

Article

Estimating the Aboveground Fresh Weight of Sugarcane Using Multispectral Images and Light Detection and Ranging (LiDAR)

Charot M. Vargas ¹, Muditha K. Heenkenda ^{2,*} and Kerin F. Romero ¹

¹ Department of Agricultural Engineering, Instituto Tecnológico de Costa Rica, Cartago 30101, Costa Rica; charot@estudiantec.cr (C.M.V.); kromero@itcr.ac.cr (K.F.R.)

² Department of Geography and the Environment, Lakehead University, Thunder Bay, ON P7B 5E1, Canada

* Correspondence: muditha.heenkenda@lakeheadu.ca; Tel.: +1-807-343-8010 (ext. 8746)

Abstract: This study aimed to develop a remote sensing method for estimating the aboveground fresh weight (AGFW) of sugarcane using multispectral images and light detection and ranging (LiDAR). Remotely sensed data were acquired from an unmanned aerial vehicle (drone). Sample plots were harvested and the AGFW of each plot was measured. Sugarcane crown heights and volumes were obtained by isolating individual tree crowns using a LiDAR-derived digital surface model of the area. Multiple linear regression (MLR) and partial least-squares regression (PLSR) models were tested for the field-sampled AGFWs (dependent variable) and individual canopy heights and volumes, and spectral indices were used as independent variables or predictors. The PLSR model showed more promising results than the MLR model when predicting the AGFW over the study area. Although PLSR is well-suited to a large number of collinear predictor variables and a limited number of field samples, this study showed moderate results ($R^2 = 0.5$). The visual appearance of the spatial distribution of the AGFW map is satisfactory. The limited no. of field samples overfitted the MLR prediction results. Overall, this research highlights the potential of integrating remote sensing technologies in the sugarcane industry, thereby improving yield estimation and effective crop management.



Citation: Vargas, C.M.; Heenkenda, M.K.; Romero, K.F. Estimating the Aboveground Fresh Weight of Sugarcane Using Multispectral Images and Light Detection and Ranging (LiDAR). *Land* **2024**, *13*, 611. <https://doi.org/10.3390/land13050611>

Academic Editors: Massimo Fabris and Djamil Al-Halbouni

Received: 3 April 2024

Revised: 28 April 2024

Accepted: 29 April 2024

Published: 1 May 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/>).

1. Introduction

Sugarcane (*Saccharum* spp.) is one of the largest broadacre crops in the world. It is grown in 106 countries across the tropics and subtropics and accounts more than 70% of global sugar consumption [1]. Additionally, sugarcane plays a fundamental role in renewable energy production, ranking as the second major feedstock for ethanol production [2]. Moreover, the crop is identified as the most efficient bioenergy crop in tropical regions, and biotechnological tools have been emphasized in maximizing its potential for sustainable energy crop development [3]. These factors underscore the economic and scientific significance of this crop. As sugarcane remains an economic mainstay for many regions of the world, the indirect estimation of sugarcane yields is of crucial importance for the induction of this crop [4].

Sugarcane cultivation is one of Costa Rica's main agricultural products and plays a key role in boosting the national economy; for example, the sugarcane sector, through production and agroindustry, contributes to 0.33% of the gross national product and 3.83% of the gross agricultural domestic product [5]. The sugarcane industry generates 25,000 direct jobs and more than 100,000 indirect jobs [6]. Given the significant importance of sugarcane cultivation in the agricultural sector in Costa Rica, crop yield assessment plays a vital role. The accurate estimation of factors such as the aboveground fresh weight of crops before harvesting enables effective crop management while increasing the capacity for

informed decision-making [7]. For instance, knowledge of the aboveground fresh weight of sugarcane allows producers to predict the amount of raw material to be harvested and, accordingly, to plan production and planning processes more efficiently in terms of managing resources and optimizing production and decision-making [8].

Traditionally, the crop yield has been determined by directly weighing the harvest from a sample of the farm and averaging the whole field. However, this is costly and laborious at the production level for large areas; therefore, there is a demand for developing an alternative method that incorporates the use of new technologies [9,10]. Predicting the sugarcane yield is not an easy task, although some studies have explored alternatives to improve this process. A study carried out in Brazil revealed significant comparative results between data obtained from multispectral images (red, green, and blue bands (RGB images)) acquired by unmanned aerial vehicles (UAVs) and field observations of the aboveground yield [11]. The study aimed to assess the degree of canopy closure in order to predict the potential yield. The crop growth was evaluated by considering the Leaf Area Index (LAI) and Green–Red Vegetation Index (GRVI). The results revealed that the GRVI ($R^2 = 0.69$) provided better results compared to the LAI ($R^2 = 0.34$) when predicting the yield [11]. Sumesh et al. [8] estimated the sugarcane yield using a consumer-grade red–green–blue (RGB) camera mounted on a UAV with a minimal field dataset. The yield was calculated by integrating the spatial variability of plant height, stalk density, and weight using ordinary least-squares regression. The study outlined a very-high-spatial-resolution RGB image and object-based image analysis (OBIA) approach for mapping the spatial variability of plant height and stalk density and estimating the sugarcane yield [8]. Another study conducted in Thailand demonstrated higher accuracies of OBIA (92% and 96%) in contrast to pixel-based methods (84% and 88%) for estimating sugarcane coverage [12]. After isolating the sugarcane coverage, the number of stems and the average height and weight were calculated to assess the yield [12]. The results indicated the effectiveness of applying OBIA techniques for estimating the sugarcane aboveground fresh weight [12]. Aleman et al. [13] estimated the sugarcane yield by integrating aerial photographs, photogrammetric methods, and field determinations. Digital surface models (DSMs) and digital terrain models (DTMs) were obtained, and based on those, the height and volume of the sugarcane were calculated. These estimates were correlated with different yield variables, such as metric tons of sugarcane per hectare. There was a high linear correlation between this variable and the height of the crop ($R^2 = 0.83$). The absolute error of the yield estimation from plant height was 3.7 ton/ha. However, although this study provided promising results, sugarcane spectral information was not considered in the analysis. Therefore, demand still exists for analyzing sugarcane's spectral information and thus estimating the aboveground fresh weight.

Some studies have evaluated sugarcane crop growth stages using light detection and ranging (LiDAR) technology. Villareal and Tongco [14] achieved 98.4% accuracy for classifying sugarcane crop growth stages and 91.7% for detecting the maturation stage. These results support the usefulness of developing sets of rules, suggesting their applicability to map and verify other sugarcane plantation areas. Furthermore, as LiDAR data easily capture information on the vegetation structure, they are suitable for predicting the aboveground fresh weight and generating spatial distribution maps accordingly [9,15]. LiDAR data provide more consistent and significant correlations of sugarcane biophysical parameters, especially volume and height, than multispectral images [15]. Aerial images provide quantitative information about the vegetation and its health conditions through spectral data analysis, reflecting the expected amount of yield [16]. Hence, an integrated approach combining these two data types might provide accurate statistics on the sugarcane aboveground fresh weight. Although existing studies have demonstrated the effectiveness of individual remote sensing technologies (multispectral imaging or LiDAR) for predicting the aboveground fresh weight (AGFW), according to the authors' knowledge, there are very limited or no studies integrating these technologies, especially in Costa Rica, where we conducted this research.

The aim of this research was to develop a remote sensing method for estimating sugarcane AGFW to streamline production at the farm level. The proposed data included multispectral images and LiDAR captured from a UAV as well as field observations (AGFW of sample plots) for calibration and validation. The specific objectives were to (1) delineate individual sugarcane tree crowns using LiDAR; (2) develop a mathematical model that estimates the AGFW of sugarcane using spectral values from remotely sensed imagery and field observations; and (3) apply the developed model to predict the AGFW over the entire study area.

2. Materials and Methods

2.1. Study Area

This study was conducted over a sugarcane field located in Guanacaste Province, Costa Rica ($10^{\circ}20'9''$ N, $85^{\circ}9'31''$ W) (Figure 1). The farm is called “Agrícola El Cántaro”, and the cultivated sugarcane variety was “CP722086”. The average temperature in this area is consistently high, averaging 28.3°C with a maximum of 40.1°C and a minimum of 15.9°C , which makes it a suitable region for sugarcane cultivation [5]. The summer months span from December to April, typically the sugarcane harvesting months, and coincide with the resowing of sugarcane during the first week of January [17].

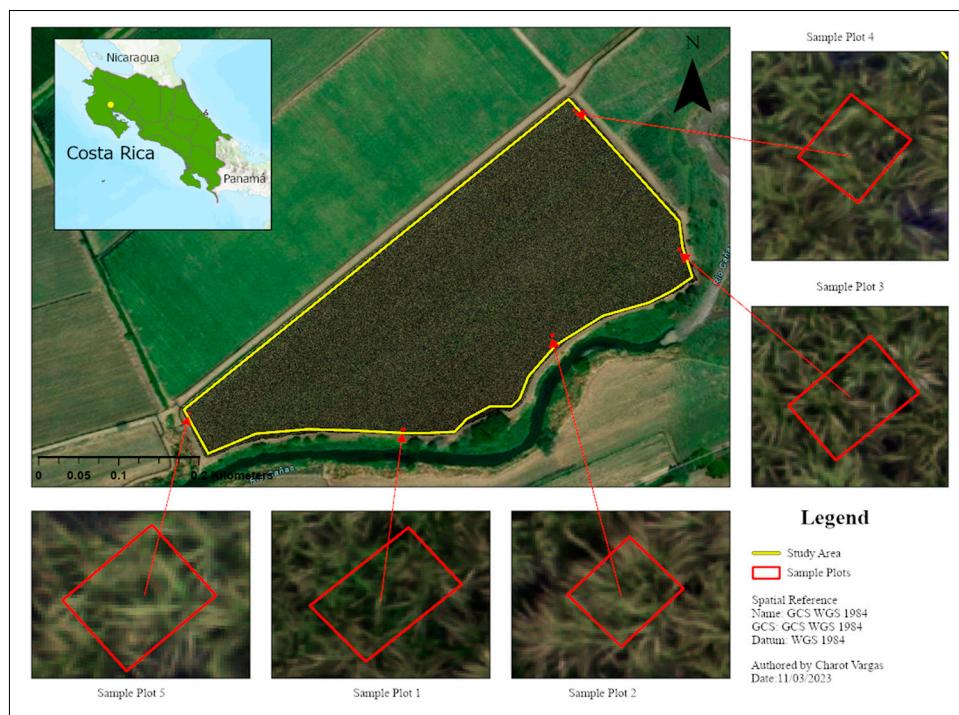


Figure 1. The study area map and the locations of field plots.

The field was in its fourth-year reproduction cycle and was eight months old at data acquisition (in the maturation stage). According to the farm management, this plot was homogeneous and productive compared to other plots within the neighborhood. The farm utilizes a gravity irrigation system and implements scheduled amendments to the crops [18].

2.2. Data Acquisition and Preprocessing

The aerial images and LiDAR and field observations were collected on 24 August 2023. A UAV: DJI Matrice 300 [19] equipped with a Real-Time Kinematic (RTK) differential correction system, Zenmuse L1 laser scanner [20], and Micasense Red Edge-P camera [21] were used for data acquisition. The flying height was selected as 60 m aboveground for LiDAR data acquisition. To capture multispectral images with a high spatial resolution (5 cm),

the flying height was 80 m above ground level. The LiDAR point cloud was converted to the LAS format for further processing, as a standard binary format is recommended for industry usage [22]. LiDAR data were cleaned by removing the data noise, and a semi-automatic classification of low points was performed, considering variations in iteration angle, iteration distance, terrain angle, and terrain slope [23]. This process was carried out using Spatix software (Version 023.002) [24].

The camera provided multispectral images with five bands and a panchromatic image. Table 1 shows the spectral sensitivity (center wavelength) of each band. Multispectral images were processed to obtain the orthomosaic (with spectral reflectance values) for further analysis using Pix4DMapper software (v 4.9) [25].

Table 1. Characteristics of the Micasense RedEdge-P images (adapted from <https://ageagle.com/drone-sensors/rededge-p-high-res-multispectral-camera/>; accessed on 10 September 2023).

Band	Wavelength (nm)	Bandwidth (nm)
Blue	475	32
Green	560	27
Red	668	14
Red edge	717	12
Near infrared (NIR)	842	57
Panchromatic	634.5	463

Five sampling plots (3 m^2 each) were identified based on the green appearance (homogeneity) of the plants, the spatial distribution, and the access to the plots. The location of each plant on each sampling plot was captured using a Global Navigation Satellite System with an RTK Fix solution. The field measurements showed that the distance between sugarcane rows on the farm was not uniform; however, the average distance between rows was 0.9 m, and the average distance between plants was 0.65 m. The total weight of millable stalks (aboveground fresh weight (AGFW)) in each harvest plot was measured; 70% of sampling was used to create the models, and 30% was used for validation. Figure 2 shows the complete workflow for this study.

2.3. Individual Tree Crown Delineation

A digital terrain model (DTM) and the digital surface model (DSM) were extracted from the LiDAR data (Figure 2). A DTM contains only information about the terrain surface, excluding objects and surface features, and a DSM consists of an elevation dataset that includes both terrain information and information about the objects and surface features present [26]. The canopy height model (CHM) was created for a spatial resolution of 0.2 m using Equation (1).

$$\text{CHM} = \text{DSM} - \text{DTM} \quad (1)$$

The individual tree crowns were isolated using the “lidR” package in R software [27]. First, the CHM’s highest points (local maxima) were extracted as sugarcane tree top locations. A threshold value of 0.5 m was applied to avoid small plants and stalks in this calculation. Then, the algorithm developed by Dalponte and Coomes [28] was used to extract individual tree crowns. The approximate crown diameter was selected as 1.7 m by manually measuring several crowns on-screen within the study area. Heights of individual trees and the canopy coverage were extracted (Figure 2).

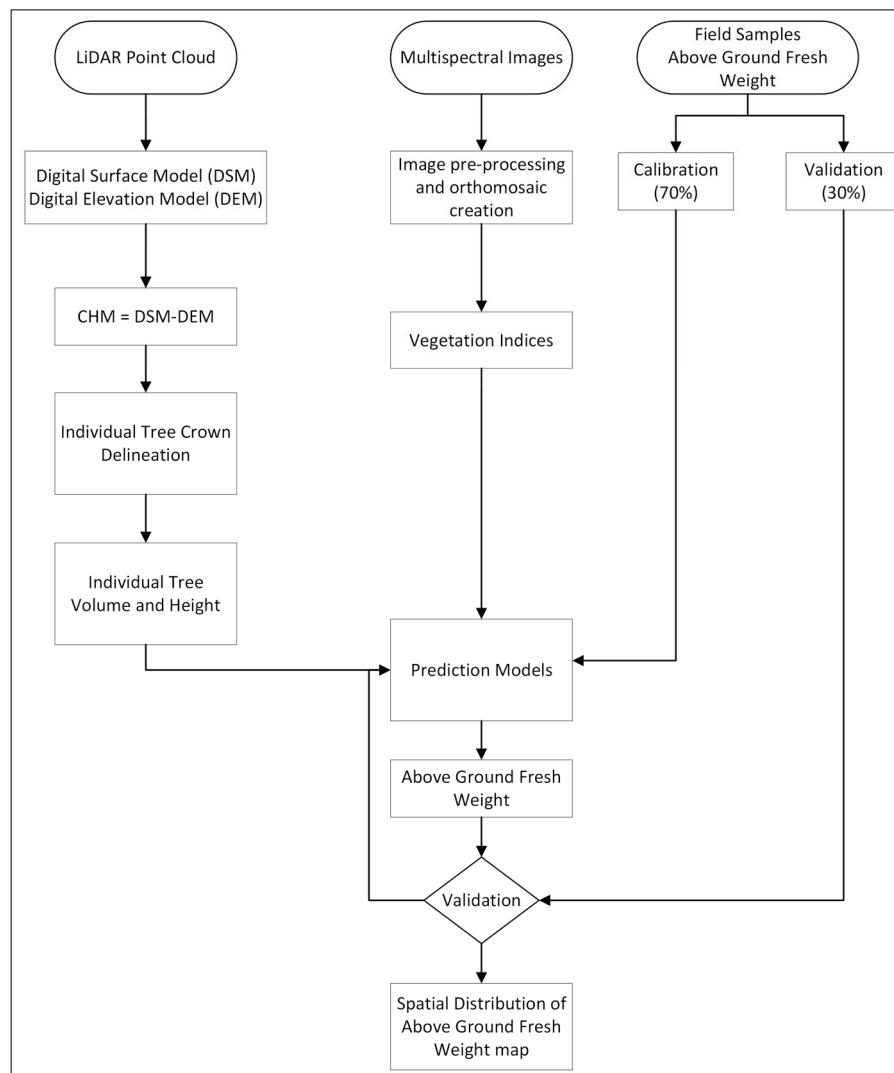


Figure 2. Flowchart explaining the overall workflow for this study.

2.4. Estimating the Aboveground Fresh Weight

As shown in Figure 2, vegetation indices were derived from the multispectral image. A vegetation index transforms the spectral values in a meaningful way to enhance the contribution of vegetation properties and thus ease the identification of vegetation vigor [29]. Nine vegetation indices (Table 2) were selected for this study based on their ability to detect the greenness of vegetation and biomass [30]. For instance, the Normalized Difference Vegetation Index (NDVI) is well-known for analyzing sugarcane biomass [10,12]. Additionally, there were some yield gaps in the study area due to replanting (small plants—early stages and bare soil), so the SAVI and MSAVI were selected considering their effectiveness in this situation, especially when there is abundant soil and small plants [31]. Meanwhile, the positive values of the Triangular Greenness Index (TDI) correspond to green vegetation and thus the leaf chlorophyll content [32], while the negative values may represent soil. Since the TDI is calculated using blue, green, and red wavelengths, it is suitable for RGB cameras in agricultural applications such as this one [33].

Table 2. Details of vegetation indices selected for this study.

Name	Equation	References
Green Normalized Difference Vegetation Index (GNDVI)	$\frac{NIR - Green}{NIR + Green}$	[34]
Normalized Difference Red Edge Index (NDRE)	$\frac{NIR - Red\ Edge}{NIR + Red\ Edge}$	[35]
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR - Red}{NIR + Red}$	[36]
Soil-Adjusted Vegetation Index (SAVI)	$\left(\frac{(NIR - red)}{(NIR + red+L)} \right) \times (1 + L); where L = 0.5$	[31]
Modified Soil-Adjusted Vegetation Index (MSAVI)	$2 \times NIR + 1 - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - Red)}$	[37]
Simple Ratio (SR)	$\frac{NIR}{Red}^2$	[38]
Normalized Green–Red Difference Index (NGRDI)	$\frac{Green - Red}{Green + Red}$	[33,37]
Normalized Green–Red Edge Difference Index (NDGRed)	$\frac{Green - Red\ Edge}{Green + Red\ Edge}$	[37]
Triangular Greenness Index (TGI)	$-0.5[(\lambda_r - \lambda_b)(Red - Green) - (\lambda_r - \lambda_g)(Red - Blue)]^*$	[33,37]

* λ_r , λ_b , and λ_g represent the center wavelengths for the red, blue, and green bands, respectively.

Once vegetation indices were created, two algorithms, multiple linear regression (MLR) and partial least-squares regression (PLSR), were tested to calculate the AGFW. For the MLR method, the correlations between the AGFW of each plot and independent variables were tested to select the best predictor variables. Each plot's canopy cover, mean canopy height, volume, and mean vegetation indices (NDVI, NDRE, GNDVI, SAVI, MSAVI) provided the highest correlations with field samples. Equation (2) shows an example of the MLR model.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \quad (2)$$

where Y is the dependent variable to be predicted, X_1 , X_2 , and X_n are the independent variables, β_0 is the intercept, β_1 , β_2 , and β_n are the coefficients for each independent variable, and ϵ is the error.

The accuracy of the model was assessed using the coefficient of determination R-squared (R^2). R^2 is a measure of the goodness of a fit of a model and indicates how the AGFW varies with predictor variables.

Since we had limited AGFW samples, the PLSR algorithm was also selected to predict the AGFW. PLSR is an efficient and optimal regression model based on covariances. The PLSR model generalizes and combines principal component analysis and multiple linear regression, producing a linear regression model [39]. It reduces the independent variables to a smaller set of predictors internally. The algorithm adds complexity and utility to the model, as it is designed to deal with situations where there are many independent variables, possibly correlated, and relatively few samples as dependent variables [40].

Random points were generated inside individual tree crowns of each sample plot to ensure the homogeneity of spectral values. Corresponding spectral band values (Table 1 except panchromatic band) and different vegetation indices (Table 2), canopy height, and volume were extracted for each point. The initial PLSR model was created using 15 predictor variables, assuming that more variables would improve the prediction power. The “pls” package in R software was used [41]. The model performance and the best number of components were evaluated using cross-validation (root mean square error of prediction (RMSEP)). Once the best number of components was identified, the final model was created using four components only.

Finally, the PLSR model was selected to predict the spatial distribution of AGFW over the study area. A low-pass filter was applied to the resultant raster to remove data noise. The accuracy of the predicted AGFW raster was assessed using R-squared (R^2).

3. Results

3.1. Individual Tree Crown Delineation

Individual tree crowns were detected successfully. Their accuracy was assessed visually compared to the multispectral image. For example, Figure 3a shows a part of the study area with delineated tree tops and crowns. However, some areas do not show tree tops or crowns (Figure 3b). These areas have trees shorter than 0.5 m (threshold applied in the algorithm) or soil.

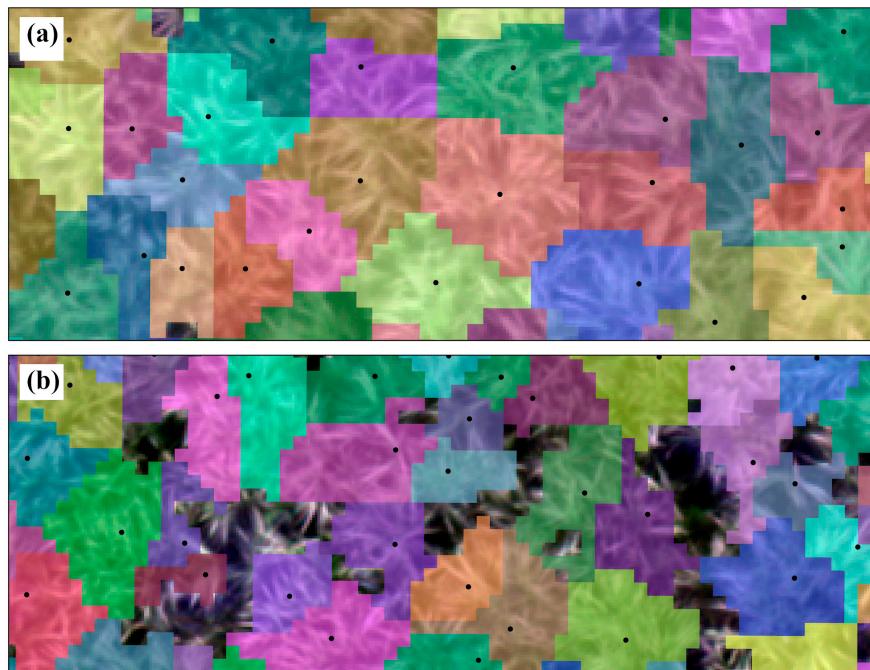


Figure 3. (a) Individual treetops and crowns detected (black dots represent treetops and polygons represent tree crowns). (b) Examples of areas not detected as tree crowns (lower than the height threshold value).

Also, the number of trees in each sample plot was compared with the number of trees delineated in each lot plot (Figure 4). There was a 100% agreement between these data for three plots. In plots 2 and 5, an extra crown was detected, mainly due to the foliage of neighboring plants.

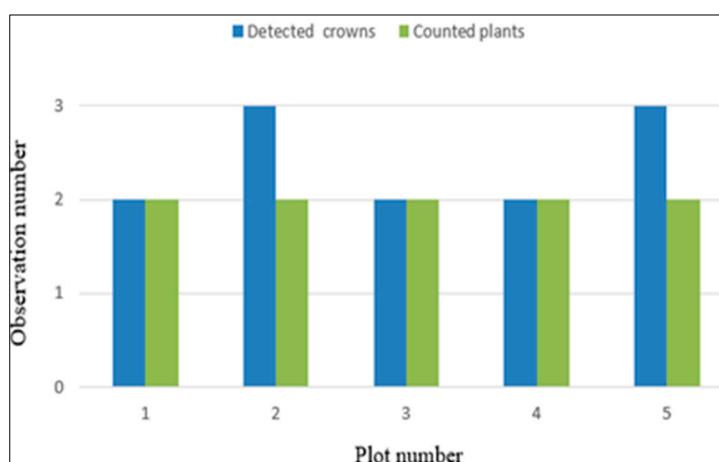


Figure 4. The total number of plants was manually counted in the field and delineated using remote sensing images in each sample plot.

3.2. Estimating Aboveground Fresh Weight

3.2.1. Multiple Linear Regression (MLR)

Table 3 shows the correlation between the AGFW of each sample plot and the mean value of each predictor variable. All predictor variables where we obtained a correlation (r) greater than 0.7 were selected for the final regression model—the MSAVI, SAVI, NDVI, and crown area (Table 3).

Table 3. The correlation between aboveground fresh weight and the predictor variable of each plot.

Variable	Correlation Coefficient (r)
Canopy area (m^2)	0.86
Canopy height (m)	0.63
NDVI	0.72
NDRE	0.67
GNDVI	0.63
SAVI	0.91
MSAVI	0.95
Volume	0.16

Figure 5 shows the R^2 values of each chosen predictor variable. In this study, the NDVI had a lower R^2 than the MSAVI and SAVI. Although we expected a high correlation between the crown area, canopy height, and aboveground fresh weight, they showed a moderate relation. The final MLR model showed a high correlation between sample data and selected predictors ($R^2 = 0.7$).

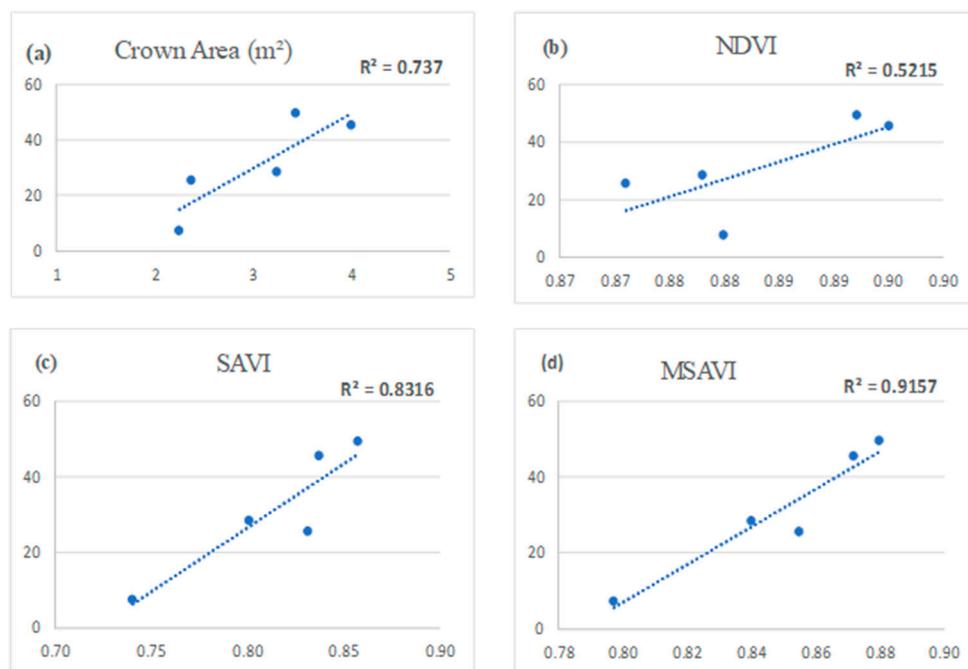


Figure 5. Correlations between aboveground fresh weight and predictor variables.

3.2.2. Partial Least-Squares Regression (PLSR) Model

As shown in Figure 6, the lowest RMSEP was obtained with only four components. Although three to seven components were within the lowest RMSEP range, increasing the number of components did not improve the prediction performance. The cross-validation results also confirmed the variability of RMSEP with a changing number of components.

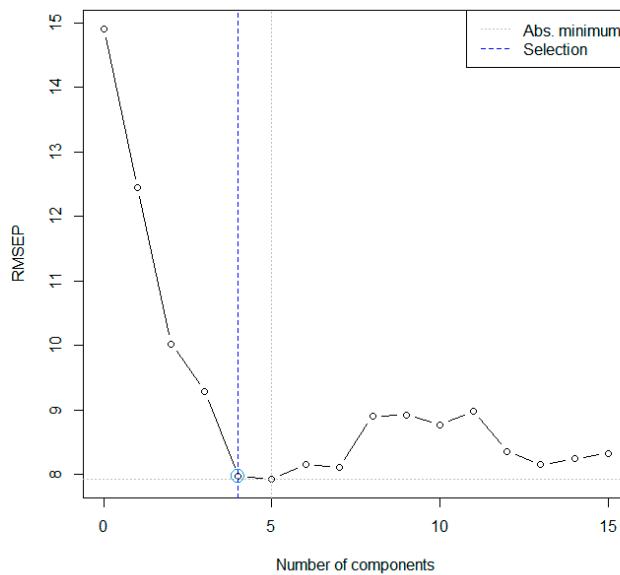


Figure 6. Evaluating the number of components for the PLSR model.

3.3. Spatial Distribution of Aboveground Fresh Weight

The prediction results of the MLR model were unreliable and exhibited over prediction. The resultant AGFW over the study area ranged from 200 kgm^{-2} to 300 kgm^{-2} , indicating model overfitting.

The PLSR model predicted the AGFW over the study area (Figure 7a). The results ranged from 33.1 kgm^{-2} to 70.0 kgm^{-2} . The R^2 compared to field samples was 0.5, indicating moderate prediction results. The AGFW showed higher values closer to areas with high precipitation, especially south of the plot, adjacent to a large river. Lower values were exhibited closer to the access road, where the ground elevation was higher than that of other areas. As such, the visual analysis of the map indicated a reasonable pattern related to the geography and availability of water in the study area. Figure 7b illustrates the distribution of the mean AGFW of each crown in the study area, and the red density curve fitted to the data reflects normality in the distribution of these and the concentration of values around the mean, which has a value of 47 kgm^{-2} . Overall, the study area has 11.48 ha and an aboveground fresh weight of 184 ton/ha.

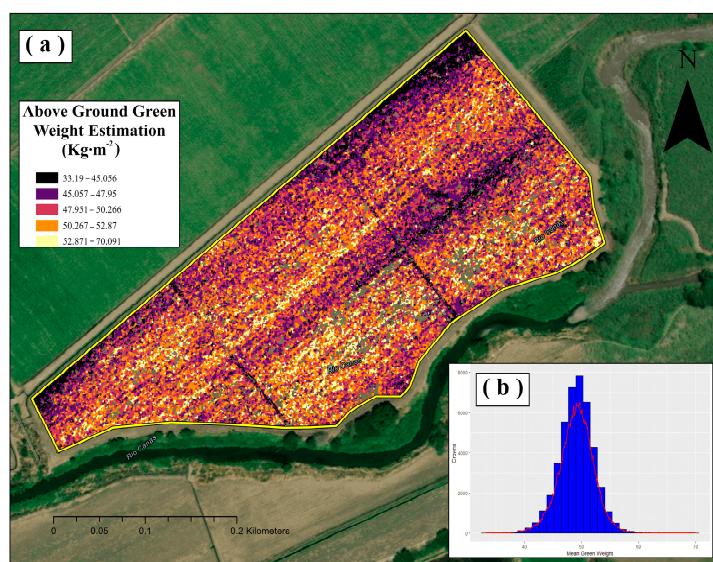


Figure 7. (a) Spatial distribution of the aboveground fresh weight estimation. (b) A histogram of the mean fresh weights of canopies.

4. Discussion

4.1. Individual Tree Crown Delineation

This study tested the tree crown delineation method developed by Dalponte and Coomes [28] for forest tree crowns. Since it was successful in the sugarcane environment with some customization, we demonstrate the method's applicability in sugarcane studies and its advantages over image segmentation. For instance, when identifying individual tree crowns using aerial image classification algorithms (pixel-based or object-based) without incorporating CHM, tree heights are not considered. Hence, small, juvenile plants are identified as mature trees in the calculation of the potential yield. Pixel-based image analysis has long been the mainstay approach for classifying remotely sensed imagery due to its simplicity, effectiveness in numerous studies, and wide range of applications [42]. However, the pixel-based classification experienced difficulties in segmenting sugarcane and non-sugarcane (noise, soil, and grass). It cannot classify the leaf structure and canopy due to limitations when considering texture and spatial structure only, leading to erroneous stem counting [12]. Object-based approaches offer a significant improvement in this regard, integrating a wide range of features that go beyond spectral information, such as shape and texture, resulting in a more comprehensive representation of the study areas [43]; however, the latter require more detailed parameterization and can be more susceptible to errors in heterogeneous environments [12]. Nonetheless, this study has demonstrated the appropriateness of the method developed by Dalponte and Coomes [28] to overcome these challenges in sugarcane studies.

4.2. Estimating Aboveground Fresh Weight

The results indicated a significant correlation between some predictor variables and aboveground fresh weight; however, some showed high multicollinearity, especially between the SAVI and MSAVI. This occurred because these indices use reflectance from the same set of bands but are based on different mathematical combinations. Canopy volume and height did not correlate well with the AGFW samples, suggesting they may not be as descriptive as vegetation indices based on spectral reflectance. Furthermore, it is well-known that the vegetation indices indirectly represent plants' biophysical variation [44]. However, this could be influenced by the variability of the sample plots and the procedure to measure each plot's aboveground fresh weight. For instance, once each plot was harvested, the top part of each plant (leaves) was removed before weighing. The mean canopy volume and height calculated from LiDAR included that information. Hence, we recommend removing the corresponding slice of LiDAR data points from the top before calculating canopy volumes and heights.

Although the MLR model performed well ($R^2 = 0.737$) relative to sample data, the prediction results were unreliable. The R^2 was relatively lower than those in existing studies that mapped several crop yields using MLR models. For example, Niedbała [45] used the MLR model to predict the winter wheat yield, achieving between 0.518 and 0.5457, indicating moderate results. The study tested three MLR models with large number of predictor variables (13, 17, and 19) and field samples. Abrougui et al. [46] used a large number of field samples and achieved high performance with an MLR model ($R^2 = 0.89$) for agricultural yield prediction. Aleman et al. [13] achieved an R^2 of 0.83 using images acquired from a drone and MLR algorithm when estimating the sugarcane yield in metric ton/ha; their observation was that the model could present a better fit if the training data were more varied and larger in quantity. However, the study's limited number of field samples made it difficult to conclude on the performance. The model was overfitted with a limited number of field samples, as MLR models are known to be sensitive to the number of sample points [47].

Partial least-squares regression reduces the given predictors to a set of uncorrelated components, which will be used for least-squares regression instead of original data [48]. Hence, PLSR is identified as a fast, efficient, and optimal regression method that provides a high performance based on covariance [40]. With four components, this study's PLSR

model provided moderate prediction results. This was mainly due to the multicollinearity of predictor variables and the limited number of samples. However, PLSR is primarily recommended when there are more predictor variables than the number of samples [40]. Although PLSR is commonly used in the chemical, drug, food, and plastic industries [48], increasing applications can be seen in natural resources management and agriculture. Lopez et al. [49] obtained an R^2 of 0.66 and RMSE of 10% when estimating yield losses using the PLSR model, with results similar to those in this study. Jeong et al. [50] indicated that overfitting can be a problem when the training data are not sufficiently varied or large enough, which highlights the importance of a well-designed training dataset that covers a wide range of predictor variability to minimize this problem. Xu et al. [9] showed a significant positive correlation between the height and weight of sugarcane field samples. Their prediction results were also promising.

To improve the accuracy of these models, we recommend increasing the number of samples and considering the spatial variability of sugarcane when selecting plants. Sanches et al. [11] suggested the inflection point for biomass accumulation by sugarcane as a good indicator for use to estimate the yield. Hence, we recommend combining images from different growing stages (multitemporal) to capture the spectral variations over time and decide on the inflection point, thus potentially enhancing the robustness of the results.

5. Conclusions

Sugarcane is one of Costa Rica's main agricultural products and is key in boosting the national economy. Traditionally, the crop yield has been determined by directly weighing the harvest from a sample of the farm and averaging the whole field. However, this is costly and laborious at the production level for large areas; therefore, there is a demand for developing an alternative method that incorporates the use of new technologies such as remote sensing. Hence, this research aimed to develop a remote sensing method for estimating the sugarcane aboveground fresh weight to streamline the production at the farm level.

This study used multispectral images acquired from a Micasense Rededge-P camera and LiDAR data. The aboveground fresh weight measured in the field for sample plots were used for model calibration and validation. The digital terrain and digital surface models extracted from LiDAR data were used to create a canopy height model (CHM). Individual sugarcane tree crowns were delineated using the method developed by Dalponte and Coomes [28]. The extent of each crown area was calculated. The average height of each crown was extracted from the CHM. Two regression models (multiple linear regression (MLR) and partial least-squares regression (PLSR)) were developed to check the correlation between aboveground fresh weight samples and vegetation indices derived from images. The MLR model obtained higher accuracy ($R^2 = 0.7$) compared to the field samples of aboveground fresh weight. However, the PLSR model showed more promising results than the MLR model when predicting aboveground fresh weight over the study area. The accuracy of the prediction raster was calculated as $R^2 = 0.5$ compared to field samples.

This research has highlighted the potential of integrating remote sensing technologies into precision agriculture, thereby improving yield estimation and effective crop management. However, this study faced challenges when assessing the model accuracy and the prediction performance. Despite the successful adaptation of the individual tree crown delineation method, there were limitations mainly due to the limited number of samples collected in the field. Also, there was low correlation between canopy volume, height, and the aboveground fresh weight of samples. This could have been influenced by the variability of the sample plot and the procedure for measuring each plot's aboveground fresh weight. For instance, once each plot was harvested, the top part of each plant (leaves) was removed before weighing. We recommend removing the corresponding slice of LiDAR data points from the top before calculating canopy volumes and heights. Furthermore, we suggest combining images from different growing stages (multitemporal) to capture the spectral variations over time and potentially help decide on the inflection point of

sugarcane biomass production, to enhance the robustness of the results. Lastly, diversifying training samples will help address challenges related to multicollinearity and prediction performance of model accuracy.

Author Contributions: Conceptualization, C.M.V. and K.F.R.; methodology, C.M.V. and M.K.H.; software, C.M.V., K.F.R. and M.K.H.; formal analysis, C.M.V. and M.K.H.; writing—original draft preparation, C.M.V.; writing—review and editing, M.K.H.; supervision, M.K.H.; project administration, K.F.R.; funding acquisition, K.F.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: We would like to express our sincere gratitude to José Torres and the team of the farm “Agrícola El Cántaro” for their invaluable collaboration in the data collection for this study. In addition, we would like to thank our colleagues who have provided support and guidance during the development of this project.

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

References

- Pan, S.; Zabed, H.M.; Wei, Y.; Qi, X. Technoeconomic and environmental perspectives of biofuel production from sugarcane bagasse: Current status, challenges and future outlook. *Ind. Crops Prod.* **2022**, *188*, 115684. [[CrossRef](#)]
- Garcia Tavares, R.; Lakshmanan, P.; Peiter, E.; O’Connell, A.; Caldana, C.; Vicentini, R.; Sérgio Soares, J.; Menossi, M. ScGAI is a key regulator of culm development in sugarcane. *J. Exp. Bot.* **2018**, *69*, 3823–3837. [[CrossRef](#)]
- Waclawovsky, A.J.; Sato, P.M.; Lembke, C.G.; Moore, P.H.; Souza, G.M. Sugarcane for bioenergy production: An assessment of yield and regulation of sucrose content. *Plant Biotechnol. J.* **2010**, *8*, 263–276. [[CrossRef](#)]
- 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](#)]
- Vignola, R.; Poveda, K.; Watler, W.; Vargas, A.; Berrocal, Á. Prácticas efectivas para la reducción de impactos por eventos climáticos: Cultivo de caña de azúcar en Costa Rica. Costa Rica. 2018. Available online: <https://www.mag.go.cr/bibliotecavirtual/F01-8327.pdf> (accessed on 15 October 2023).
- León, J.; Arroyo, N. *Desarrollo Histórico del Sector Agroindustrial de la Caña de Azúcar en el Siglo XX: Aspectos Económicos, Institucionales y Tecnológicos*; Universidad de Costa Rica, Instituto de Investigaciones en Ciencias Económicas: San Pedro, Costa Rica, 2012; Volume 1.
- Jin, X.; Kumar, L.; Li, Z.; Feng, H.; Xu, X.; Yang, G.; Wang, J. A review of data assimilation of remote sensing and crop models. *Eur. J. Agron.* **2018**, *92*, 141–152. [[CrossRef](#)]
- Sumesh, K.C.; Ninsawat, S.; Som-ard, J. Integration of RGB-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. *Comput. Electron. Agric.* **2021**, *180*, 105903. [[CrossRef](#)]
- 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](#)]
- Oliveira, R.P.; Júnior, M.R.B.; Pinto, A.A.; Oliveira, J.L.P.; Zerbato, C.; Furlani, C.E.A. Predicting Sugarcane Biometric Parameters by UAV Multispectral Images and Machine Learning. *Agronomy* **2022**, *12*, 1992. [[CrossRef](#)]
- Sanches, G.M.; Duft, D.G.; Kölln, O.T.; Luciano AC, D.S.; De Castro SG, Q.; Okuno, F.M.; Franco, H.C.J. The potential for RGB images obtained using unmanned aerial vehicle to assess and predict yield in sugarcane fields. *Int. J. Remote Sens.* **2018**, *39*, 5402–5414. [[CrossRef](#)]
- Som-ard, J.; Hossain, M.D.; Ninsawat, S.; Veerachitt, V. Pre-harvest Sugarcane Yield Estimation Using UAV-Based RGB Images and Ground Observation. *Sugar Tech* **2018**, *20*, 645–657. [[CrossRef](#)]
- Aleman, B.; Henriquez, C.; Ramirez, T.; Largaespada, K. Estimación De Rendimiento En El Cultivo De Caña De Azúcar (*Saccharum Officinarum*) A Partir De Fotogrametría Con Vehículos Aéreos No Tripulados (Vant). *Agron. Costarric.* **2021**, *45*, 67–80. [[CrossRef](#)]
- Villareal, M.K.; Tongco, A.F. Remote Sensing Techniques for Classification and Mapping of Sugarcane Growth. *Eng. Technol. Appl. Sci. Res.* **2020**, *10*, 6041–6046. [[CrossRef](#)]
- Sofonia, J.; Shendryk, Y.; Phinn, S.; Roelfsema, C.; Kendoul, F.; Skocaj, D. Monitoring sugarcane growth response to varying nitrogen application rates: A comparison of UAV SLAM LiDAR and photogrammetry. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101878. [[CrossRef](#)]

16. Zhang, Y.; Zhao, D.; Liu, H.; Huang, X.; Deng, J.; Jia, R.; He, X.; Tahir, M.N.; Lan, Y. Research hotspots and frontiers in agricultural multispectral technology: Bibliometrics and scientometrics analysis of the Web of Science. *Front. Plant Sci.* **2022**, *13*, 955340. [[CrossRef](#)]
17. CATSA. Caña de Azucar: Generalidades. Available online: <https://www.catsa.net/service/generalidades-del-area-agricola/> (accessed on 28 January 2024).
18. Torres, J. Presented at the Meeting with the Administrator of Agricola El Cantaro, Cañas, Costa Rica. 24 August 2023.
19. DJI Matrice. Drone DJI Matrice 300 RTK [Divece]. Available online: <https://enterprise.dji.com/matrice-300?site=enterprise&from=nav> (accessed on 20 February 2024).
20. DJI Enterprise. Zenmuse L1 Laser Scanner: Lidar and RGB [Device]. 2023. Available online: <https://enterprise.dji.com/> (accessed on 20 February 2024).
21. AgEagle. Micasense Red Edge-P Camera [Devise]. Available online: <https://ageagle.com/drone-sensors/rededge-p-high-res-multispectral-camera/> (accessed on 20 February 2024).
22. ASPRS. LASER (LAS) FILE FORMAT EXCHANGE ACTIVITIES. The Imaging and Geospatial Information Society. Available online: <https://www.asprs.org/divisions-committees/lidar-division/laser-las-file-format-exchange-activities> (accessed on 22 November 2023).
23. Zaque, W.B.; Rey, L.K.E.; García, L. Obtención de parámetros óptimos en la clasificación de nubes de puntos LiDAR, a partir de sensores aerotransportados. *Av. Investig. Ing.* **2017**, *14*, 9–20. [[CrossRef](#)]
24. GISware Integro. Spatix (Version 023.002) [Software]. Available online: <https://spatix.com/> (accessed on 20 February 2024).
25. Pix4D. Pix4DMapper [Software]. Available online: <https://www.pix4d.com/product/pix4dmapper-photogrammetry-software/> (accessed on 20 February 2024).
26. Zhang, Y.; Zhang, Y.; Zhang, Y.; Li, X. Automatic extraction of dtm from low resolution dsm by twosteps semi-global filtering. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *III-3*, 249–255. [[CrossRef](#)]
27. Roussel, J.; Auty, D.; Boissieu, F.; Sánchez, A. lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications (4.1.0) [R Package]. Available online: <https://cran.r-project.org/web/packages/lidR/index.html> (accessed on 21 February 2024).
28. Dalponte, M.; Coomes, D.A. Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods Ecol. Evol.* **2016**, *7*, 1236–1245. [[CrossRef](#)]
29. Kharuf, S.; Hernández, L.; Orozco, R. Análisis de imágenes multiespectrales adquiridas con vehículos aéreos no tripulados. *Ing. Electrónica Automática Comun.* **2018**, *39*, 79–91. Available online: <http://ref.scielo.org/cd8p4f> (accessed on 9 November 2023).
30. Alemán, B.; Serra, P.; Zabala, A. Modelos para la estimación del rendimiento de la caña de azúcar en Costa Rica con datos de campo e índices de vegetación. *Rev. Teledetección* **2023**, *61*, 1–13. [[CrossRef](#)]
31. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]
32. NV5 Geospatial. Broadband Greenness: Triangular Greenness Index (TGI). Available online: <https://www.nv5geospatialsoftware.com/docs/BroadbandGreenness.html#Triangul> (accessed on 24 April 2024).
33. Todd, J.; Johnson, R. Prediction of Ratoon Sugarcane Family Yield and Selection Using Remote Imagery. *Agronomy* **2021**, *11*, 1273. [[CrossRef](#)]
34. 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](#)]
35. Thompson, C.N.; Guo, W.; Sharma, B.; Ritchie, G.L. Using Normalized Difference Red Edge Index to Assess Maturity in Cotton. *Crop Sci.* **2019**, *59*, 2167–2177. [[CrossRef](#)]
36. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
37. Hunt, E.R.; Daughtry, C.S.T.; Eitel, J.U.H.; Long, D.S. Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index. *Agron. J.* **2011**, *103*, 1090–1099. [[CrossRef](#)]
38. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. [[CrossRef](#)]
39. Abdi, H. Partial Least Squares (PLS) Regression. *Encycl. Res. Methods Soc. Sci.* **2003**, *6*, 792–795.
40. Mevik, B.H.; Wehrens, R. Introduction to the pls Package. University Center for Information Technology, University of Oslo Norway & Biometris, Wageningen University & Research The Netherlands. Available online: <https://cran.r-project.org/web/packages/pls/vignettes/pls-manual.pdf> (accessed on 5 December 2023).
41. Hovde, K.; Mevik, B.; Wehrens, R.; Hiemstra, P. pls: Partial Least Squares and Principal Component Regression (2.8-3). 2023. Available online: <https://cran.r-project.org/web/packages/pls/index.html> (accessed on 21 February 2024).
42. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* **2012**, *118*, 259–272. [[CrossRef](#)]
43. Nasiri, V.; Hawrylo, P.; Janiec, P.; Socha, J. Comparing Object-Based and Pixel-Based Machine Learning Models for Tree-Cutting Detection with PlanetScope Satellite Images: Exploring Model Generalization. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *125*, 103555. [[CrossRef](#)]
44. Banerjee, B.P.; Joshi, S.; Thoday-Kennedy, E.; Pasam, R.K.; Tibbits, J.; Hayden, M.; Spangerberg, G.; Kant, S. High-throughput phenotyping using digital and hyperspectral imaging-derived biomarkers for genotypic nitrogen response. *J. Exp. Bot.* **2020**, *71*, 4604–4615. [[CrossRef](#)]

45. Niedbała, G. Application of multiple linear regression for multi-criteria yield prediction of winter wheat. *J. Res. Appl. Agric. Eng.* **2018**, *63*, 125–131. Available online: <https://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-38e3a710-0dbc-4ba2-80ee-3e3be6dcb15f> (accessed on 22 February 2024).
46. 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 Tillage Res.* **2019**, *190*, 202–208. [CrossRef]
47. Nahhas, R. Introduction to Regression Methods for Public Health Using R. Sensitivity Analysis. Available online: <https://www.bookdown.org/rwnahhas/RMPH/mlr-sensitivity.html> (accessed on 28 February 2024).
48. Minitab. What Is Partial Least Squares Regression. Available online: <https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/regression/supporting-topics/partial-least-squares-regression/what-is-partial-least-squares-regression/> (accessed on 27 February 2024).
49. Lopez-Fornieles, E.; Brunel, G.; Rancon, F.; Gaci, B.; Metz, M.; Devaux, N.; Taylor, J.; Tisseyre, B.; Roger, J.M. Potential of Multiway PLS (N-PLS) Regression Method to Analyse Time-Series of Multispectral Images: A Case Study in Agriculture. *Remote Sens.* **2022**, *14*, 216. [CrossRef]
50. 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.; et al. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *11*, e0156571. [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.