







Article

Evaluating Sugarcane Yield Estimation in Thailand Using Multi-Temporal Sentinel-2 and Landsat Data Together with Machine-Learning Algorithms

Jaturong Som-ard ^{1,2,*} , Savittri Ratanopad Suwanlee ^{1,2} , Dusadee Pinasu ³, Surasak Keawsomsee ^{1,2} , Kemin Kasa ^{1,2}, Nattawut Seesanhao ^{1,2}, Sarawut Ninsawat ⁴ , Enrico Borgogno-Mondino ⁵  and Filippo Sarvia ^{5,6,*} 

- ¹ Department of Geography, Faculty of Humanities and Social Sciences, Maharakham University, Maha Sarakham 44150, Thailand; savittri.s@msu.ac.th (S.R.S.); surasak.ke@msu.ac.th (S.K.); kemin.k@msu.ac.th (K.K.); 63010117034@msu.ac.th (N.S.)
 - ² Earth Observation Technologies for Land and Agricultural Development Research Unit, Faculty of Humanities and Social Sciences, Maharakham University, Maha Sarakham 44150, Thailand
 - ³ Technology and Informatics Institute for Sustainability, National Metal and Materials Technology Center, National Science and Technology Development Agency, Thailand Science Park, Pathum Thani 12120, Thailand; dusadee.pin@ncr.nstda.or.th
 - ⁴ Remote Sensing and GIS, School of Engineering and Technology, Asian Institute of Technology, Klong Luang, Pathum Thani 12120, Thailand; sarawutn@ait.ac.th
 - ⁵ Department of Agricultural, Forest and Food Sciences, University of Turin, Grugliasco L.go, 10095 Braccini, Italy; enrico.borgogno@unito.it
 - ⁶ Food and Agriculture Organization of the United Nations, 00153 Rome, Italy
- * Correspondence: jaturong.s@msu.ac.th (J.S.-a.); filippo.sarvia@unito.it (F.S.)



Citation: Som-ard, J.; Suwanlee, S.R.; Pinasu, D.; Keawsomsee, S.; Kasa, K.; Seesanhao, N.; Ninsawat, S.; Borgogno-Mondino, E.; Sarvia, F. Evaluating Sugarcane Yield Estimation in Thailand Using Multi-Temporal Sentinel-2 and Landsat Data Together with Machine-Learning Algorithms. *Land* **2024**, *13*, 1481. <https://doi.org/10.3390/land13091481>

Academic Editors: Zeyuan Qiu and Subhasis Giri

Received: 3 July 2024

Revised: 5 September 2024

Accepted: 10 September 2024

Published: 13 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Updated and accurate crop yield maps play a key role in the agricultural environment. Their application enables the support for sustainable agricultural practices and the formulation of effective strategies to mitigate the impacts of climate change. Farmers can apply the maps to gain an overview of the yield variability, improving farm management practices and optimizing inputs to increase productivity and sustainability such as fertilizers. Earth observation (EO) data make it possible to map crop yield estimations over large areas, although this will remain challenging for specific crops such as sugarcane. Yield data collection is an expensive and time-consuming practice that often limits the number of samples collected. In this study, the sugarcane yield estimation based on a small number of training datasets within smallholder crop systems in the Tha Khan Tho District, Thailand for the year 2022 was assessed. Specifically, multi-temporal satellite datasets from multiple sensors, including Sentinel-2 and Landsat 8/9, were involved. Moreover, in order to generate the sugarcane yield estimation maps, only 75 sampling plots were selected and surveyed to provide training and validation data for several powerful machine-learning algorithms, including multiple linear regression (MLR), stepwise multiple regression (SMR), partial least squares regression (PLS), random forest regression (RFR), and support vector regression (SVR). Among these algorithms, the RFR model demonstrated outstanding performance, yielding an excellent result compared to existing techniques, achieving an R-squared (R^2) value of 0.79 and a root mean square error (RMSE) of 3.93 t/ha (per 10 m × 10 m pixel). Furthermore, the mapped yields across the region closely aligned with the official statistical data from the Office of the Cane and Sugar Board (with a range value of 36,000 ton). Finally, the sugarcane yield estimation model was applied to over 2100 sugarcane fields in order to provide an overview of the current state of the yield and total production in the area. In this work, the different yield rates at the field level were highlighted, providing a powerful workflow for mapping sugarcane yields across large regions, supporting sugarcane crop management and facilitating decision-making processes.

Keywords: sugarcane; yield estimation; earth observation; Sentinel-2; Landsat; machine learning

1. Introduction

Accurate crop yield maps at the regional and global scale are an important decision-making tool in the precision agriculture context [1]. Field-scale yield maps provide valuable information to farmers about yield variations attributed to factors such as fertilizers, chemicals, water management, and other agricultural inputs [2]. To carry out an analysis to assess the crop yield, and biomass- and canopy-related information, the engagement of experienced agronomists is recommended. This procedure is time consuming and plays a key role during the map accuracy assessment [3]. In this context, it is still challenging for the stockholders (i.e., farmers, government, and mills) to develop cost-effective alternatives for mapping the accurate crop yield using remote sensing techniques and earth observation (EO) data [4–7].

Thailand is the world's third-largest sugarcane exporter in 2022, allowing the country to experience good economic growth and making this crop significant for the economy and agriculture sectors [8]. In recent years, national sugarcane production and productivity have increased due to the government's and sugar mills' policy support [9]. Moreover, the increase in production was also driven by the introduction of the cultivation of specific sugar cane varieties known for their high yields [10]. Sugarcane is an important crop, serving as a primary raw material for sugar production that can be used as a feedstock for ethanol production, molasses, electricity, and pulp [11]. According to the Office of the Cane and Sugar Board report [10], sugarcane cultivation areas increased in Thailand from 0.06 Mha in 1961 to 1.53 Mha in the year 2022, making sugarcane production one of the most important sources of income for farmers in the country [12].

Northeast Thailand is found to be the most suitable region covering the major cultivated sugarcane area of the country. Farmers in this area have followed agricultural policies to increase sugarcane production and productivity [11,13]. However, several problems occur during the agronomic year that affect the full yield of the crop. Specifically, in some cases, the main problems related to production losses are due to obsolete agricultural technology, a lack of labor, insufficient irrigation, crop field fires, and the impact of climate change [11,12]. Although most of these critical problems have been solved by the stakeholders [9,13], some of them still affect sustainable sugarcane management practices.

Therefore, an accurate and timely sugarcane yield map at the field scale is expected to support farmers, sugar mills, price setting, allocating quota systems, the market, and managing crop growth. Concerning the yield estimation at the field scale, recent studies from Som-ard et al. [13]; Suwanlee et al. [14]; Canata et al. [15]; and Shendryk et al. [16] highlight the idea that EO technology can be a key tool for mapping and monitoring field spatial variability, especially in areas with low production potential.

EO-based satellites are well known for providing timely agricultural crop information over large areas and fulfilling the requirements of sugar mills and governments worldwide [11,15]. Currently, one of the main advantages of EO data based on high spatial-temporal resolutions is the ability to assess and monitor crop phenological development and crop health. For the last two decades, the sugarcane crop yield estimation was assessed by combining remote sensing techniques with EO data, frequently resulting in maps with a rough spatial resolution and, in some cases, not enough accuracy to make these maps useful for decision making within the agricultural context [11,17–19]. Currently, several studies have demonstrated the high potential for mapping crop yields through the joint use of EO data from multiple sensors, such as Sentinel-1 (S1), Sentinel-2 (S2), and Landsat with powerful machine-learning-based classification approaches [14–16,20]. The efficiency in combining the dense time series of EO from multi-sensor data was highlighted in order to obtain a sugarcane yield map estimation in traditional smallholder crop systems and complex landscapes.

Additionally, recent studies have also demonstrated that sugarcane yield estimation can be improved by combining EO data from different multiple sensors and machine-learning methods (i.e., random forest regression (RFR), support vector regression (SVR), and linear regression). For instance, Rahman and Robson [20] estimated the sugarcane

yields by utilizing a time series of S2 and Landsat-8 (L8) data within the Bundaberg region (southeastern Queensland, Australia), obtaining an R^2 equal to 0.87 and RMSE equal to 11.33 t/ha. Canata et al. in [15] mapped and estimated sugarcane yield in the Botucatu municipality, São Paulo state, Brazil, using multi-temporal S2 data, and they obtained accuracy values with an R^2 of 0.70 and RMSE of 4.63 t/ha. Shendryk et al. [16] analyzed sugarcane crop yield in the northeast Queensland, Australia, using multi-temporal S1 and S2 data, achieving an R^2 up to 0.80 and an RMSE of 16 t/ha. Abeb et al. [21] mapped sugarcane yield southeast of Addis Ababa, Ethiopia, using multi-temporal S2 and L8 data, obtaining an R^2 of 0.85 and an RMSE of 12.95 t/ha. Dimov et al. [22] generated a sugarcane yield map southeast of Lake Basaka in Metahara, Ethiopia, using S2 time series, obtaining an R^2 value of 0.84 and an RMSE of 20.40 t/ha. Som-ard et al. [13] assessed sugarcane yield in Udon Thani province, Thailand, using multi-temporal S2 data, obtaining an R^2 of 0.84 and an RMSE equal to 6.88 t/ha. Finally, Suwanlee et al. [14] accurately estimated sugarcane biomass in Udon Thani province, Thailand, using S1 and S2 time-series datasets, achieving high accuracy with an R^2 equal to 0.86 and an RMSE of 9.61 t/ha. Their results demonstrated high accuracies and excellent estimates of sugarcane yield using multi-temporal data from multiple sensors across different geographic regions.

However, in small fields (<1 ha), estimating sugarcane yield at the field scale using multi-temporal data from multiple sensors remains a challenge. Furthermore, recently, Som-Ard et al. [11] mentioned that numerous machine-learning methods still need to be determined for mapping sugarcane yield, particularly with small field sizes and cloudy regions, such as the Thailand context.

Accordingly, to address these issues, the main aims of this work are outlined below:

- (i) To evaluate the best machine-learning method for sugarcane yield estimation (e.g., multiple linear regression (MLR), stepwise multiple regression (SMR), partial least squares regression (PLS), RFR and SVR);
- (ii) To apply the identified algorithm using joint multi-temporal S2 and L8/9 data to map the estimated sugarcane yield in 2022 within the Tha Khan Tho district of Kalasin province.

An accurate and timely sugarcane yield map from this study provides valuable information for sustainable management practices and environmental considerations. It will enable farmers to estimate their yield before harvest in different years and geographical regions, helping to enhance their production process before the agricultural season ends.

2. Materials and Methods

The implemented methodology was carried out according to the workflow reported in Figure 1. In particular, three main phases were defined: (i) field data collection and image preprocessing; (ii) predictive models' evaluation; and (iii) sugarcane yield map generation using the best machine-learning predictive model identified for the entire Tha Khan Tho District, Kalasin Province.

2.1. Study Area

The study area is located within the Tha Khan Tho District, Kalasin Province, Thailand, spanning from 16°47'30'' to 16°57'30'' N and 103°5'00'' to 103°25'00'' E, covering approximately 39,360 ha (Figure 2). The area is primarily characterized by flat terrain with undulating valleys. The soil is mostly sandy loam or brown-gray loam, particularly in the low-lying areas. The climate is tropical semi-humid dry-savanna (Köppen climate, classification: Aw), with major seasons including the summer (mid-February to mid-May), rainy season (mid-May to mid-October), and winter (mid-October to mid-February). The average temperature is 27 °C, and the average annual precipitation is about 1368.4 mm, making the area highly suitable for sugarcane cultivation and plantations.

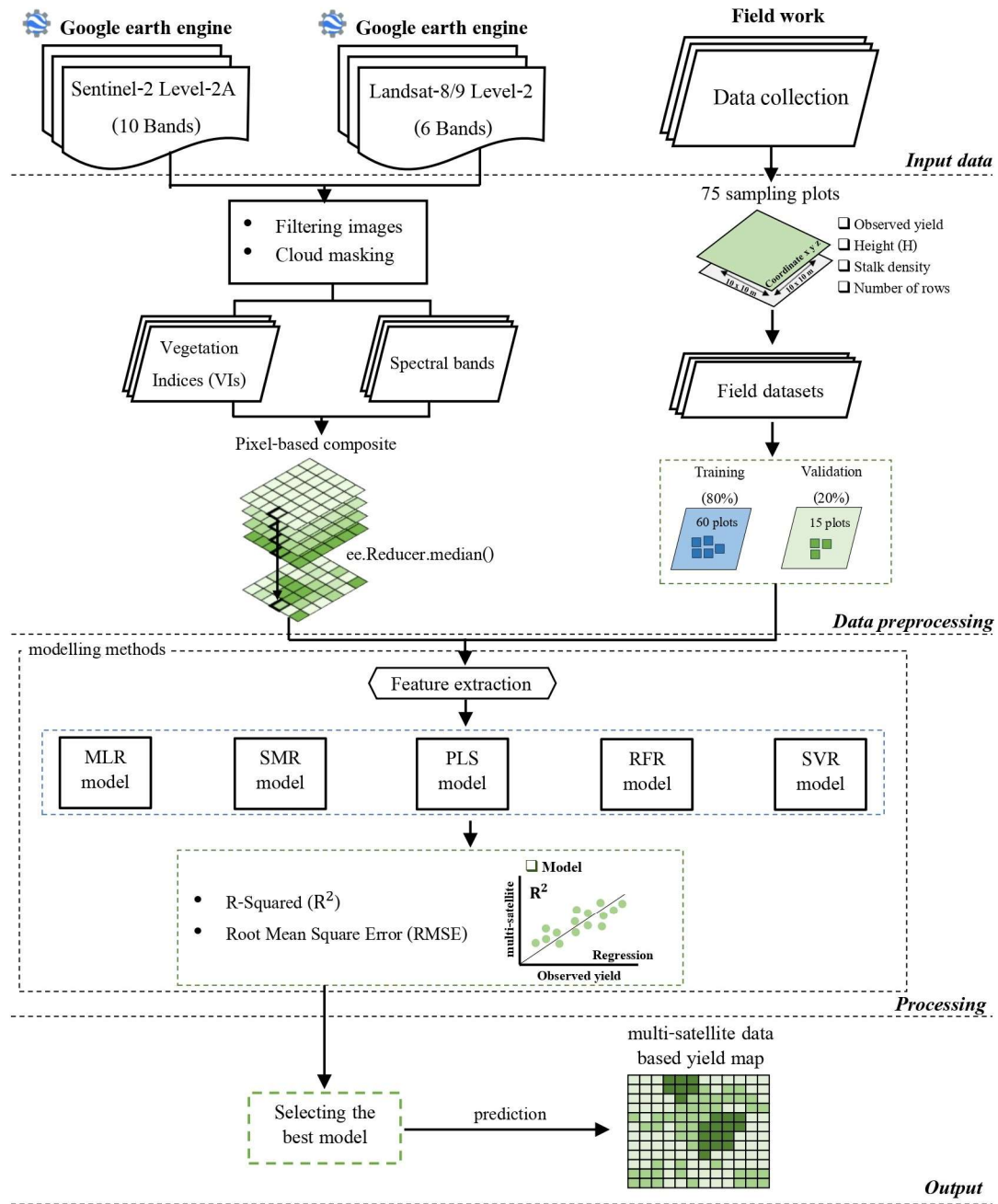


Figure 1. Flow chart of the implemented methodology for mapping sugarcane yield in 2022 at Tha Khan Tho District, Thailand using multi-temporal Sentinel-2 (S2) and Landsat 8/9 (L8/9) dataset together with the several machine-learning methods.

Although farmers in the area cultivate various crops, including cassava, rubber and mango, sugarcane has shown a continuous annual increase, holding significant economic value. This growth could be attributed to government policy that promoted the cultivation of crops well suited to the region. Moreover, it is worth to highlight that sugarcane development and cultivation follow two main seasons in this area: (i) the sugarcane crop rainy season from May to July; and (ii) the sugarcane crop summer season from November to February.

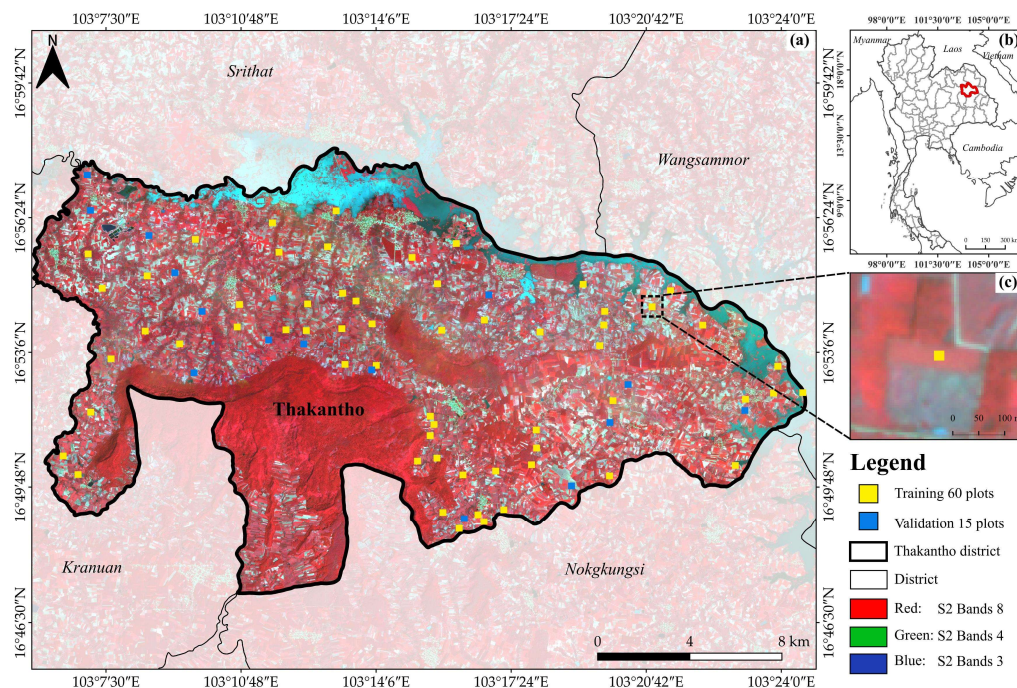


Figure 2. Study area (a): background shows Sentinel-2 (S2) imagery (image composites: during November 2022) with false color (Red = band 8: Green = band 4: Blue = band 3). The 60 yellow sampling plots are used for training datasets and remaining 15 blue plots were used for validating the mapped results. (b) is a location of the Tha Khan Tho District, Kalasin Province, Thailand (study region), (c) shows sampling plot with size of 10 m × 10 m.

2.2. Field Data Collection

A field data campaign was conducted from November 20 to December 15, 2022 in order to select, collect and map a total of 75 sugarcane plantations for training and validating the proposed machine-learning predictive model. Specifically, the actual yield (t/ha) was measured and collected within a 100 m² area (10 m × 10 m). Field selection was based on a random sampling criterion to cover the entire area of interest. During the field survey, plot locations were recorded using a UniStrong G10 global navigation satellite system (GNSS) receiver from Beijing UniStrong Science & Technology company, Beijing, China, as shown in Figure 2. It is worth highlighting that the sugarcane cultivar Khonkaen-3 was found in all assessed fields [13]. Considering the crop management, sugarcane cultivation is characterized by having a row spacing between 1.2 and 1.5 m with inter-row spacing ranging from 0.2 to 0.4 m. Statistics from the field data are reported in Table 1. Specifically, for the 2022 year, sugarcane yield values ranged from 43.75 to 112.50 t/ha with a mean yield value of 96.00 t/ha and a median value of 100 t/ha. For use in the prediction model, the ground data were subsequently split into 60 sampling plots (80%) for training the models and 15 plots (20%) for validation.

Table 1. The summary descriptive statistics of the observed sugarcane yield datasets in the year 2022 in this study. This dataset was collected from field work between 20 November and 15 December 2022.

Year	n (10 m × 10 m)	Minimum (t/ha)	Median (t/ha)	Standard Deviation (t/ha)	Mean (t/ha)	Maximum (t/ha)
2022	75	43.75	100.00	12.45	96.00	112.50

Abbreviation: n is the number of plots.

Moreover, photo-interpretation work was conducted to identify all sugarcane fields within the study area. The goal was to map these fields for the application of yield models and to estimate sugarcane production. Specifically, this process utilized a high-resolution Planet image selected from November 2022. A total of 2364 sugarcane fields were visually identified and recorded (Figure 3). This dataset was subsequently cross-validated during the field survey.

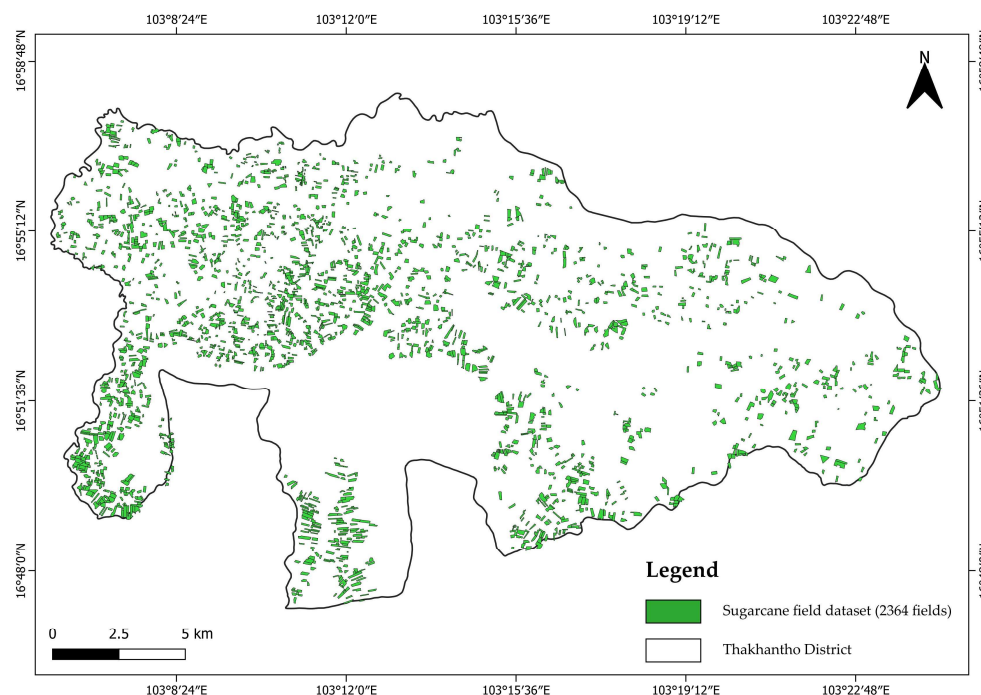


Figure 3. The sugarcane field dataset (2364 fields) was visually interpreted using very high-resolution imagery as Planet imagery during November 2022.

Finally, for 15 of the 75 sugarcane fields collected during the field data campaign, actual yield (t/ha) at the plot level was measured and provided by the farmer after the harvest between 15 and December 30, 2022. This dataset, which was recorded from the farmer using the satellite data from this study, was compared with the output from models calibrated with field surveys in order to assess and evaluate the accuracy.

2.3. Satellite Dataset

In this analysis, multi-temporal S2 and L8/9 data from October to December 2022 were combined and involved as input for the sugarcane yield estimation. This time period was selected according to the phenological development of sugarcane crops in northeast Thailand, where peak growth typically occurs between October and December [11].

2.3.1. Sentinel-2 Data

Sentinel-2 Level-2A images acquired from October to December 2022 were downloaded from the Google Earth Engine (GEE) platform [23]. A total of 49 scenes were selected after applying a cloud cover filter to remove images with more than 40% cloud cover. Spectral bands with spatial resolutions of 10 m (B2, B3, B4 and B8) and 20 m (B5, B6, B7, B8a, B11 and B12) were selected. Moreover, to minimize the impact of meteorological issues, the Scene Classification Layer (SCL) and Quality Indicators (QIs) were taken into consideration to remove pixels affected by cloud and snow conditions. Furthermore, all 20 m resolution spectral bands from S2 and L8/9 (B5, B6, B7, B8a, B11 and B12) were resampled to 10 m using the nearest-neighbor (NN) method [24].

2.3.2. Landsat Data

Landsat 8/9 (L8/9) Level 2, Collection 2, Tier 1 products were acquired from October to December 2022. The images were preprocessed through the GEE using the same steps applied to the S2 dataset [25]. Therefore, a total of 7 scenes with less than 40% cloud cover were selected. The L8/9 dataset has a geometric resolution of 30 m, and the bands B2, B3, B4, B5, B6, and B7 were selected and resampled to match the 10 m resolution of S2 bands.

2.3.3. Vegetation Index Generation

From the S2 and L8/9 datasets, several vegetation indices (VIs) that characterize sugarcane phenological dynamics were computed. Specifically, the following VI were used: normalized difference vegetation index (NDVI) [26], normalized difference water index (NDWI) [27], enhanced vegetation index (EVI) [28], normalized difference infrared index (NDII) [29], green normalized difference vegetation index (GNDVI) [30], soil-adjusted vegetation index (SAVI) [31], ratio vegetation index (RVI) [32] and moisture index (MSI) [33,34]. These VIs served as additional features for monitoring and mapping sugarcane. Further details on VIs computation are provided in Table 2.

Table 2. The derived vegetation indices (VIs) of Sentinel-2 (S2) and Landsat 8/9 (L8/9) data used for sugarcane yield estimation and formula with references.

Index	Formula	Reference
Normalized Difference Vegetation Index (NDVI)	$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}}$	[26]
Maximum Normalized Difference Vegetation Index (MaxNDVI)	$\text{MaxNDVI} = \max\{\text{NDVI}(i,j)_1, \text{NDVI}(i,j)_2, \dots, \text{NDVI}(i,j)_n\}$	[35]
Green Normalized Difference Vegetation Index (GNDVI)	$\text{GNDVI} = \frac{\text{NIR} - \text{GREEN}}{\text{NIR} + \text{GREEN}}$	[30]
Normalized Difference Water Index (NDWI)	$\text{NDWI} = \frac{\text{GREEN} - \text{NIR}}{\text{GREEN} + \text{NIR}}$	[27]
Enhanced Vegetation Index (EVI)	$\text{EVI} = 2.5 \frac{\text{NIR} - \text{RED}}{(\text{NIR} + 6\text{RED} - 7.5\text{BLUE}) + 1}$	[28]
Normalized Difference Infrared Index (NDII)	$\text{NDII} = \frac{\text{NIR} - \text{SWIR1}}{\text{NIR} + \text{SWIR1}}$	[29]
Soil-Adjusted Vegetation Index (SAVI)	$\text{SAVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED} + L} (L + 1)$	[31]
Ratio Vegetation Index (RVI)	$\text{RVI} = \frac{\text{NIR}}{\text{RED}}$	[32]
Moisture Index (MSI)	$\text{MSI} = \frac{\text{SWIR}}{\text{NIR}}$	[34]

2.4. The Median Compositing Approach

Although cloudy images were filtered during the image selection process, the median compositing (MC) method was applied in order to ensure the quality of the images involved for sugarcane yield estimation. This method reduced pixel uncertainty by generating gap-free image composites, ensuring spatio-temporal consistency. It is well regarded for its effectiveness in generating cloud-free and consistent phenological data across large areas [36,37]. Furthermore, this approach reduces the number of images needed for classification, enhancing image quality and shortening classification time. Therefore, spectral bands and VIs were composited for each month, and the final images were visually interpreted by comparing them with Google Earth Pro products to ensure high quality.

In addition, the maximum NDVI composite (MaxNDVI) was computed by selecting the pixel with the highest NDVI value from multi-temporal NDVI datasets of S2 and L8/9 images (see Table 2). This method mitigates issue related to clouds and poor environmental conditions, and it is known as an effective parameter for monitoring vegetation especially in areas affected by high cloud cover [35,38].

2.5. Sugarcane Yield Estimation

Sugarcane yield estimation is a well-explored topic in the literature (Som-Ard, Atzberger, Izquierdo-Verdiguier, Vuolo and Immitzer [11]; de França e Silva, Chaves, Luciano, Sanches, de Almeida and Adami [19]). However, the accuracy achieved to date often requires improvement before advancing technology transfer despite the availability of advanced machine-learning algorithms. To fill this gap, this study employed five different machine-learning algorithms: multiple linear regression (MLR), stepwise multiple regression (SMR), partial least squares regression (PLS), random forest regression (RFR), and support vector regression (SVR). These algorithms were trained with the training data, and their performance was evaluated using the validation set with R software version 4.2.2. Details of the models used in this work are as follows:

- Multiple Linear Regression (MLR) is a mathematical algorithm that leverages multiple independent variables to derive a predictive value. The MLR model is widely used in several applications across scientific disciplines [39,40]. Their adaptability and robust theoretical underpinnings make MLR an invaluable tool for researchers and practitioners seeking accurate predictions in various scientific contexts (see the details in Formula (1) below). To fit the model, we implemented the “leaps” package in R.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n \quad (1)$$

where y represents the estimated sugarcane yield in this study, where β_0 to β_n denote the unknown linear regression coefficients, and X_1 to X_n represent predictor variables.

- Stepwise Multiple Regression (SMR) is a multiple linear model designed to incorporate or exclude variables from the regression equation based on their impact on the dependent variable. This facilitates examining linear relationships among multiple variables [41]. Prior investigations have indicated the superiority of the stepwise Akaike Information Criterion (AIC) method for variable selection compared to other stepwise techniques. Consequently, our strategy involved identifying crucial variables by all input datasets, which was guided by the AIC principle [42]. This work utilized the ‘stepwise’ package in R for the implementation, ensuring a comprehensive exploration of variable selection in the MLR model.
- Partial Least Squares Regression (PLSR) is an approach that integrate the principles of principal components analysis (PCA) and multiple linear regression and identifies components from the explanatory (X) variables that are pertinent to the response (Y) variables, reducing the dimensionality of the explanatory variables [43,44]. PLSR solves significant problems such as overfitting, multicollinearity, and outliers, simultaneously identifying factors that elucidate a substantial portion of the information shared between all ‘X’ and ‘Y’ variables [43,45,46]. Specifically, the PLSR process utilizes ‘X’ variables to predict ‘Y’. Incorporating all variables in constructing PLSR factors does not specify which variable or group of variables comparatively contributes more to predicting the ‘Y’ variable [47]. We implemented the PLSR model using the PLS library in the R software [48].
- Random Forest Regression (RFR) is a machine-learning technique introduced by Breiman [49] that leverages multiple decision trees for the classification and regression method. It is renowned for its robust performance, particularly in handling noise in training data and addressing nonlinear problems. RFR involves optimizing two key parameters during model construction: the number of trees (*n_{tree}*) and the randomly chosen predictor for each node in the binary tree (*m_{try}*). The final predictors are derived by averaging the outputs of all regression trees [49]. Our implementation of the RFR model utilized the ‘randomForest’ package in R. Moreover, the RFR model was optimized through a ‘Grid Search’ process to find the best parameter settings. Specifically, the parameters of trees were tuned from 100 to 1500 (at an interval of 50) with entry from 1 to 10 (at the step of 1). The result fitted the best parameters of 500 *trees* and 3 *mtry*, resulting in the most excellent hyperparameters.

- Support Vector Regression (SVR) is a kernel-based machine-learning approach proposed by Cortes and Vapnik [50] and further developed by Vapnik [51]. This method typically utilizes a kernel function to transform training datasets into a high-dimensional feature space, enabling the identification of an optimal hyperplane that maximizes the margin between the hyperplane and the nearest positive and negative reference datasets. For the kernel function, we used the radial basis function (RBF). SVR requires two parameters, including cost (c) and γ values, to determine the most effective configuration for the SVR model. Similar for the RFR model, also in this case, the SVR hyperparameter were optimized. In particular, the c value was set from -1 to 5 (at the interval of 1), and many γ values were set from 0.1 to 5 (at the step of 1). This analysis resulted in the final optimization of the hyperparameter for the model with $c = 1$ and $\gamma = 0.1$. Implementation of the SVR model was accomplished using the 'e1071' package in R.

2.5.1. Important Feature Selection

It is important to highlight that the features used during the training phase play a key role in the classification process. Depending on the information they contribute to the model, they may have different weight in achieving the result. Understanding these weights helps identify the most important variables or those providing the most valuable information. This approach allows for the selection of key variables, reducing the number of variables used during the classification process and decreasing processing time. Consequently, to enhance the efficiency of the yield prediction model, an analysis of the weights of the 234 variables was conducted. This analysis included all spectral information bands and VIs from S2 and L8/9. The feature selection process involves an elimination procedure algorithm, focusing on the "Increase of Mean Squared Error" (%IncMSE) based on a random forest approach. This method has proven effective in selecting important model features, aiming to reduce the number of features while maintaining their relevance and significance [13,15].

2.5.2. Sugarcane Yield Accuracy Evaluation

The performance of the optimal predictor models for mapping sugarcane yield was rigorously evaluated using the validation set based on the mean square error (MSE), the root mean square error (RMSE) and the coefficient of determination (R^2), as described by Equations (2), (3) and (4), respectively.

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}} \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{x}_i - \bar{x}_i)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (4)$$

where \hat{x}_i is the estimated yield, x_i is the observed yield for training datasets, \bar{x}_i is the mean value of observations, and n is the number of samples or number of observations.

The estimated yield at the field level was compared with the yield provided by the farmers after the 2022 harvesting season. Subsequently, the scatter plots for each of the five predictive models used in the sugarcane yield estimation were generated. These plots display the relationship between the observed yield (x-axis) and estimated yield (y-axis) for each sampling plot.

2.5.3. Sugarcane Yield Mapping

Sugarcane yield mapping allows to understand and manage the agronomic season, support farmers in enhancing production through specific agronomical practices, and

predict the final obtainable yield. Therefore, once the best machine-learning algorithm was identified (the one with the lowest error values), it was applied to estimate the potential sugarcane yield within the area of interest. Specifically, yield estimation was performed at the pixel level with a geometric resolution of 10 m across nearly 2300 sugarcane fields. Additionally, the frequency distribution of sugarcane production was computed, and a frequency histogram was generated to provide an overview of the yield within the study area.

3. Results

3.1. Ranking Important Features

In this work, starting from the RF model and the 234 input features, it was possible to identify and select the optimal features for sugarcane yield estimation. This approach encompasses all spectral information bands and VIs. Based on the RMSE (t/ha) values, the most important features were selected and used to predict sugarcane yield across the study region, as shown in Figure 4.

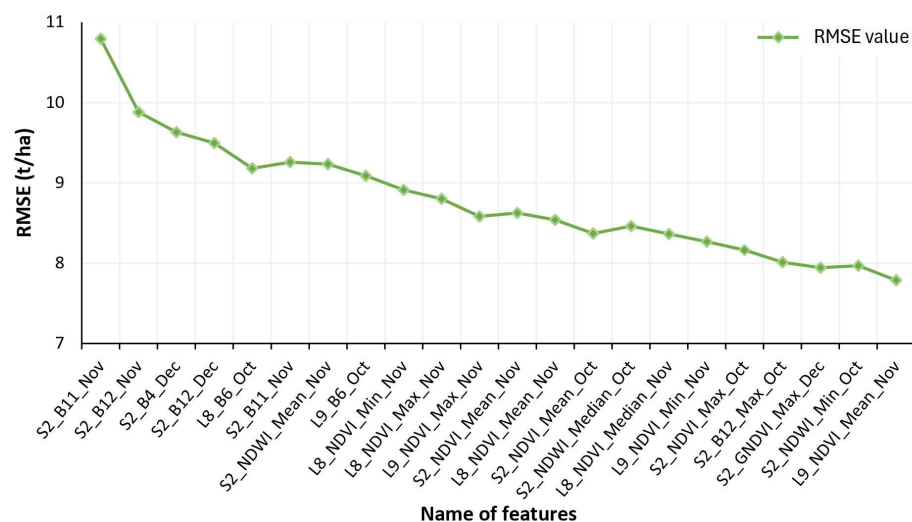


Figure 4. The ranking importance of the features using the random forest (RF) method for the year 2022 with sampling plots and the multi-temporal Sentinel-2 (S2) and Landsat 8/Landsat 9 (L8/9) data.

Considering Figure 4, it can be highlighted that among the first 10 selected features, the spectral bands and indices do not recur. This result suggests an absence of spectral monospecificity, indicating that sugarcane yield estimation involves several bands and spectral indices, including NDWI and NDVI.

Once the main satellite spectral features were defined, the training of all the algorithms involved could proceed. Subsequently, sugarcane yield maps were generated. Figure 5a–e shows the estimated yield maps (unit: t/ha at S2-10 m × 10 m) at the field scale in the zoom box for five predictive models.

Looking at Figure 5, it can be noted that sugarcane yields ranged between 59 and 108 t/ha. However, the different algorithms applied generated varying outcomes. Figure 5 shows that the MLR and SMR algorithms are characterized by high dispersion in sugarcane yield estimations with very productive plots marked by dark green (southern zone) and very unproductive plots marked by white (northern zone). In contrast, the SVR algorithm shows a moderate distribution of sugarcane yield densities with clustered bright green areas in the eastern zone at the field level. As expected, PLS and RFR exhibit the lowest variability in sugarcane yield estimation across different fields. Both methods provide a highly detailed spatial distribution of estimated sugarcane yields at the field scale, corresponding to the varying greenness of the sugarcane canopy, as captured by the S2 image compositing from November 2022.

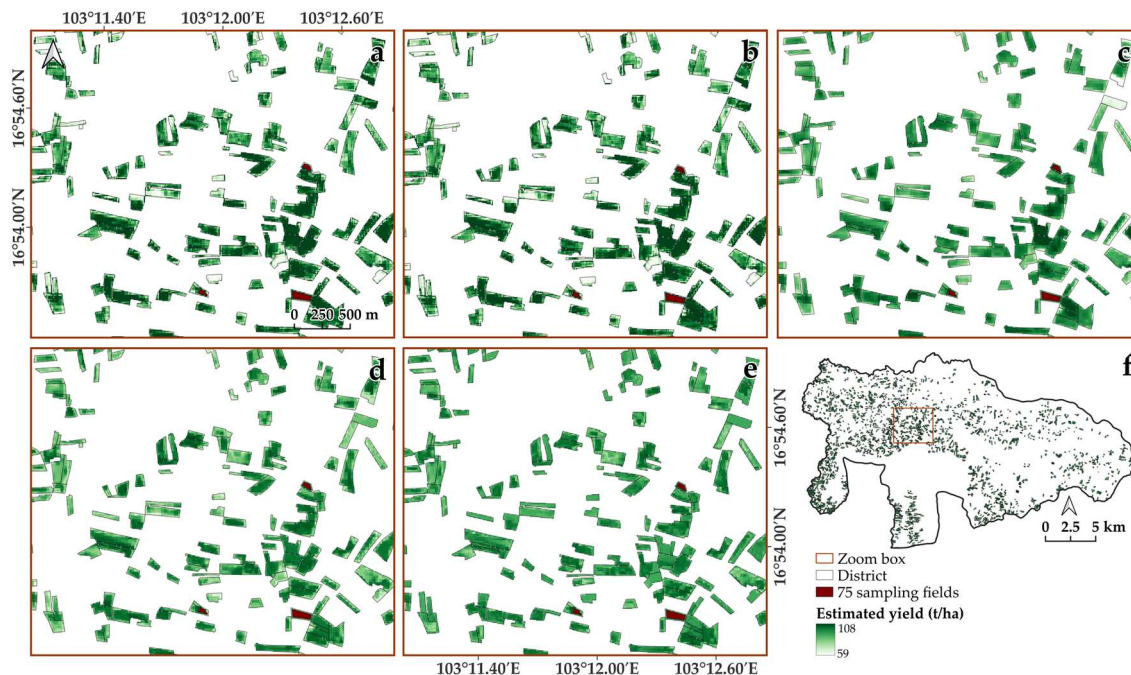


Figure 5. Zoom box of the estimated yield maps with yield value from 59 to 108 based on productive models: multiple linear regression (MLR) (a); stepwise multiple regression (SMR) (b); partial least squares regression (PLS) (c); random forest regression (RFR) (d); and support vector regression (SVR) (e). The sugarcane fields the entire study area (f) are shown.

3.2. Sugarcane Yield Modelling

Once the sugarcane yield estimation was mapped, the validation dataset was used to assess the quality of the models. A summary of the various model performances, based on R^2 , MSE, and RMSE values derived from an independent validation dataset, is presented in Table 3 and Figure 6.

Table 3. The summary of the coefficient of determination (R^2), the mean square error (MSE), and the root mean square error (RMSE) results from different machine-learning methods for estimating sugarcane yield in 2022 using independent validation dataset.

Statistics	Methods				
	MLR	SMR	PLS	RFR	SVR
R^2	0.40	0.46	0.64	0.79	0.72
MSE (t/ha) ²	173.41	159.75	23.71	15.42	17.99
RMSE (t/ha)	13.17	12.64	4.87	3.93	4.24

Considering the R^2 values in Table 3, it can be observed that they vary significantly with a minimum of 0.4 for MLR and a maximum of 0.79 for RFR. Similarly, the MSE values also show considerable variation. Specifically, the two linear regression models, MLR and SMR, exhibit high MSE values of 173 and 159, respectively. In contrast, the machine learning models achieved better MSE values, approximately 23, 15, and 18 for PLS, RFR, and SVR, respectively. A similar trend is seen in RMSE values, where MLR and SMR report values of 13, while PLS, RFR and SVR yield lower RMSE values of 5, 4 and 4, respectively.

Figure 6 shows the estimated yield results compared to the independent dataset based on 15 sampling plots in 2022 for all the models involved.

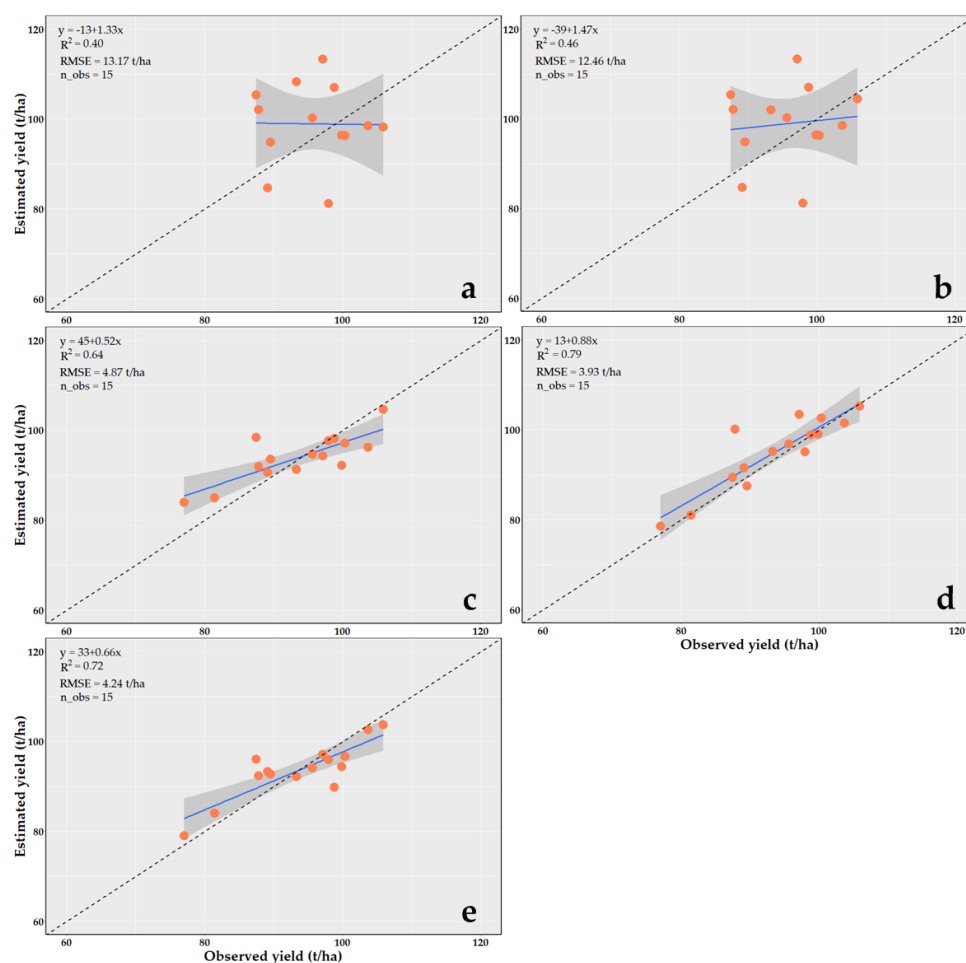


Figure 6. The scatter plots of sugarcane yield estimation results using five productive models together with multi-temporal Sentinel-2 (S2) and Landsat 8/9 (L8/9) data: multiple linear regression (MLR) (a); stepwise multiple regression (SMR) (b); partial least squares regression (PLS) (c); random forest regression (RFR) (d); and support vector regression (SVR) (e).

Considering Figure 6, it is possible to observe the results generated by the different models for sugarcane yield estimation. The scatter plots for the MLR and SMR models showed similar patterns with random dispersion. Meanwhile, the RFR and SVR models achieved a highly correlation between observed and estimated yield. According to Table 3 and Figure 6, the RFR model proved to be the best predictor, achieving high accuracy and efficiency (i.e., R^2 , MSE and RMSE equal to 0.79, 15.42 and 3.93, respectively) for mapping sugarcane yield. This demonstrates the overall performance and reliability of the RFR model, which is attributed to its flexibility and reduction of overfitting. Therefore, a comparison at the field level was made using this model to assess the differences between the estimated sugarcane yield and the yield collected by the farmer. A bar graph comparing the actual harvested yield (t/ha) and the model-predicted yield (t/ha) for individual fields in 2022, based on the RFR model, is presented in Figure 7.

Considering Figure 7, no significant differences can be observed when comparing the data from the 15 plots, indicating that the sugarcane yield predictions closely mirrored the actual yields. However, a slight overestimation by the model appears in all fields except for field 1. Specifically, the average overestimation is 3.79 t/ha, which corresponds to 6.67%, highlighting the model accuracy. In field 1, the actual sugarcane yield is 108.88 t/ha, compared to the RFR model's estimate of 107.11 t/ha, indicating an underestimation of about 1.77 t/ha. It is worth highlighting that this overestimation can be related to the model used for generating the results. Moreover, overlap with permanent vegetation pixels can

occur during the classification of sugarcane fields, leading to crop type misclassification and yield over–underestimation. In addition, pixels at the edges of fields may include other crop types, such as grassland, rubber trees or cassava, at different growth stages, which could artificially increase the area of the sugarcane fields.

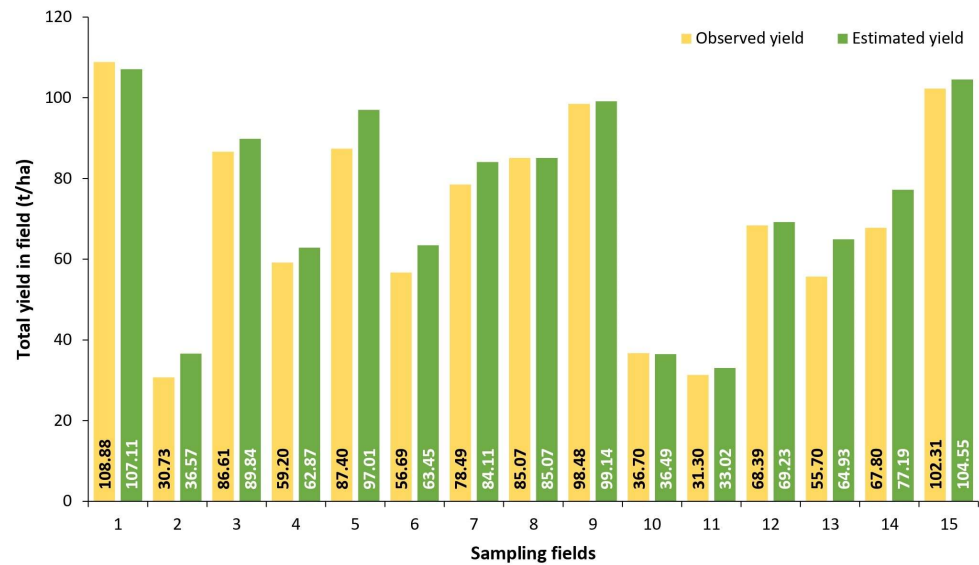


Figure 7. A comparison of observed yield (t/ha) and estimated yield (t/ha) of 15 sampling fields across the study area based on the best random forest regression (RFR)-predictive model together with Sentinel-2 (S2) and Landsat data.

3.3. Estimated Sugarcane Yield Maps

The optimized RFR model, which exhibited the highest performance in predicting sugarcane yield, was therefore used to generate a yield map for 2364 sugarcane fields. This map showcases the spatial distribution of different yield densities estimated for 2022 (Figure 8).

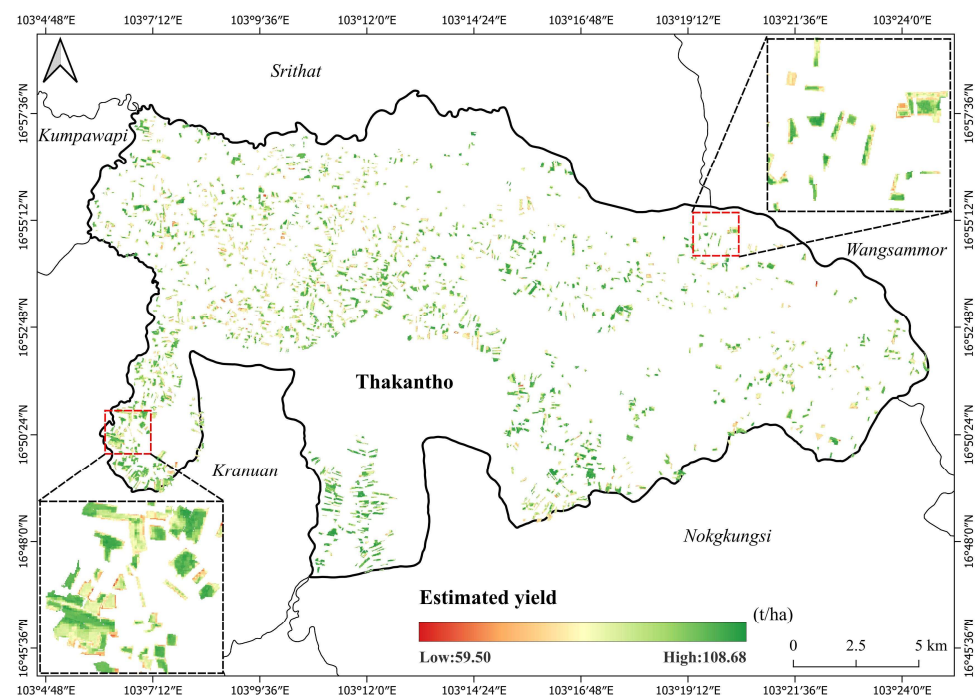


Figure 8. The spatial distribution of estimated yield in 2022 using the best random forest regression (RFR) together with Sentinel-2 (S2) and Landsat data.

Considering Figure 8, the yield spatial distribution exhibited significant heterogeneity, which was likely due to the varied yield fluctuations within the study area. This variation may be attributed to differences in crop management practices and the use of diverse inputs. Furthermore, in Figure 8, which represents the year 2022, the highest yield densities are primarily distributed in the central, west and southwestern regions of the study area. Sugarcane yield prediction serves as a valuable tool for farmers and stakeholders, enabling informed decisions regarding field management. By using map data to identify areas of high and low production, this information helps optimize resource allocation, ultimately increasing the overall yield and profitability in sugarcane cultivation.

Figure 9 shows the correspondent histogram, depicting the frequency distribution of estimated sugarcane yields across the study area for the 2022 harvest season.

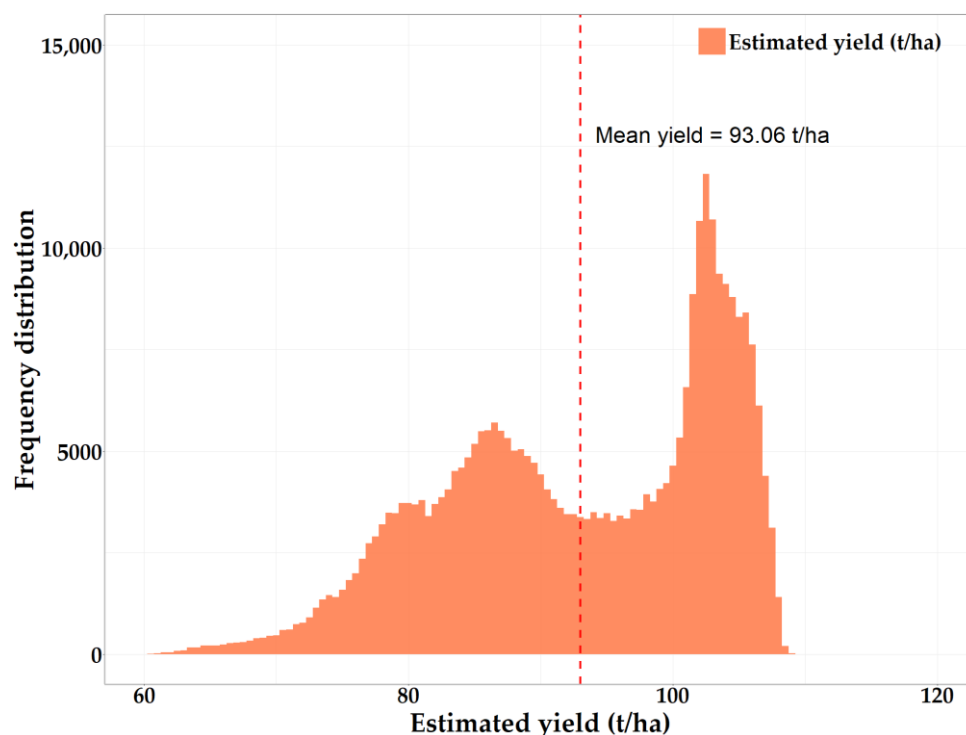


Figure 9. Histogram of the frequency distribution of estimated yield (t/ha) across the Tha Khan Tho District, Thailand, from 10 m × 10 m Sentinel-2 (S2). The red dotted line is the mean value of the estimated yield in this region.

Considering Figure 9, a range of yield estimates was found with values ranging from 59.50 to 108.68 t/ha. The average yield was 93.06 t/ha with a standard deviation (SD) of 10.25 t/ha. Based on the yield obtained and the area of the sugarcane fields, the total sugarcane production in the area was computed. The total mapped yield was 236606 t, while official statistical data report 210,327 t. Thus, the estimated yield across the region closely aligned with the official statistical data from the OCSB for 2022.

Moreover, it appears that two main types of sugarcane crops are present in the area. Specifically, two peaks are observed: one at a yield value of about 87 t/ha, and the second at around 102 t/ha, with a difference of approximately 15 t/ha. This variation is likely due to differences in agronomic management practices by farmers and is probably influenced by physical factors such as soil moisture and topography, which could lead to crop growth and yield densities at the field levels. In the first case, farmers have limited access to water and mechanization, making their operations less effective. On the contrary, in the second case, farmers have better access to mechanization and water, allowing them to achieve a significantly higher sugarcane yield: an increase of nearly 13%.

4. Discussion

4.1. Evaluated Performance of Yield Estimation

This study assessed the performance of various machine learning regression models, including MLR, SMR, PLS, RFR, and SVR algorithms, based on 75 sampling plots involved during the training steps. Using ground-based data alongside multi-temporal S2 and L8/L9 data, the predictive models demonstrated an accuracy with an R^2 greater than 0.40 and an RMSE less than 13.17 t/ha. Notably, the RFR model showed higher accuracy ($R^2 = 0.79$ and RMSE = 3.93 t/ha) compared to other models. The high performance of the models is attributed to the dense time-series data provided by the S2 and L8/L9 datasets [13,15,20]. Additionally, the RFR model proved to be more flexible in handling different datasets with limited ground data. These results are consistent with recent research by Som-Ard et al. [11], who also applied the RFR model to estimate sugarcane yields in small field sizes and cloudy regions, achieving high accuracy with an R^2 exceeding 0.65. Similarly, studies by Canata et al. [15]; Xu et al. [52]; Jeong et al. [53]; and Sakamoto [54] have successfully applied the RFR model to map crop yields at the field scale for various crops, including sugarcane, wheat, maize, soybean, and potato. In Fusui County, Chongzui City, Guangxi Province, China, Xu et al. [52] compared several predictive models, including MLR, SMR, generalized linear regression (GLM), generalized boosted model (GBM), kernel-based regularized least squares (KRLS), and RFR, combined with UAV-mounted LiDAR for estimating sugarcane yield at the field scale. Their findings highlighted the superior performance of the RFR model in delivering accurate predictions. In this study, the effectiveness of the RFR model, based on a limited number of sampling datasets, was confirmed with a lower RMSE value (<7 t/ha) compared to previous research [20]. This approach also proved efficient in complex landscapes and small crop field sizes (<1 ha), as observed in the Tha Khan Tho District, Thailand. These findings align with the results of Canata et al., [15], who demonstrated that the RFR model outperformed MLR in estimating sugarcane yield on farms in Brazil.

The study showed consistent and efficient results in estimating sugarcane yield by combining multi-temporal Landsat and S2 datasets particularly for mapping sugarcane yield in small fields and cloudy regions. This aligns with previous studies [11,20,21], which highlighted that the accuracy of sugarcane yield mapping improves significantly when combining multiple EO data sources (e.g., S1, S2 and Landsat) due to the availability of a dense time-series dataset. Moreover, the key features in this study were mostly identified in November, during the final growth stage of the sugarcane crop, when the highest greenness is observed. This stage is closely linked to biomass and yield, as indicated in the studies of Som-ard et al. [13] and Suwanlee et al. [14]. These findings offer valuable insights for selecting satellite imagery data to estimate sugarcane crop yield in a timely and accurate manner.

This study demonstrates the robustness and reliability of the RFR approach in mapping sugarcane yield especially in regions with cloud cover and small crop field sizes. Additionally, this approach enhances the spatial diversity of estimated crop yield maps, benefiting not only sugarcane but also other crops like cassava, rice, corn, and wheat. This underscores the versatility and applicability of multi-temporal satellite imagery in agricultural yield estimation and management practices. However, further research is needed to explore the applicability of the RFR method across diverse geographic regions, incorporating varied training datasets and utilizing advanced parameter tuning techniques. Future studies could also benefit from integrating dense time-series data from additional sources, such as S1 or Planet imagery, to capture variations in sugarcane and other crop yields. Additionally, this study relied on a limited number of training datasets for model development; future works should focus on increasing the quantity and quality of ground data to better evaluate and compare different machine learning predictive models.

4.2. Spatial Distribution of Sugarcane Yield Map

Our findings provided valuable insights into the variability of sugarcane yield densities at the field scale, assisting farmers and researchers in crop management tasks such as monitoring crop health, identifying stress factors, and estimating yield levels across different fields. The yield densities were well smoothed and distributed across the study region at the field scale. The results of this study are aligned with previous research, confirming that the RFR model consistently outperforms other models, as demonstrated by its lowest RMSE and highest R^2 values [13,52]. Moreover, numerous studies have highlighted the efficacy of utilizing temporal S2 and L8 datasets for accurately estimating sugarcane yield, showing smoothed spatial distribution patterns across entire sugarcane fields [13,20,55,56]. The suitability of S2 and L8 imagery for sugarcane yield estimation is further supported by findings from other studies [15,16,20,55–57].

4.3. Insight on Sugarcane Crop Management

It is worth highlighting that the evaluation performed in this study was specifically tailored for small crop field sizes (<1 ha) and regions characterized by high cloud cover, such as the northeast region of Thailand, where sugarcane serves as a primary economic crop cultivated extensively. Despite the economic significance of sugarcane cultivation in this region, there is a notable scarcity of district-level maps depicting sugarcane production and estimated yield. Accurate yield estimation is important for precision decision making in agricultural management, as emphasized in recent studies [11,13].

The precise yield maps generated in this study offer local farmers and stakeholders detailed spatial information on expected sugarcane yields across their fields. Additionally, these maps provide insights into areas with potential yield variations, aiding in the identification of areas requiring targeted interventions [13]. Previous research by Sawaengsak and Gheewala [58] identified various factors contributing to reduced sugarcane production in Nakhon Ratchasima Province, including misunderstandings of good agricultural practices, inadequate soil management, limited knowledge of fertilization, and adverse weather conditions such as drought.

In Thailand, sugarcane cultivation faces numerous challenges, as outlined in reports from the Department of Agricultural Extension [59]. These challenges include deficiencies in field management practices such as soil maintenance, water management, and the use of fertilizers and pesticides. Moreover, there are gaps in the development of agricultural technology aimed at improving sugarcane yield. Outbreaks of diseases like white leaf disease have also been reported, leading to decreased yield, reduced sweetness, and weakened ability to retain sugarcane stumps [59].

The methodology proposed in this work, by providing precise estimations of sugarcane yield, can assist in evaluating optimal planting practices to ensure favorable yields and mitigate the impact of these challenges. Furthermore, the capability to reliably assess fields affected by sugarcane-related issues was demonstrated, offering insight into potential alternatives to address these impacts.

In the context of global sustainable agriculture, the optimization of crop management practices through the integration of EO datasets from multiple sensors, combined with the RFR model, presents local farmers with opportunities to reduce the environmental footprint of sugarcane cultivation. This includes minimizing chemical usage, reducing water consumption, and mitigating greenhouse gas emissions, all while maintaining or even improving productivity [60,61].

Mapping and monitoring sugarcane yield represents a transformative step toward smart agriculture and precision agricultural systems in Thailand. The proposed approach also promotes environmental sustainability by encouraging resource-efficient agricultural practices. By accurately mapping sugarcane yield, farmers can minimize waste and reduce the need for burning sugarcane fields. Additionally, these accurate yield maps enable farmers to mitigate the negative effects of agricultural activities on soil health and water quality.

5. Conclusions

The integration of Sentinel-2 (S2) and Landsat 8/9 (L8/L9) datasets with advanced machine-learning (ML) algorithms, based on the small number of training sites, represents a significant advancement in estimating sugarcane yield at the field scale in the northeast region of Thailand. In this study, the use of the random forest regression (RFR) algorithm proved to be an accurate and efficient method for mapping sugarcane yield. The resulting maps provided detailed information on sugarcane yield densities in small field sizes for the year 2022, demonstrating the approach's ability to capture variations in yield across different areas. The accurate yield map generated in this study offers valuable insights into sustainable agricultural practices and enhances our understanding of sugarcane yield dynamics.

The main findings and further suggestions of this work can be concluded as follows:

- The random forest regression (RFR) model demonstrated high accuracy in sugarcane yield estimation compared to other predictive models like MLR, SMR, PLS and SVR. The RFR model proved particularly effective in smallholder crop systems and regions with persistent cloud cover, such as Thailand's northeast. It achieved an R^2 of 0.79 and an RMSE of 3.93 t/ha;
- This study evaluated machine-learning performance for sugarcane yield estimation using a limited samples size (60 sampling plots) due to cost and time constraints. Despite the small dataset, the RFR model achieved highly accurate results for mapping sugarcane yield in small field sizes (<1 ha) and complex landscapes;
- Variability in sugarcane yield within fields was observed with results showing smooth and evenly distributed yield estimates across the region. The average estimated yield was 93.06 t/ha with a standard deviation (SD) of 10.25 t/ha;
- The RFR method, even with a limited ground dataset, showed efficiency in mapping sugarcane fields;
- The high accuracy of yield estimates at the field level provides valuable information for local farmers and stakeholders, aiding in the understanding of crop conditions and the optimizing fertilizer and irrigation practices. This approach to sugarcane yield mapping represents a significant advancement in smart agriculture for Thailand, offering farmers actionable insights to enhance productivity, profitability and environmental management.

Author Contributions: Conceptualization, F.S. and J.S.-a.; methodology, S.R.S. and J.S.-a.; software, K.K., N.S., S.K. and J.S.-a.; validation, K.K., N.S., S.K., F.S. and J.S.-a.; formal analysis, K.K., N.S., S.K. and J.S.-a.; writing—original draft preparation, S.R.S., D.P., S.N., E.B.-M., F.S. and J.S.-a.; writing—review and editing, S.R.S., D.P., S.K., S.N., E.B.-M., F.S. and J.S.-a.; visualization, F.S., D.P., S.R.S., S.K., S.N. and J.S.-a.; supervision, S.R.S. and J.S.-a.; project administration, S.R.S. and J.S.-a.; funding acquisition, S.R.S. and J.S.-a. All authors have read and agreed to the published version of the manuscript.

Funding: The project is funded by Mahasarakham University.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: This research project was financially supported by Mahasarakham University.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Barbosa Júnior, M.R.; Tedesco, D.; Corrêa, R.d.G.; Moreira, B.R.d.A.; Silva, R.P.d.; Zerbato, C. Mapping gaps in sugarcane by UAV RGB imagery: The lower and earlier the flight, the more accurate. *Agronomy* **2021**, *11*, 2578. [[CrossRef](#)]
2. Momin, M.A.; Grift, T.E.; Valente, D.S.; Hansen, A.C. Sugarcane yield mapping based on vehicle tracking. *Precis. Agric.* **2019**, *20*, 896–910. [[CrossRef](#)]
3. Molin, J.P.; Veiga, J.P.S. Spatial variability of sugarcane row gaps: Measurement and mapping. *Cienc. Agrotec.* **2016**, *40*, 347–355. [[CrossRef](#)]

4. Birrell, S.J.; Sudduth, K.A.; Borgelt, S.C. Comparison of sensors and techniques for crop yield mapping. *Comput. Electron. Agric.* **1996**, *14*, 215–233. [[CrossRef](#)]
5. Bégué, A.; Lebourgeois, V.; Bappel, E.; Todoroff, P.; Pellegrino, A.; Baillarin, F.; Siegmund, B. Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI. *Int. J. Remote Sens.* **2010**, *31*, 5391–5407. [[CrossRef](#)]
6. Palaniswami, C.; Gopalasundaram, P.; Bhaskaran, A. Application of GPS and GIS in Sugarcane Agriculture. *Sugar Tech* **2011**, *13*, 360–365. [[CrossRef](#)]
7. Noureldin, N.; Aboelghar, M.; Saady, H.; Ali, A. Rice yield forecasting models using satellite imagery in Egypt. *Egypt. J. Remote Sens. Space Sci.* **2013**, *16*, 125–131. [[CrossRef](#)]
8. FAO. Food and Agriculture Organization of the United Nations. Available online: <https://www.fao.org/faostat/en/?#data/QCL> (accessed on 13 April 2024).
9. Sriroth, K.; Vanichsiratana, W.; Sunthornvarabhas, J. The current status of sugar industry and by-products in Thailand. *Sugar Tech* **2016**, *18*, 576–582. [[CrossRef](#)]
10. Office of the Cane & Sugar Board. *Documentation Detailing the Status of Sugarcane Cultivation during the Production Year 2022/23*; Office of The Cane and Sugar Board: Bangkok, Thailand, 2023.
11. Som-Ard, J.; Atzberger, C.; Izquierdo-Verdiguier, E.; Vuolo, F.; Immitzer, M. Remote sensing applications in sugarcane cultivation: A review. *Remote Sens.* **2021**, *13*, 4040. [[CrossRef](#)]
12. Chunhawong, K.; Chaisan, T.; Rungmekarat, S.; Khotavivattana, S. Sugar industry and utilization of its by-products in Thailand: An overview. *Sugar Tech* **2018**, *20*, 111–115. [[CrossRef](#)]
13. Som-ard, J.; Immitzer, M.; Vuolo, F.; Atzberger, C. Sugarcane yield estimation in Thailand at multiple scales using the integration of UAV and Sentinel-2 imagery. *Precis. Agric.* **2024**, *25*, 1581–1608. [[CrossRef](#)]
14. Suwanlee, S.R.; Pinasu, D.; Som-ard, J.; Borgogno-Mondino, E.; Sarvia, F. Estimating Sugarcane Aboveground Biomass and Carbon Stock Using the Combined Time Series of Sentinel Data with Machine Learning Algorithms. *Remote Sens.* **2024**, *16*, 750. [[CrossRef](#)]
15. Canata, T.F.; Wei, M.C.F.; Maldaner, L.F.; Molin, J.P. Sugarcane yield mapping using high-resolution imagery data and machine learning technique. *Remote Sens.* **2021**, *13*, 232. [[CrossRef](#)]
16. Shendryk, Y.; Davy, R.; Thorburn, P. Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. *Field Crops Res.* **2021**, *260*, 107984. [[CrossRef](#)]
17. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [[CrossRef](#)]
18. Atzberger, C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sens.* **2013**, *5*, 949–981. [[CrossRef](#)]
19. de França e Silva, N.R.; Chaves, M.E.D.; Luciano, A.C.d.S.; Sanches, I.D.A.; de Almeida, C.M.; Adami, M. Sugarcane Yield Estimation Using Satellite Remote Sensing Data in Empirical or Mechanistic Modeling: A Systematic Review. *Remote Sens.* **2024**, *16*, 863. [[CrossRef](#)]
20. Rahman, M.M.; Robson, A. Integrating landsat-8 and sentinel-2 time series data for yield prediction of sugarcane crops at the block level. *Remote Sens.* **2020**, *12*, 1313. [[CrossRef](#)]
21. Abebe, G.; Tadesse, T.; Gessesse, B. Combined use of Landsat 8 and Sentinel 2A imagery for improved sugarcane yield estimation in Wonji-Shoa, Ethiopia. *J. Indian. Soc. Remote Sens.* **2022**, *50*, 143–157. [[CrossRef](#)]
22. Dimov, D.; Uhl, J.H.; Löw, F.; Seboka, G.N. Sugarcane yield estimation through remote sensing time series and phenology metrics. *Smart Agric. Technol.* **2022**, *2*, 100046. [[CrossRef](#)]
23. Louis, J.; Debaecker, V.; Pflug, B.; Main-Knorn, M.; Bieniarz, J.; Mueller-Wilm, U.; Cadau, E.; Gascon, F. Sentinel-2 Sen2Cor: L2A processor for users. In Proceedings of the ‘Living Planet Symposium 2016’, Prague, Czech Republic, 9–13 May 2016; pp. 1–8.
24. Som-ard, J. Rice security assessment using geo-spatial analysis. *Int. J. Geoinform.* **2020**, *16*, 21–38.
25. Suwanlee, S.R.; Keawsomsee, S.; Pengjunsang, M.; Homtong, N.; Prakobya, A.; Borgogno-Mondino, E.; Sarvia, F.; Som-ard, J. Monitoring Agricultural Land and Land Cover Change from 2001–2021 of the Chi River Basin, Thailand Using Multi-Temporal Landsat Data Based on Google Earth Engine. *Remote Sens.* **2023**, *15*, 4339. [[CrossRef](#)]
26. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309.
27. Serrano, L.; Ustin, S.L.; Roberts, D.A.; Gamon, J.A.; Penuelas, J. Deriving water content of chaparral vegetation from AVIRIS data. *Remote Sens. Environ.* **2000**, *74*, 570–581. [[CrossRef](#)]
28. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
29. Klemas, V.; Smart, R. The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of *Spartina Alterniflora* canopies. *Photogramm. Eng. Remote Sens.* **1983**, *49*, 77–83.
30. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
31. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]
32. Kaufman, Y.J.; Tanre, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 261–270. [[CrossRef](#)]

33. Jin, S.; Sader, S.A. Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sens. Environ.* **2005**, *94*, 364–372. [CrossRef]
34. Wilson, E.H.; Sader, S.A. Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sens. Environ.* **2002**, *80*, 385–396. [CrossRef]
35. Cihlar, J.; Manak, D.; D’Iorio, M. Evaluation of compositing algorithms for AVHRR data over land. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 427–437. [CrossRef]
36. Griffiths, P.; van der Linden, S.; Kuemmerle, T.; Hostert, P. A pixel-based Landsat compositing algorithm for large area land cover mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 2088–2101. [CrossRef]
37. White, J.C.; Wulder, M.; Hobart, G.; Luther, J.; Hermosilla, T.; Griffiths, P.; Coops, N.; Hall, R.; Hostert, P.; Dyk, A. Pixel-based image compositing for large-area dense time series applications and science. *Can. J. Remote Sens.* **2014**, *40*, 192–212. [CrossRef]
38. Maxwell, S.K.; Sylvester, K.M. Identification of “ever-cropped” land (1984–2010) using Landsat annual maximum NDVI image composites: Southwestern Kansas case study. *Remote Sens. Environ.* **2012**, *121*, 186–195. [CrossRef]
39. Araiza-Aguilar, J.; Rojas-Valencia, M.; Aguilar-Vera, R. Forecast generation model of municipal solid waste using multiple linear regression. *Glob. J. Environ. Sci. Manag.* **2020**, *6*, 1–14.
40. Abrougui, K.; Gabsi, K.; Mercatoris, B.; Khemis, C.; Amami, R.; Chehaibi, S. Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil. Till. Res.* **2019**, *190*, 202–208. [CrossRef]
41. Xu, K.; Su, Y.; Liu, J.; Hu, T.; Jin, S.; Ma, Q.; Zhai, Q.; Wang, R.; Zhang, J.; Li, Y. Estimation of degraded grassland aboveground biomass using machine learning methods from terrestrial laser scanning data. *Ecol. Indic.* **2020**, *108*, 105747. [CrossRef]
42. Yamashita, T.; Yamashita, K.; Kamimura, R. A stepwise AIC method for variable selection in linear regression. *Commun. Stat.—Theory Methods* **2007**, *36*, 2395–2403. [CrossRef]
43. Abdi, H. Partial least square regression (PLS regression). *Encycl. Res. Methods Soc. Sci.* **2003**, *6*, 792–795.
44. Wold, S.; Trygg, J.; Berglund, A.; Antti, H. Some recent developments in PLS modeling. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 131–150. [CrossRef]
45. Huang, Z.; Turner, B.J.; Dury, S.J.; Wallis, I.R.; Foley, W.J. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sens. Environ.* **2004**, *93*, 18–29. [CrossRef]
46. Hubert, M.; Branden, K.V. Robust methods for partial least squares regression. *J. Chemom.* **2003**, *17*, 537–549. [CrossRef]
47. Chun, H.; Keleş, S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Stat.* **2010**, *72*, 3–25. [CrossRef] [PubMed]
48. Abdel-Rahman, E.M.; Mutanga, O.; Odindi, J.; Adam, E.; Odindo, A.; Ismail, R. A comparison of partial least squares (PLS) and sparse PLS regressions for predicting yield of Swiss chard grown under different irrigation water sources using hyperspectral data. *Comput. Electron. Agric.* **2014**, *106*, 11–19. [CrossRef]
49. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
50. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
51. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995; Volume 10, pp. 123–160.
52. Xu, J.-X.; Ma, J.; Tang, Y.-N.; Wu, W.-X.; Shao, J.-H.; Wu, W.-B.; Wei, S.-Y.; Liu, Y.-F.; Wang, Y.-C.; Guo, H.-Q. Estimation of sugarcane yield using a machine learning approach based on uav-lidar data. *Remote Sens.* **2020**, *12*, 2823. [CrossRef]
53. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R. Random forests for global and regional crop yield predictions. *PLoS ONE* **2016**, *11*, e0156571. [CrossRef]
54. Sakamoto, T. Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 208–228. [CrossRef]
55. Morel, J.; Todoroff, P.; Bégué, A.; Bury, A.; Martiné, J.-F.; Petit, M. Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on Reunion Island. *Remote Sens.* **2014**, *6*, 6620–6635. [CrossRef]
56. Rahman, M.M.; Robson, A.J. A novel approach for sugarcane yield prediction using landsat time series imagery: A case study on Bundaberg region. *Adv. Remote Sens.* **2016**, *5*, 93–102. [CrossRef]
57. Sharifi, A. Yield prediction with machine learning algorithms and satellite images. *J. Sci. Food. Agric.* **2021**, *101*, 891–896. [CrossRef] [PubMed]
58. Sawaengsak, W.; Gheewala, S.H. Analysis of social and socio-economic impacts of sugarcane production: A case study in Nakhon Ratchasima province of Thailand. *J. Clean. Prod.* **2017**, *142*, 1169–1175. [CrossRef]
59. Department of Agricultural Extension. Sugar Cane Production. Available online: <https://esc.doae.go.th/> (accessed on 12 April 2024).
60. Burney, J.A.; Davis, S.J.; Lobell, D.B. Greenhouse gas mitigation by agricultural intensification. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12052–12057. [CrossRef]
61. Kropp, I.; Nejadhashemi, A.P.; Deb, K.; Abouali, M.; Roy, P.C.; Adhikari, U.; Hoogenboom, G. A multi-objective approach to water and nutrient efficiency for sustainable agricultural intensification. *Agric. Syst.* **2019**, *173*, 289–302. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.