, which is also indicated by the increase in the EBI VI. The Laica 05-805 and Laica 07-801 varieties alternated as the third variety with higher greenness based on NDVI and SR VIs. Both have an erected growth habit, with the important difference that Laica 05-805 decreases more sharply towards the end of the harvest cycle, as shown in the NIR band, which can be related to its mid-maturation (Figures 5 and 6). The Laica 04-809 variety

had the lowest reflectance, as seen in NIR band and NDVI, SR, $NDII_1$, and $NDII_2$. This pattern was most evident before July-August, possibly due to its semi-erect growth habit and the large-size leaves with medium fineness. Finally, RB 99-381 variety showed the highest similarities with the other varieties.

The non-parametric methods were the best classifiers, with SVM being the top performer in all three groups and RF in one. In other studies regarding sugarcane variety classification, SVM has also been the method that shows the best results (Everingham *et al.* 2007, Kai *et al.* 2022). Although SVM and RF are comparable in most results, the SVM ability to generalize complex characteristics could be the reason to better outcomes (Mountrakis *et al.* 2011). According to Galvão *et al.* (2005), OA has a general decrease trend when the number of varieties increase because the complexity of discrimination, due to less differences among varieties, also growths. Nevertheless, in our case this trend was not so clear, because some slight differences were obtained: 0.94, 0.87, 0.88, and 0.88 for the groups with three, four, five, and six varieties, respectively, at pixel scale, and 0.92, 0.90, 0.94, and 0.92, respectively, at plot scale.

In our modelling, Green and NIR bands were the most common predictor variables, appearing in all the best models, while SWIR bands were important for two best models (Appendix 3). The Red band was not selected as an explanatory variable in any model. The importance of Green, NIR and SWIR bands in classifying sugarcane varieties was also found by Galvão *et al.* (2005), Fortes and Demattê (2006), Murillo-Sandoval *et al.* (2011), Duft *et al.* (2019), Kai *et al.* (2022). The Blue single band was only important in one model. Regarding the VIs, the most common were EBI and $NDII_1$, both significant in three models, followed by SR, which was present in two models, while NDVI and $NDII_2$ just were presents in only one model. In some previous studies, spectral bands were the primary explanatory variables; however, our results highlight the importance of VIs as well, marking a relevant difference from some prior research. Additionally, Age emerged as a relevant variable, particularly in classification groups three and four, where an increasing number of varieties made its discriminative power more evident.

According to our results, the most relevant month was December (around nine months of sugarcane development), four times selected, followed by September, November, and January (around six, eight, and ten months of development, respectively), each three times selected; therefore, they were in advanced growth cycle stages. These findings differ from those obtained by Murillo-Sandoval *et al.* (2011) because they concluded that the only period for obtaining a clear discrimination of varieties was between 4^{th} and 5^{th} months of growth development.

The OA at the pixel scale was lower than at the plot scale for the groups with four, five, and six varieties, while in the groups with three varieties, the behaviour was the opposite. The best OA at the pixel scale was 0.94, 0.87, 0.88, and 0.88 for the groups with three, four, five, and six varieties, respectively. At the plot scale, the OA of the selected models was 0.92, 0.90, 0.94, and 0.92, respectively. Our results are better than those obtained by Apan *et al.* (2004), Galvão *et al.* (2005), Murillo-Sandoval *et al.* (2011), and Duft *et al.* (2019). However, our results were lower than those obtained by Fortes and Demattê (2006) who evaluated the classification of five sugarcane varieties, one of which was easy to classify just using a threshold in the NIR band. Another case with better results was Kai *et al.* (2022) who evaluated the classification of four sugarcane varieties using non-parametric methods. The main difference with our work is that in their evaluated varieties, the growth cycle started at different times, an advantage for improving classification, while in our study, the growth cycle of all varieties started at the same time (from January to April), decreasing their separability.

5. Conclusions

This work contributes to improve the sugarcane varieties classification using a harmonization of Sentinel and Landsat images, an additive variety approach (with four groups for six varieties), and a combination of parametric and non-parametric methods at pixel and plot scale, with an OA higher than 90 %.

In cloudy regions, the harmonization process of different satellite imagery is a rewarding technique that enhances RS time series. In this work, a complete Sentinel/Landsat time series aggregated by months was essential to understand the sugarcane growth cycle by varieties. Each variety has different physiological and morphological characteristics that can be identified from spectral information, such as the bloom phenomenon in RB 86-7515.

In our study, the non-parametric methods provided the best results, with SVM performing better in three groups (with an OA from 0.87 to 0.94) and RF in one case (0.88). The Green and NIR bands, along with the EBI, were more suitable for identifying differences among varieties. Regarding temporal information, September, November, and December contributed to the best selected models, corresponding to advanced growth cycle stages. The validation results were successful at both pixel and plot scales but the latter being much better in almost all cases. Plot scale validation is a valuable approach because agronomic management practices are carried out at this scale, becoming more relevant in crop management.

Our future work will be to evaluate the inclusion of other satellite sensors, such as Synthetic Aperture Radar, and other bands, for instance, the Red Edge. The second objective will be the evaluation of other non-parametric methods, for example, k-Nearest Neighbours (kNN) or Artificial Neural Networks (ANN) as applied by Kai *et al.* (2022) among others. The last aim will be to apply these models to years or farms not included in the current analysis.

6. Acknowledgments

The authors would like to express their gratitude to the Agricultural Department of CoopeAgri R.L. for its invaluable support in the development of this research.

7. Funding

This research was funded by Oficina de Asuntos Internacionales y Cooperación Externa (OAICE) of the Universidad de Costa Rica through contract number OAICE-59-2021.

References

Abdel-Rahman, E.M., Ahmed, F.B. (2008). The application of remote sensing techniques to sugarcane (Saccharum spp. hybrid) production: A review of the literature. *International Journal of Remote Sensing*, 29(13), 3753–3767. **https://doi.org/10.1080/01431160701874603**

Alemán-Montes, B., Serra, P., Zabala, A. (2023). Modelos para la estimación del rendimiento de la caña de azúcar en Costa Rica con datos de campo e índices de vegetación. *Revista de Teledetección*, 2023(61), 1–13. **https://doi.org/10.4995/raet.2023.18705**

Allison, J.C.S., Pammenter, N.W., Haslam, R.J. (2007). Why does sugarcane (Saccharum sp. hybrid) grow slowly? *South African Journal of Botany*, 73(4), 546–551. **https://doi.org/10.1016/j.sajb.2007.04.065**

Apan, A., Held, A., Phinn, S., Markley, J. (2004). Spectral discrimination and classification of sugarcane varieties using EO-1 Hyperion Hyperspectral Imagery. In *Proceedings of the 25th Asian Conference on Remote Sensing (ACRS 2004)*, 1–5. https://core.ac.uk/download/pdf/211499836.pdf

Berra, E.F., Fontana, D.C., Yin, F., Breunig, F.M. (2024). Harmonized Landsat and Sentinel-2 Data with Google Earth Engine. *Remote Sensing*, 16(15). **https://doi.org/10.3390/rs16152695**

Breiman, L. (2001). Random forests, in R.E. Schapire (Ed.), *Machine learning.* 45, 5–32. **http://dx.doi.org/10.1023/A:1010933404324**

Breiman, L., Cutler, A., Liaw, A., Wiener, M. (2024). Package 'randomForest'. In Entomologia Experimentalis et Applicata. **https://doi.org/10.1023/A**

Chen, B., Jin, Y., Brown, P. (2019). An enhanced bloom index for quantifying floral phenology using multi-scale remote sensing observations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 156, 108–120. **https://doi.org/10.1016/j.isprsjprs.2019.08.006**

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S V., Justice, C. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219(August 2017), 145–161. **https://doi.org/10.1016/j.rse.2018.09.002**

Cock, J.H. (2003). Sugarcane growth and development. International Sugar Journal, 105(1259), 540–552.

Cutler, A., Cutler, D.R., Stevens, J.R. (2012). Random Forests. In Ensemble machine learning: Methods and applications (pp. 157–175). **https://doi.org/10.1007/978-1-4419-9326-7**

Duft, D.G., Luciano, A.C.S., Fiorio, P.R. (2019). Sentinel-2B and Random Forest algorithm potential for sugarcane varieties identification. In *XX GEOINFO*, November, 188–193.

Everingham, Y.L., Lowe, K.H., Donald, D.A., Coomans, D.H., Markley, J. (2007). Advanced satellite imagery to classify sugarcane crop characteristics. *Agronomy for Sustainable Development*, 27(2), 111–117. **https://doi.org/10.1051/agro:2006034**

FAO, (Food and Agriculture Organization). (2024). Crop and livestock products. Accessed 01-03-2025. Available at https://www.fao.org/faostat/en/#data/QCL

Fortes, C., Demattê, J.A.M. (2006). Discrimination of sugarcane varieties using Landsat 7 ETM+ spectral data. *International Journal of Remote Sensing*, 27(7), 1395–1412. **https://doi.org/10.1080/01431160500383863**

Galvão, L.S., Formaggio, A.R., Tisot, D.A. (2005). Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. *Remote Sensing of Environment*, 94(4), 523–534. **https://doi.org/10.1016/j.rse.2004.11.012**

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sensing of Environment*, 58(3), 289–298. **https://doi.org/10.1016/S0034-4257(96)00072-7**

Hardisky, M.A., Klemas, V., Smart, R.M. (1983). The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of Spartina alterniflora canopies. *Photogrammetric Engineering & Remote Sensing*, 49(1), 77–83.

Inman-Bamber, N.G. (1994). Temperature and seasonal effects on canopy development and light interception of sugarcane. *Field Crops Research*, 36(1), 41–51. **https://doi.org/10.1016/0378-4290(94)90051-5**

Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology*, 50(4), 663–666. **https://doi.org/10.2307/1936256**

Kai, P.M., de Oliveira, B.M., da Costa, R.M. (2022). Deep Learning-Based Method for Classification of Sugarcane Varieties. *Agronomy*, 12(11), 2722. **https://doi.org/10.3390/agronomy12112722**

LAICA. (2023). Ficha Técnica de Variedades. Accessed 15-01-2025. Available at https://servicios.laica.co.cr/laica-asistencia/

LAICA. (2025). LAICA. Accessed 10-02-2025. Available at https://laica.cr/

Liu, C., Frazier, P., Kumar, L. (2007). Comparative assessment of the measures of thematic classification accuracy. *Remote Sensing of Environment*, 107(4), 606–616. **https://doi.org/10.1016/j.rse.2006.10.010**

Mandanici, E., Bitelli, G. (2016). Preliminary comparison of Sentinel-2 and Landsat 8 imagery for a combined use. *Remote Sensing*, 8, (12), 1014. **https://doi.org/10.3390/rs8121014**

Manzoor, M., Khan, M.Z., Ahmad, S., Alqahtani, M.D., Shabaan, M., Sarwar, S., Hameed, M.A., Zulfiqar, U., Hussain, S., Ali, M.F., Ahmad, M., Haider, F.U. (2023). Ecological Zones and Irrigation Sources Amidst Environmental Stressors. *Plants*, 12(20), 3526. **https://doi.org/10.3390/plants12203526**

Mata, R., Rosales, A., Sandoval, Da., Vindas, E., Alemán, B. (2020). Subórdenes de suelos de Costa Rica [mapa digital]. Escala 1:200000. Accessed 15-01-2024. Available at http://www.cia.ucr.ac.cr/es/mapa-de-suelos-de-costa-rica

Moré, G., Serra, P., Pons, X. (2011). Multitemporal flooding dynamics of rice fields by means of discriminant analysis of radiometrically corrected remote sensing imagery. *International Journal of Remote Sensing*, 32(7), 1983–2011. **https://doi.org/10.1080/01431161003645816**

Mountrakis, G., Im, J., Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. **https://doi.org/10.1016/j.isprsjprs.2010.11.001**

Mueller-Wilm, U., Devignot, O., Pessiot, L. (2019). Sen2Cor Configuration and User Manual. Accessed 07-06-2023. Available at http://step.esa.int/thirdparties/sen2cor/2.3.0/[L2A-SUM]_S2-PDGS-MPC-L2A-SUM [2.3.0].pdf

Murillo-Sandoval, P.J., Carbonell-Gonzalez, J.C., Osorio-Murillo, C.A. (2011). Evaluation of Landsat 7 ETM + Data for Spectral Discrimination and Classification of Sugarcane Varieties in Colombia. *Journal of Agricultural Science and Technology*, 1, 101–107.

Raftery, A., Hoeting, J., Volinsky, C., Painter, I., Yeung, K.Y., Sevcikova, M.H. (2025). Package 'BMA'.

Rahman, M.M., Robson, A. (2020). Integrating Landsat-8 and Sentinel-2 Time Series Data for Yield Prediction of Sugarcane Crops at the Block Level. *Remote Sensing*, 12(8), 1313. **https://doi.org/doi:10.3390/rs12081313**

Ramayah, T., Ahmad, N.H., Halim, H.A., Rohaida, S., Zainal, M., Lo, M. (2010). Discriminant analysis: An illustrated example. *African Journal of Business Management*, 4(9), 1654–1667.

Razaque, A., Ben Haj Frej, M., Almi'ani, M., Alotaibi, M., Alotaibi, B. (2021). Improved support vector machine enabled radial basis function and linear variants for remote sensing image classification. *Sensors*, 21(13), 1–26. **https://doi.org/10.3390/s21134431**

Rouse, J.W., Hass, R.H., Schell, J.A., Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. Third Earth Resources Technology Satellite (ERTS) Symposium, 351, 309–317. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740022614.pdf

Rwanga, S.S., Ndambuki, J.M. (2017). Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS. *International Journal of Geosciences*, 08(04), 611–622. **https://doi.org/10.4236/ijg.2017.84033**

Saini, R., Ghosh, S.K. (2018). Crop classification on single date Sentinel-2 imagery using Random Forest and Support Vector Machine. In Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-5, 683–688, **https://doi.org/10.5194/isprs-archives-XLII-5-683-2018**

Schulthess, U., Rodrigues, F., Taymans, M., Bellemans, N., Bontemps, S., Ortiz-Monasterio, I., Gérard, B., Defourny, P. (2023). Optimal Sample Size and Composition for Crop Classification with Sen2-Agri's Random Forest Classifier. Remote Sens, 15(3), 608. **https://doi.org/10.3390/rs15030608**

Serra, P., Pons, X., Saurí, D. (2003). Post-classification change detection with data from different sensors: Some accuracy considerations. *International Journal of Remote Sensing*, 24(16), 3311–3340. **https://doi.org/10.1080/0143116021000021189**

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6308–6325. **https://doi.org/10.1109/JSTARS.2020.3026724**

Solano, J., Villalobos, R. (2001). Aspectos fisiográficos aplicados a un bosquejo de regionalización geográfico climático de Costa Rica. *Top Meteor Oceanog*, 8, 26–39.

Som-ard, J., Atzberger, C., Izquierdo-Verdiguier, E., Vuolo, F., Immitzer, M. (2021). Remote sensing applications in sugarcane cultivation: A review. Remote Sensing, 13(20), 1–46. **https://doi.org/10.3390/rs13204040**

Suykens, J.A.K., Vandewalle, J. (1999). Least Squares Support Vector Machine Classifiers. *Neural Processing Letters*, 9, 293-300. **https://doi.org/10.1023/A:1018628609742**

USGS, (United States Geological Survey). (2022). Landsat 8-9 Collection 2 (C2) Level 2 Science Product (L2SP) Accessed 06-06-2023. Available at https://www.usgs.gov/media/files/landsat-8-9-collection-2-level-2-science-product-guide

Vignola, R., Poveda, K., Watler, W., Vargas, A., Berrocal, Á. (2018). Cultivo de Caña de Azúcar en Costa Rica. Accessed 06-06-2023. Available at https://www.mag.go.cr/bibliotecavirtual/F01-8327.pdf

Wang, X., Zhang, F., Kung, H.T, Johnson, V.C., Latif, A. (2020). Extracting soil salinization information with a fractional-order filtering algorithm and grid-search support vector machine (GS-SVM) model. *International Journal of Remote Sensing*, 41(3), 953–973. **https://doi.org/10.1080/01431161.2019.1654142**

Appendix 1. Literature review of sugarcane variety classifications from remote sensing.

Reference	Evaluated varieties	Method	RS source	Used variables	OA
Apan *et al*. (2004)	8 (Q121, Q124, Q136, Q138, Q185, Q190, Q195, and Q20)	Discriminant analysis.	EO-1 Hyperion	152 bands and 40 vegetation indices.	74.0 %
Galvao *et al.* (2005)	5 (RB72-454, SP80-1816, SP80-1842, SP81-3250, and SP87-365)	Discriminant analysis.	EO-1 Hyperion	146 bands (10 to 57, 79 to 115, 135 to 163, and 185 to 224). All ratios of reflectance between the selected bands. 13 vegetation indices.	87.5 %
Fortes and Demattê (2006)	4 (RB835486, RB855536, RB855113, and SP81-3250)	Analysis of individual bands, pixel dispersion plots, and discriminating equations.	Landsat-7	6 bands (B1, B2, B3, B4, B5, and B7). 6 vegetation indices (NDVI, GVI, SAVI, RVI, RATIO, and GNDVI)	93.6 %
Everingham *et al.* (2007)	9 (20, 121, 124, 135, 136, 138, 159, 185, and 190)	Discriminant analysis, Support Vector Machine, and Random Forest.	EO-1 Hyperion	150 bands	85.0 %
Murillo-Sandoval *et al.* (2011)	2 (CC85-92 and CC84-75)	Jeffries-Matusita.	Landsat-7	6 bands (B1, B2, B3, B4, B5, and B7). 6 vegetation indices (RVI, NDVI, GNDVI, SAVI, ARVI, GVI). 2 principal components.	80.8 %
Duft *et al.* (2019)	25	Random Forest.	Sentinel-2	10 bands (B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12). 3 vegetation indices (NDBI, RENDWI and RENDVI)	86.0 %
Kai *et al.* (2022)	4 (RB867515, RB92579, RB966928, and RB988082)	k-Nearest Neigh-bours algorithm, Support Vector Machine, Random Forest, and Artificial Neural Network.	Sentinel-2	12 bands (B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12). 10 vegetation indices (NDVI, GNDVI, NDWI, NDCI, NDMI, EVI, NDRE, SAVI, SIPI, and CI). 10 bands combination.	99.5 %

Appendix 2. Results of variable selection using *MeanDecreaseGini* and Bayesian Model Averaging methods for sugarcane variety classification with Discriminant Analysis.

Varieties	Selection method	Selected variables	Kappa index	Overall accuracy	User accuracy	Producer accuracy
Laica 05-805	MeanDecreaseGini	**The bands** Green (July) and NIR (July). **The VIs** $NDII_1$ (September), $NDII_2$ (December), and EBI (January).	0.79	0.86	0.81	0.8
RB 86-7515					0.88	0.98
RB 98-710					0.89	0.82
Laica 05-805	Bayesian Model Averaging	Age and Sector. **The bands** Blue (November), Green (July), $SWIR_1$ (April and January), and $SWIR_2$ (October). **The VIs** GNDVI (October), EBI (August and June), and $NDII_2$ (May).	0.70	0.80	0.76	0.73
RB 86-7515					0.79	0.87
RB 98-710					0.86	0.81
Laica 05-805	MeanDecreaseGini	**The bands** Green (July), NIR (September), and $SWIR_1$ (December). **The VIs** $NDII_1$ (October), EBI (November), and SR (January).	0.70	0.77	0.68	0.68
RB 86-7515					0.72	0.91
RB 98-710					0.84	0.76
RB 99-381					0.84	0.74
Laica 05-805	Bayesian Model Averaging	Age. **The bands** Blue (May), SWIR1 (April and January), and $SWIR_2$ (January). **The VIs** NDVI (June), EBI (July), and $NDII_1$ (October).	0.64	0.73	0.73	0.72
RB 86-7515					0.73	0.76
RB 98-710					0.81	0.72
RB 99-381					0.65	0.73
Laica 05-805	MeanDecreaseGini	Age. **The bands** Green (July and November) and NIR (January). **The VIs** $NDII_2$ (September) and NDVI (December).	0.56	0.64	0.47	0.48
RB 86-7515					0.84	0.74
RB 98-710					0.7	0.81
RB 99-381					0.84	0.76
Laica 07-801					0.37	0.4
Laica 05-805	Bayesian Model Averaging	Age and Sector. **The bands** Blue (December), Red (September), $SWIR_1$ (October), and $SWIR_2$ (August and January). **The VIs** GNDVI (May), NDVI (June), SR (April, July, and November), and $NDII_2$ (September).	0.62	0.70	0.57	0.69
RB 86-7515					0.76	0.76
RB 98-710					0.72	0.82
RB 99-381					0.77	0.68
Laica 07-801					0.69	0.59

Laica 05-805	MeanDecreaseGini	Age. **The bands** NIR (August), Blue (September), $SWIR_1$ (November), and Green (December). **The VIs** SR (June) and EBI (November).	0.51	0.59	0.40	0.58
RB 86-7515					0.77	0.71
RB 98-710					0.78	0.71
RB 99-381					0.71	0.59
Laica 07-801					0.65	0.54
Laica 04-809					0.25	0.33
Laica 05-805	Bayesian Model Averaging	Age and Sector. **The bands** Blue (June), Green (December), $SWIR_1$ (April and January), and SWIR2 (August). **The VIs** GNDVI (June), NDVI (October and December), EBI (September and October), $NDII_1$ (August), and $NDII_2$ (May).	0.55	0.63	0.36	0.5
RB 86-7515					0.65	0.71
RB 98-710					0.5	0.52
RB 99-381					0.69	0.84
Laica 07-801					0.84	0.76
Laica 04-809					0.71	0.45

Appendix 3. Summary of the best variables for variety classification groups.

Variety groups	**Age**	**Blue** Sep	**Green** Jul	**Green** Nov	**Green** Dec	**NIR** Jul	**NIR** Aug	**NIR** Sept	**NIR** Jan	**SWIR$_1$** Nov	**SWIR$_1$** Dec	**EBI** Nov	**EBI** Jan	**NDVI** Dec	**SR** Jun	**SR** Jan	**NDII$_1$** Sep	**NDII$_1$** Oct	**NDII$_2$** Dec
Laica 05-805 RB 86-7515 RB 98-710			X			X							X				X		X
Laica 05-805 RB 86-7515 RB 98-710 RB 99-381			X					X			X	X				X		X	
Laica 05-805 RB 86-7515 RB 98-710 RB 99-381 Laica 07-801	X		X	X					X					X			X		
Laica 05-805 RB 86-7515 RB 98-710 RB 99-381 Laica 07-801 Laica 04-809	X	X			X		X			X		X			X				