land

MDPI

Article

Estimating Groundnut Yield in Smallholder Agriculture Systems Using PlanetScope Data

Daniel Kpienbaareh [1], Kamaldeen Mohammed [2,*], Isaac Luginaah [2], Jinfei Wang [2], Rachel Bezner Kerr [3], Esther Lupafya [4] and Laifolo Dakishoni [4]

[1] Department of Geography, Geology and the Environment, Illinois State University, 104 Felmley Hall, Normal, IL 61790-4000, USA
[2] Department of Geography, University of Western Ontario, 151 Richmond St, London, ON N6A 3K7, Canada
[3] Department of Global Development, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY 14853, USA
[4] Soils, Food and Healthy Communities (SFHC), Ekwendeni P.O. Box 36, Malawi
* Correspondence: kmoham27@uwo.ca

Abstract: Crop yield is related to household food security and community resilience, especially in smallholder agricultural systems. As such, it is crucial to accurately estimate within-season yield in order to provide critical information for farm management and decision making. Therefore, the primary objective of this paper is to assess the most appropriate method, indices, and growth stage for predicting the groundnut yield in smallholder agricultural systems in northern Malawi. We have estimated the yield of groundnut in two smallholder farms using the observed yield and vegetation indices (VIs), which were derived from multitemporal PlanetScope satellite data. Simple linear, multiple linear (MLR), and random forest (RF) regressions were applied for the prediction. The leave-one-out cross-validation method was used to validate the models. The results showed that (i) of the modelling approaches, the RF model using the five most important variables (RF5) was the best approach for predicting the groundnut yield, with a coefficient of determination (R^2) of 0.96 and a root mean square error (RMSE) of 0.29 kg/ha, followed by the MLR model (R^2 = 0.84, RMSE = 0.84 kg/ha); in addition, (ii) the best within-season stage to accurately predict groundnut yield is during the R5/beginning seed stage. The RF5 model was used to estimate the yield for four different farms. The estimated yields were compared with the total reported yields from the farms. The results revealed that the RF5 model generally accurately estimated the groundnut yields, with the margins of error ranging between 0.85% and 11%. The errors are within the post-harvest loss margins in Malawi. The results indicate that the observed yield and VIs, which were derived from open-source remote sensing data, can be applied to estimate yield in order to facilitate farming and food security planning.

Keywords: food security; yield prediction; Malawi; random forest regression; PlanetScope; vegetation indices

check for updates

Citation: Kpienbaareh, D.; Mohammed, K.; Luginaah, I.; Wang, J.; Bezner Kerr, R.; Lupafya, E.; Dakishoni, L. Estimating Groundnut Yield in Smallholder Agriculture Systems Using PlanetScope Data. *Land* **2022**, *11*, 1752. https://doi.org/10.3390/land11101752

Academic Editors: Liangjie Xin and Xue Wang

Received: 12 September 2022
Accepted: 5 October 2022
Published: 9 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

Groundnuts, or peanuts (*Arachis hypogaea* L.), which are commonly called the poor man's nut, are an important oilseed and food crop for the majority of the global population. The global yield of groundnuts was estimated to be 47.09 million metric tons in 2019 [1]. Although they are native to South America, China and India together produce most of the world's groundnuts, accounting for about 60% of the production and 52% of the crop area [2]. Recent data show that more than half of the groundnuts that are produced globally are processed into oil, but a substantial quantity of groundnuts that are produced in developing countries are marketed domestically [1]. Groundnuts also play a crucial role in smallholder agricultural systems, including crop rotation, intercropping, and soil health

through nitrogen fixation [3]. International trade in groundnuts is mainly in the form of in-shell (pods), shelled (kernels), and meal (cake) [4].

In sub-Saharan Africa (SSA), groundnut yields are generally low, at about 964 kg/ha, which is far less than the potential yield of up to 3500 kg/ha [5]. The low yield levels of groundnuts among smallholder farmers in SSA are attributable to various factors, including abiotic factors (drought and low soil fertility) and biotic factors (including pests, such as aphids and termites, and diseases, such as groundnut rosette disease, leaf spot, and rust [6]). In Malawi, as in other countries in SSA, groundnuts are mostly cultivated on a smallholder basis and account for up to 25% of household agricultural income [7]. The AICC (2016) reports that in the last decade, groundnut production has been increasing, even though recent trends show decreasing yield trends. Additionally, during periods of increasing yield (e.g., from 2009 to 2012 production increased from 985 kg/ha to 1083 kg/ha), the rate of increase remained significantly below the realizable potential yield of 2500 kg/ha [8]. The underperformance of the groundnut subsector in Malawi is attributable to several factors, including frequent droughts, low soil fertility, the unavailability of ready markets for groundnuts and their products, and poor insect, pest, and disease control [7].

In the context of low yields resulting from multiple factors, which are compounded by increasing food and nutrition insecurity in SSA in general, and in Malawi in particular, monitoring and estimating groundnut yields is important for making effective farming decisions [9], for predicting market behavior, and for formulating government policies in order to help to improve income and address food insecurity [10]. Groundnuts contain high amounts of protein, which makes them an essential component in the diet of households, especially in places such as Malawi where the nutrition insecurity rate is about 37% among children [11]. There is, thus, an increasing need for adopting techniques that allow for effective monitoring of groundnuts on small farms and the timely forecasting of their yield for effective farm management, food security planning, and marketing decisions. Remote sensing has emerged as a cost-effective effective tool for predicting crop yield in a timely and accurate manner [12,13]. Remote sensing techniques have been used in several studies in order to predict the yield of various crops, including wheat [9,14], maize [15,16], rice [17,18], soybean [10,19], and potato [20,21].

These prediction analyses are often conducted using vegetation indices (VIs) that are derived from either satellite images or unmanned aerial vehicles (UAVs). VIs are mathematical models that are widely used to evaluate canopy characteristics, such as vigor, growth dynamics, nutrient levels, and chlorophyll content in vegetation [22,23]. The study of [24] describes VIs as transformations of spectral bands and wavelengths to allow for the interpretation of variations in the canopy structure and photosynthetic activities. Since crop yield is a function of such canopy characteristics, VIs can be used to estimate their yield [25]. Several studies have taken advantage of the relationship between different VIs and crop canopy characteristics in order to estimate crop yield. For instance, [14] found good predictive ability between the soil adjusted vegetation index (SAVI), the normalized difference vegetation index (NDVI), and soybean grain yield in Brazil, while [26] found a positive linear relationship between the NDVI and cotton yield. The study of [27] also found a strong positive correlation between the green normalized vegetation index (GNDVI) and sugar cane crop. The leaf area index (LAI), which corresponds to one half of the total green leaf area per unit horizontal ground surface area [28], has been widely used in order to estimate crop yield. Indeed, [29] concluded that the LAI is a better measurement for plant density and growth compared to the other VIs. As such, it is an important parameter reflecting the biochemical status and physical processes of the crop [30]. The LAI can provide crucial theoretical indicators for crop management, water and fertilizer regulation, growth and health monitoring, and yield estimation. Several studies have, thus, used the LAI as a variable in order to estimate crop yield. For instance, [31] monitored and estimated the yield of wheat using the LAI in Jiangsu Province, China, whereas [32] used the LAI to predict maize yield with high accuracy in China. In terms of the prediction methods, a variety of statistical analysis methods, including simple linear regression, multiple linear

regression, and non-linear regression, have been used [31,33]; however, machine learning methods, such as the random forest regression, support vector regression, deep learning, and neural networks, have gained importance [34–37].

While these studies have generally shown a good predictive capacity for crop yield, it is important to highlight that variations exist in the behavior of the computed indices based on the location, the climatic conditions (e.g., rainfall, temperature, and illumination), and the farm management practices. These factors could likely affect the relationship between the index values and the crop yield. For instance, [38] found that there is a strong correlation between the illumination conditions and the enhanced vegetation index (EVI) when they are observed temporally. The study of [39] found spatio-temporal variations of the NDVI, the SAVI, and the EVI in response to rainfall and temperature in semi-arid regions, whereas [40] found spatial differences in responses of forests to the El Niño–Southern Oscillation (ENSO). Furthermore, different soil enhancement methods affect the crop phenological stages [41,42]. In this case, VIs would likely vary depending on the management practices on the farm, such as the manure application, the weed control, and the disease and pest control. In addition, in SSA, where farms are generally small [43], it is important to understand the characteristics of VIs for their potential for estimating crop yield. There is, however, a paucity of literature on the use of such VIs for estimating groundnut yield, especially in the sub-tropical region of SSA, despite the significance of groundnuts to household nutrition. Present studies have not exhaustively explored the conditions within the sub-tropical climatic zone.

Consequently, there is a need for yield prediction analysis for farming systems in SSA. Understanding crop productivity in SSA will allow for a deeper understanding of the behavior of VIs for developing models that are context-specific and can more accurately predict the groundnut yield. This study aims to demonstrate the potential for using satellite-derived VIs in order to conduct a pre-harvest yield prediction in smallholder agricultural systems in the sub-tropical climatic zone of SSA, such as in northern Malawi. Our goal is to identify the ideal growth stage and the appropriate method to predict the groundnut yield within the season in northern Malawi. While some studies have predicted the yield of other crops using UAVs and Sentinel-2 data in large-scale commercial farming systems in SSA [44,45], similar studies in smallholder contexts using PlanetScope data to predict yield are lacking in smallholder agricultural systems. Yet, such studies are needed in order to guide smallholder farmers and local agriculture officers to plan for perennial food challenges. This present study contributes to the literature on the application of geospatial techniques for supporting smallholder food security planning and building community resilience.

2. Materials and Methods

2.1. Study Sites

The study was conducted on two farmlands located ~25 km apart in the Mzimba district in northern Malawi (Figure 1). Site 1 is located on latitude −11°23′41.63″ S and longitude 33°45′56.11″ E. Site 2 is located on latitude −11°15′58.24″ S and longitude 33°50′53.46″ E. The soils in the district are uniformly moderately fertile, with generally medium to light texture, mostly sandy-loam, and loamy, with moderate to good drainage [46,47]. The sub-tropical climate type prevails in the area and is characterized by average monthly maximum summer temperatures ranging from 27 °C to 33 °C. Such temperatures are suitable for groundnut cultivation [7]. In the winter months, the temperatures drop to between 0 °C and 10 °C, with June and July being the coldest months. The unimodal rainfall pattern in the area starts in November/December and ends in April/May, with an annual range from 650 mm to 1300 mm [48]. The agriculture in the district is primarily small scale and is highly dependent on the unimodal rainfall pattern, making it persistently vulnerable to seasonal food insecurity resulting from rainfall variability. The farmers in the district primarily cultivate on small farms (i.e., farms with an average size of two hectares). As such, smallholder farming is crucial for local food supply and nutrition

in the districts and other rural areas in Malawi, where households essentially eat what they produce. Generally, most crops are planted in early to late December and are harvested in April/May [46]. The government's Farm Input Subsidy Program (FISP) has increased the reliance on chemical fertilizers to enhance soil fertility, even though most smallholder farmers have difficulty affording it [49]. The heavy reliance on fertilizer has not led to increased crop yield because fertilizers do not address underlying issues with soil erosion, deteriorating soil organic matter, and declining soil microorganisms, which all influence soil health and, ultimately, crop productivity [50,51].

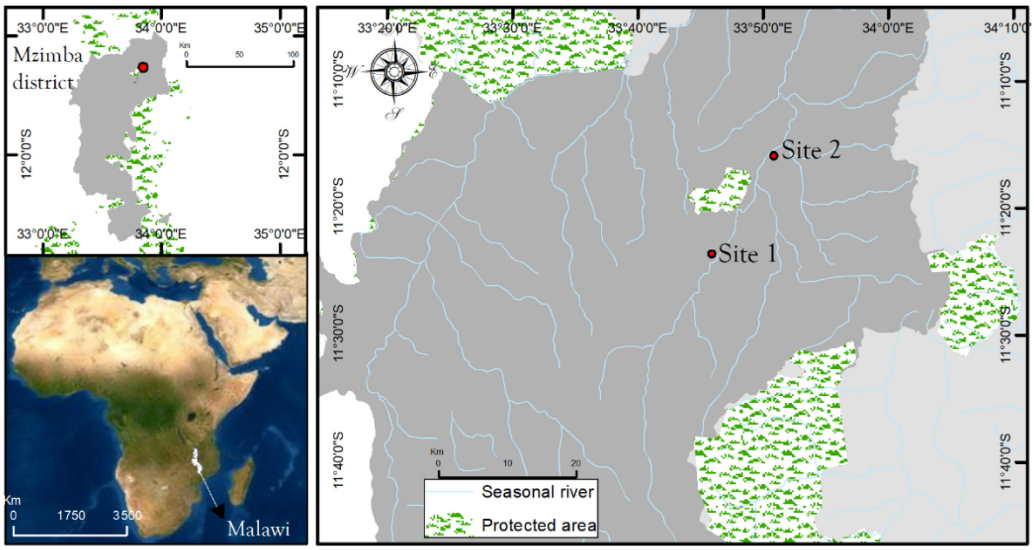

Figure 1. Location of study sites. The base map of the inset map of Africa is from ESRI.

2.2. Study Design

The two farms on which the yield was measured were selected from two rural communities in the Mzimba district. The farms were small, with areas of 0.78 ha and 0.83 ha for Sites 1 and 2, respectively. The farm sizes were measured by walking around the boundaries of the farms with a handheld global positioning system (GPS) device and the .gpx files were converted to feature polygons. Each field was planted with 20 kg of CG 7 groundnut variety. CG 7 is a high-performing variety that is being promoted for cultivation in Malawi because of its high yield returns [7]. Both farmers were provided with the same seed variety for uniformity so that any variation in the yield could be explained by other factors, including the soils or the behavior of VIs. Six sampling points were identified and geolocated on each field, with pegs placed at the locations for easy identification. The number of sampling points was decided based on the size of the farms and the spatial resolution of the satellite image that was to be used. Since VIs are computed on a per-pixel basis [52] sampling more than one point in the same pixel would be redundant. The soil on the farms in both locations was similar [47], therefore, any variation in the yield may be explained by factors other than the soil. The locations of the sampling points on the fields were determined by the width and length of the fields.

The planting was carried out on 8–9 December 2019 on Sites 1 and 2, respectively, during the 2019/2020 planting season. A planting distance of 50 cm by 50 cm was used in both fields, although the spacing on Site 1 was more erratic. A 5 m^2 area was delineated around each sample point, within which hemispherical photographs were captured during the fieldwork. Figure 2 shows portions of the field that were captured on the specified dates.

Figure 2. Photographs of the groundnut fields were captured during the flowering stages of crop growth.

3. Data Collection and Processing

3.1. Digital Hemispherical Photographs' Capture and Processing

Between 24 December and 27 February 2020, nine field campaigns were conducted in order to capture the crop canopy characteristics. By 27 February the crops had passed the flowering stage and the peak LAI was attained. The peak LAI has been used to accurately predict crop yield in prior studies [53,54]. During each field visit, downward viewing digital hemispherical photographs (DHPs) were captured using a NIKON D500 camera that was fitted with an AF DX Fisheye-Nikkor lens. Fourteen DHPs were captured within each of the 5 m^2 areas around each geolocated sample point. The field campaigns were scheduled every seven days, and whenever there was cloud-free weather, in order to facilitate the selection of satellite images for analysis. The photographs were captured early in the morning or late in the afternoon, under diffuse light conditions, in order to avoid shadow effects. In addition to the DHPs, from each field, three pictures were taken from the above-head perspective, across, and along the row in order to allow for a visual inspection and documentation of the crop characteristics and to observe the phenological stages of the crops. Weeds are common on farms in the area and can reduce the yield, conflate with vegetation index values, and affect the accuracy of the LAI and VIs that are extracted from the remote sensing images. Farmers were, therefore, encouraged to weed the farms biweekly in order to remove all of the weeds. At each field visit, management practices, such as pest/disease control and manure or fertilizer application, were recorded in a fieldwork book.

The DHPs were processed using the CAN-EYE software version 6.4.9 [55]. The software allows for up to twenty hemispherical photographs to be processed at the same time. The photographs were analyzed iteratively in the red, green, and blue color space. The data processing was limited to view zenith angles that were smaller than 60° in order to minimize the number of mixed pixels [56]. We maintained the default values that were set by the CAN-EYE software for the zenith (2.5°) and azimuthal (5°) angle resolutions because those angles produced the best results. The CAN-EYE software estimates the LAI using the measures of the gap fraction, which is the transmission of light through the canopy when considering the vegetation elements as opaque [57]. The software uses a look-up table (LUT) in order to estimate the LAI from the gap fraction, assuming a random distribution of phyto-elements within the canopy that is without clumping [58]. The CAN-EYE software allows for the computation of an effective LAI (LAI_{eff}) from the gap fraction that is estimated from the hemispherical photographs by averaging the gap fraction across the azimuth and the photographs for each zenithal ring. The true LAI that was measured using a planimeter was derived from the LAI_{eff} in the following function:

$$LAI = \frac{LAI_{eff}}{\lambda_o} \quad (1)$$

where λ_o is the aggregation or dispersion parameter [58] or the clumping index [28].

3.2. Groundnut Yield Data

The harvesting of the groundnuts was carried out on 6 May 2020, at Site 1 and 4 May 2020, at Site 2. The harvesting was performed when the farmers were ready to harvest the whole field. In order to assess the yield, all of the crops within the 5 m^2 area around each sampling point were harvested by digging them up with a hoe or pulling the entire plant out. The number of plants in each 5 m^2 area was counted and the pods were hand plucked. Both the pods and the crop residue were weighed (in kilograms, kg). The pods and the residue were weighed again after being sun-dried for 10 to 14 days on farmers' property. The dried pods were shelled, their moisture content was measured, and both the shelled nut and the dried shells were weighed. The unshelled weight was used for the modelling, as has been carried out in other studies [45]. The crops were returned to the farmers after the yield assessment.

3.3. Satellite Image Acquisition and Vegetation Index Selection

High spatial resolution (3-m resolution) multispectral surface reflectance images were downloaded freely under the Education and Research programme of Planet Labs. The PlanetScope constellation of satellites presently has about 130+ CubeSats (4-kg satellites) that are operated by Planet Labs [59] The majority of these CubeSats are in a sun-synchronous orbit, and cross the equator, between 9:30 and 11:30 (local solar time) [60]. The PlanetScope images have the following four spectral bands: blue (455–515 nm), green (500–590 nm), red (590–670 nm), and near-infrared (NIR, 780–860 nm). The Level-3B surface reflectance products that have been atmospherically corrected by Planet Labs using the 6S radiative transfer model with ancillary data from moderate resolution imaging spectroradiometer (MODIS) [60,61] were used for this study. Such high temporal and high spatial resolution data are useful for computing plant biomass and crop biophysical characteristics for precision agriculture purposes [62,63]. The 10 PlanetScope images that have been used for this analysis (Table 1) were selected based on the availability of the images, as there was dense cloud cover in the study location [64]. The images that were captured on dates that were closer to the harvest date were not included because of our goal of forecasting the yield before harvest. Moreover, the images at the start of the germination period were excluded because of the high reflectance from the exposed soil during the inchoate stages of crop growth. High soil reflectance affects the computed VI values [65,66]. The goal was to extract the Vis from a combination of all of the images and to use the model that was developed to predict the yield by using a single image that was captured within the growing season. The images were, therefore, co-registered using the geoscience extended flow optical Lucas–Kanade iterative [67] method in the sentinel application platform (SNAP) toolbox. The GeFolki algorithm allows the registration of images in a nonparametric and dense way. The images were then stacked into a single multi-band image using the collocation tool in the SNAP. After stacking all of the images into a multi-band image, a subset was taken using a defined polygon covering the location of the farmlands. The subset images were then used for deriving the VIs.

Table 1. Satellite data used for the analysis. All images were at 3-m spatial resolution; they had 4 bands each. All image dates were in 2020.

Site 1	Site 2	Testing Data	Yield Prediction
February 22	January 2	February 23	March 7
March 15	February 4		March 31
April 1	February 23		
	March 7		
	March 31		
	April 7		

A total of eight spectral VIs and the in situ LAI were explored in order to determine their potential for predicting the groundnut yield. These indices were chosen based on the capacity of the images that were available to compute them, therefore, only the indices that can be derived from the visible band and the NIR band were chosen for the analysis. Table 2 presents a description and the authorship of the VIs.

Table 2. Details of selected vegetation indices used to predict groundnut yields.

Index	Equation	Reference	Description
Green Normalized Difference Vegetation Index (GNDVI)	$\frac{nir - green}{nir + green}$	[68]	More sensitive than NDVI to identifying different concentration rates of chlorophyll.
Normalized Difference Vegetation Index (NDVI)	$\frac{nir - red}{nir + red}$	[69]	Has the capacity for assessing regional and global vegetation fluctuations. It is more sensitive to reflectance from soil background and tends to saturate for high biomass canopies.
Soil Adjusted Vegetation Index (SAVI)	$\frac{(nir - red)}{(nir + red + L)} \times (1 + L)$	[70]	Used to correct NDVI for the influence of soil brightness in areas with low vegetative cover.
Modified Soil Adjusted Vegetation Index 2 (MSAVI2)	$\frac{2 \times nir + 1 - \sqrt{(2nir + 1)^2 - 8 \times (nir - red)}}{2}$	[71]	Does not rely on a soil line principle and has a simpler algorithm. It is mainly applied in plant growth analysis.
Ratio Vegetation Index (RVI)	$\frac{nir}{red}$	[72]	Sensitive to green vegetation and is highly correlated with LAI and leaf biomass.
Difference Vegetation Index (DVI)	$nir - red$	[69]	Very sensitive to the amount of vegetation and changes in the soil background.
Transformed Normalized Difference Vegetation Index (TNDVI)	$\sqrt{\left[\frac{nir - red}{nir + red}\right] + 0.5}$	[73]	Indicates a better correlation between the amount of green biomass that is found in a pixel.
Weighted Difference Vegetation Index (WDVI)	$nir - \lambda \times red$	[74]	Very sensitive to atmospheric variations and has a correction factor on the slope of the soil line.

L is the soil brightness correction factor, which was set at 0.5 because of the intermediate vegetation cover [71] that was provided by the groundnut plants, λ is the slope of the soil line set at 1.5.

4. Analytical Approach

Three modelling methods were used in order to identify the best approach for predicting the within-season groundnut yield. First, simple linear regressions (SLR) were used to fit the yield data with the individual VIs and the in situ LAI. Previous studies have shown that VIs, including the NDVI, the RVI, the GNDVI, the SAVI, and the MSAVI2, are good predictors of the yields for various crops [10,14,75,76]. The in situ LAI has also been found to be a good predictor of crop yields in several studies using SLR [54,77,78]. Secondly, a random forest (RF) machine learning regressions method was used in order to explore the possible non-linear relationships between the yield, the in situ LAI, and the VIs. The RF regression is a bagging method that is based on CART (classification and regression tree) that employs recursive partitioning in order to divide data into many homogenous subsets (the regression trees) [79]. The algorithm then averages the results of all of the trees. Each tree is independently grown to its maximum size based on a bootstrap sample from the training data set (about 70%) without any pruning. In each tree, RF uses randomness in the regression process by selecting a random subset of variables in order to determine the split (the random selection of the variables) at each node [80]. In each tree, the ensemble predicts the data that are not in the tree (the out-of-bag, OOB data, approximately 30%), and by calculating the difference in the mean square errors between the OOB data and the data that are used to grow the regression trees, the RF algorithm gives an error of prediction that is called the OOB error of estimate for each variable [79,80]. The OOB error measures the importance of the variables by comparing how much the OOB error of estimate increases when a variable is permuted while all of the others are left unchanged [81]. According

to [79], the variable importance is evaluated based on how much worse the prediction would be if the data for that variable were permuted randomly. In this study, the variable importance plot has been used in order to determine the most influential variables for predicting the groundnut yields. Two RF models were fitted, one with all of the VIs (RF All model) and the second with only the five most influential VIs (RF5 model), which were ranked by the variance importance plot in the varImplot() function in R. Random forests are easy to train; they have a low sensibility to outliers, a high computational efficiency, and a robustness against over-fitting [82]. RF regression modelling was performed using the "randomForest" package. In the RF models, the *ntree* was set at 500 and the *mtry* was set to 3. Tuning the *ntree* and the *mtry* parameters did not change the results in any way, therefore, these default values were maintained for the RF All and RF5 models.

Finally, a multiple linear regression (MLR) was fitted in order to determine the linear relationship between the observed yield data and multiple variables. In order to address the problem of multicollinearity, the Pearson correlation coefficient (PCC) of the relationship between the VIs was examined in order to determine the strength of the correlation among them. The PCCs that were higher than 0.70 showed possible collinearity with other predictors [83]. The predictors with PCC > 0.70 were eliminated from the MLR. All of the analyses were performed using the R statistical package, version 4.0.2 [84]. The linear regression models were built using the lm() function.

The leave-one-out cross-validation (LOOCV) method [85] was used in order to evaluate the model accuracy. The LOOCV is an effective technique for evaluating the generalizability of regression models and is especially useful in situations with a low amount of field-measured data [86,87]. The LOOCV allows for the use of the same data for calibration and validation. In the LOOCV method, the data are divided into k samples (k = total number of samples that were used for the analysis), and the samples are then removed one by one. The models were calibrated k times using all of the k samples, except for the omitted one, and were used to predict the excluded sample. The R^2 (Equation (2)) and root mean square error (RMSE) (Equation (3)) metrics were used to evaluate the model accuracy. Mathematically, a higher R^2 and a lower RMSE represent better model accuracy [88]. The R^2 metric represents the proportion of the variance for a dependent variable that is explained by a predictor variable or the variables that are in a regression model, while the RMSE characterizes the mean differences between the observed and the estimated values. The R^2 and RMSE values were computed as follows:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \overline{y})^2} \quad (2)$$

where y_i is the associated observed value in the dataset or is formed in a vector as $y = [y_1, \ldots . y_n]^T$, $\hat{y}$ is the predicted value of the associated y_i, and $\overline{y}$ is the mean value of the observed data.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}} \quad (3)$$

where N is the number of samples and $Predicted_i$ is the predicted value for the ith sample point. $Actual_i$ is the actual value of the ith sample point. The RMSE can be interpreted more precisely when it is compared to the overall value range of the input variables.

The model generating the lowest RMSE from the cross-validation process was used to generate the yield prediction maps for two growth stages of groundnuts. These were proposed by [89] to be the R3/beginning pod stage and the R5/beginning seed stage, or codes 71 and 79 in the development of the fruit and seed phenological stages, respectively of the BBCH-scale [90]. The two stages are critical points in the phenological cycle of groundnuts that eventually determine the final yield. The raster:predict() function in R Studio was applied in order to predict the yield per pixel for the two stages using the satellite images that are described in Table 2. The writeRaster() function was used to generate the prediction maps in the .tif format. The final maps were visualized, and the

cartographic elements were added using ArcGIS 10.7. The sampling points were used to extract the yield values and the results were compared with the observed yields. The R^2 and RMSE metrics were used to assess the accuracy of the prediction.

Additionally, in order to test the usefulness of the model in providing critical information for household yields, the model was used to estimate the yield of four farms (labelled A, B, C, and D) near the study site. The zonal statistics tool in ArcGIS 10.7 was used to extract the total yield per farm and the results were compared with the total yields that were reported by the farmers of the four fields. The differences between the two values were computed in order to determine the degree of accuracy of the model that was used.

5. Results

5.1. Descriptive Statistics for Yield and In Situ LAI Data

There were variations in the groundnut yield for the respective sampling locations on the two fields. The average yield from all of the sample points was 1.96 ± 0.50 kg/ha (minimum = 1.28 kg/ha, maximum = 2.56 kg/ha) for Site 1 and 4.25 ± 1.66 kg/ha (minimum = 1.78 kg/ha, maximum = 6.61 kg/ha) for Site 2. Figure 3 shows the within-field yield variations, with sample point B recording the lowest and sample point F recording the highest yield of the assessed groundnut.

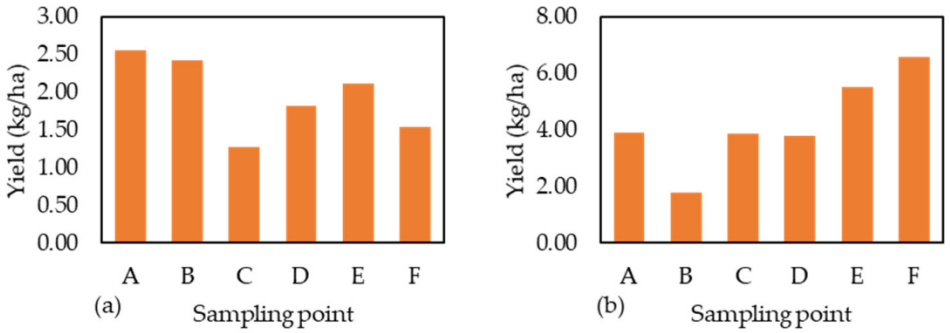

Figure 3. Within-field variations of groundnut yield for (**a**) Site 1 and (**b**) Site 2.

The in situ LAI for both of the fields showed similar variations. The average in situ LAI for the period from 24 December 2019, to 27 February 2020, for Site 1 was 0.15 ± 0.02 m^2/m^2 (minimum = 0.12 m^2/m^2, maximum = 0.17 m^2/m^2) whereas that of Site 2 was 0.48 ± 0.10 m^2/m^2 (minimum = 0.34 m^2/m^2, maximum = 0.56 m^2/m^2).

5.2. Relationship between Average Observed Yield and Predictor Variables

Table 3 presents the results of the simple linear regression analysis between the seasonal average groundnut yield and the nine predictor variables. For Site 1, the results show R^2 values ranging between 0.21 and 0.45, with the in situ LAI generating the highest R^2 of 0.45 and the GNDVI producing the lowest R^2 of 0.21. The results were not significant at the 95% confidence level, indicating a poor predictive power for the data on Site 1.

On the other hand, all of the VIs showed a strong relationship with the groundnut yield for Site 2. The R^2 values ranged from 0.56 (DVI) to 0.89 (WNDVI). The relationship between the yield and the VIs varied for the two sites, likely because the number of images that were combined in order to compute the VIs for Site 1 were only three, compared to six images for Site 2. Figure 4 shows the relationship between the observed and the predicted groundnut yield for each of the nine variables based on the SLR, using data for only Site 2. The RMSE values that were derived from comparing the observed yield with the predicted yield ranged from 0.32 kg/ha for the GNDVI to 0.82 kg/ha (LAI). Since low RMSE values represent a better accuracy, the results suggest the GNDVI is the best single VI for predicting the groundnut yields in such smallholder settings. Generally, the result suggests that all of the indices can predict the yield to a reasonable degree of accuracy. All of the coefficients for Site 2 were significant at the 95% confidence level.

Table 3. R^2 indicating the relationship between yield (kg/ha) and the predictor variables.

	Site 1	Site 2
Variable	R^2	R^2
LAI	0.45	0.70
GNDVI	0.21	0.70
NDVI	0.26	0.65
SAVI	0.26	0.79
DVI	0.26	0.56
MSAVI2	0.25	0.69
RVI	0.27	0.73
TNDVI	0.26	0.59
WNDVI	0.27	0.89

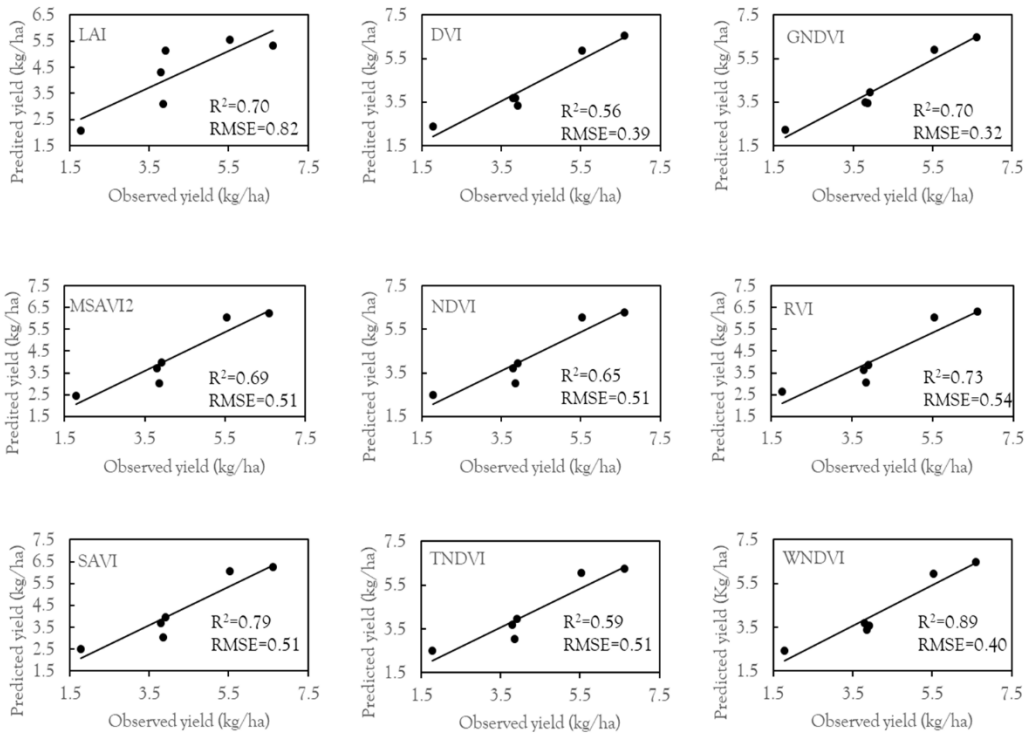

Figure 4. Groundnut yield (kg/ha) was predicted from the observed groundnut yield and the nine predictor variables for Site 2.

Only three of the VIs fitted the MLR model due to multicollinearity. The results in Table 4 show a good relationship between the yield and the TNDVI, the NDVI, and the RVI (R^2 = 0.98, RMSE of 0.18 kg/ha). The variable importance plot (Figure 5), which was computed in order to identify the most important variables, showed that the TNDVI was the most influential variable for predicting the groundnut yield, while the DVI was the least influential variable. Table 4 shows that the RF5 model resulting from the five most influential variables (TNDVI, NDVI, RVI, MSAVI2, and SAVI) from the variance importance plot produced a higher R^2 (0.82) and a lower RMSE (1.00 kg/ha) and was statistically significant ($p < 0.01$, $\alpha < 0.05$). On the other hand, the RF All model with all nine of the predictors had a relatively higher RMSE (1.23 kg/ha), a lower R^2 (0.35), and was not statistically significant ($p = 0.22$, $\alpha < 0.05$), implying the less important variables degraded the RF All model. The RF regression results suggest that the models that were fitted with only the most influential variables were more accurate for predicting the groundnut yield. Overall, the results suggest that the MLR model outperformed the RF All and RF5 models in the calibration process, even though it used fewer variables.

Table 4. Statistics of model calibration using MLR and RF for Site 2.

Variables	Model	R^2	RMSE	*p*-Value
All variables	RF All	0.35	1.23	0.22
TNDVI, NDVI, RVI, MSAVI2, and SAVI	RF5	0.82	1.00	0.01 **
TNDVI, NDVI, and RVI	MLR	0.98	0.18	0.02 **

** Model is significant at the 95% confidence level.

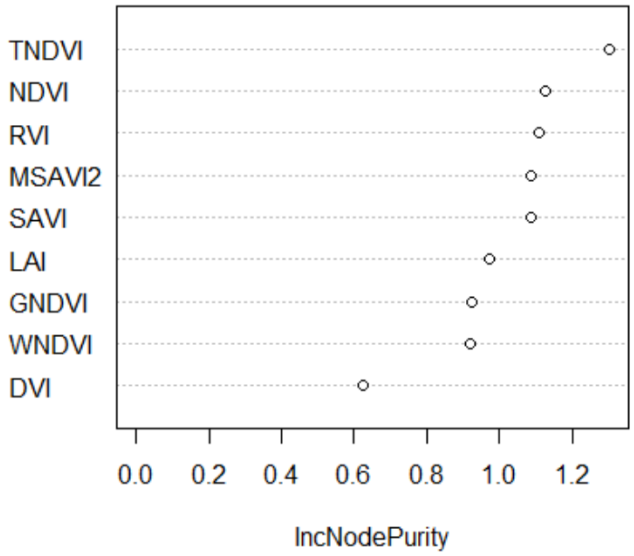

Figure 5. Variance importance plot using varImpPlot() in R Studio for study Site 2. Higher IncNodePurity signifies more contribution of the variable to the model.

Table 5 shows the results of the LOOCV validation for the MLR and RF models. The results show the RF5 model to have an R^2 of 0.96 ($p < 0.001$, $\alpha < 0.05$) and an RMSE of 0.29 kg/ha. The results for the MLR model produced an R^2 of 0.84 and the RMSE was 0.84 kg/ha. Additionally, the coefficients for the variables were statistically significant ($p = 0.02$, $\alpha < 0.05$). The RF All model in Table 4 was not validated because it produced insignificant coefficients during the calibration stage. The results suggest that the RF5 model outperformed the MLR during the validation.

Table 5. Statistics for the LOOCV of the statistically significant models for Site.

Variables	Method	R^2	RMSE	*p*-Value
TNDVI, NDVI, RVI, MSAVI2, and SAVI	RF5	0.96	0.29	0.001 **
TNDVI, NDVI, and RVI	MLR	0.84	0.84	0.02 **

** Model is significant at 95% confidence level.

Table 6 shows the results of the application of the two models on different PlanetScope images that were chosen to coincide with the R3/beginning pod and the R5/beginning seed growth stage. The results from this test were compared with the observed yield from experimental Site 2. Only the RF5 model produced significant results ($p < 0.001$, $\alpha < 0.05$); the highest R^2 was 0.95 and lowest RMSE was 0.35 kg/ha during the R5 growth stage/code 79 on the BBCH-scale phenological stage. The models were not statistically significant for the R3 growth stage, and they had high RMSEs and low R^2 values compared to the R5 stage (Table 6). The results suggest that the best stage of growth for the within-season prediction of the groundnut yield in the study location, in order to obtain accurate results, is during the beginning seed/R5 stage when using a random forest regression model with relevant VIs.

Table 6. Statistics of the relationship between observed yield and predicted yield for Site 2.

Date	Growth Stage	Model	R^2	RMSE (kg/ha)	*p*-Value
7 March 2020	R3/beginning pod stage	MLR	0.57	0.99	0.08
		RF5	0.59	0.97	0.070
31 March 2020	R5/beginning seed stage	MLR	0.70	0.66	0.370
		RF5	0.95	0.35	0.001 **

** Model is significant at 95% confidence level.

Figure 6 presents the distribution of the predicted groundnut yield per pixel for the two growth stages for which there were cloud-free satellite images. Similar to the observed yield, the predicted results have revealed variations at different parts of Site 2. Note the relationship between the sampling points in Figures 6b,c and 3b above. The sampling locations where the observed yields were low also appeared to be low on the prediction maps, indicating a significant correlation between the predictions and the observed yield during the R5 stage. The areas of high yield also coincide with the areas with high predicted values. Taken together, Table 6 and Figure 6 explain the potential of using satellite-based remote sensing data from PlanetScope for predicting the groundnut yield on small farms in northern Malawi.

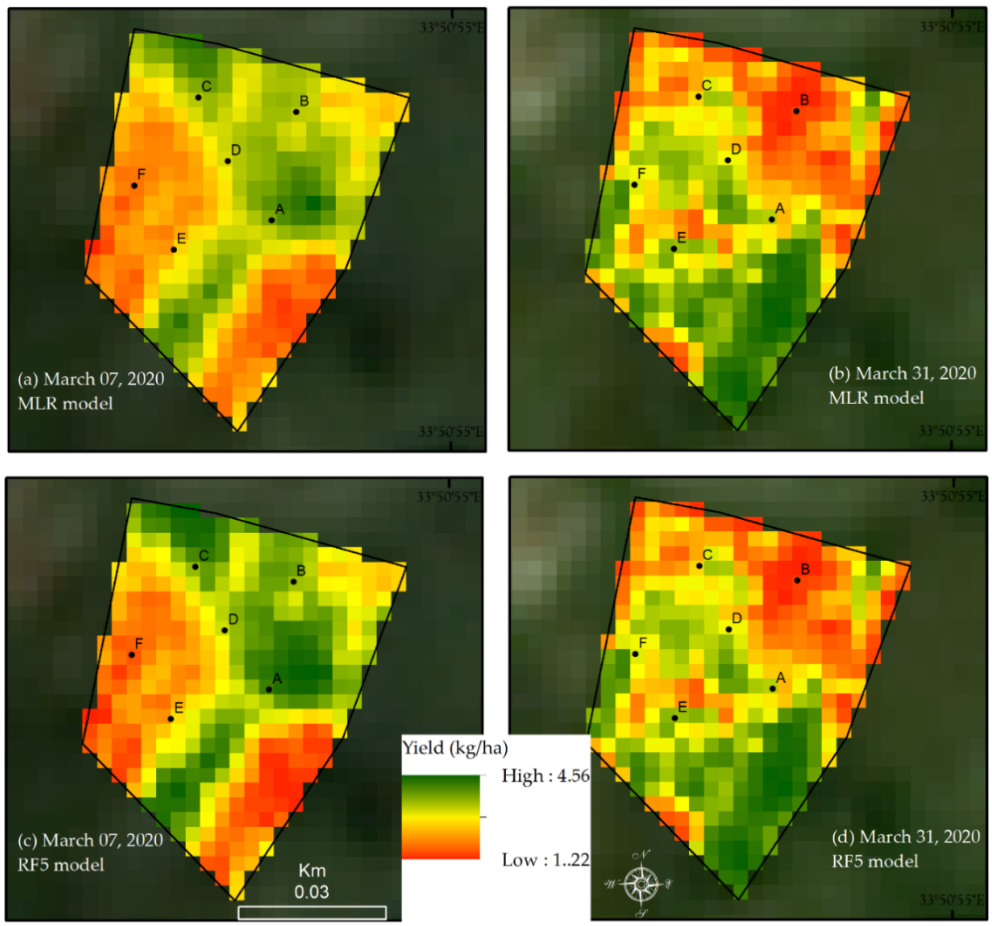

Figure 6. Yield prediction maps using the final MLR models during the (**a**) R3/beginning pod stage and the (**b**) R5/beginning seed stage (7 March 2020, image for Site 2), and the final RF5 models for the (**c**) R3/beginning pod stage and the (**d**) R5/beginning seed stage (31 March 2020, image for Site 2). The letters on the maps represent sampling points where the LAI and yield were measured.

Finally, Table 7 presents the results of the comparison between the estimated yield from the RF5 model using an image that was captured on 23 February 2020, and the total field reported yield from the four farms. The yield per pixel from the prediction map ranged from 1.22 kg/ha to 4.56 kg/ha. The results show varying degrees of accuracy between the estimates and the field reported yield. The model slightly overestimated field A by about 1.44%, B by 0.85%, C by 2.55%, and D by about 11.46%.

Table 7. Comparison of yield collected from the field and the predicted yield based on the RF5 model. The negatives represent the amount of overestimation by the model compared to the field reported yields from the four farms.

Farm	Reported Yield (kg/ha)	Predicted Yield (kg/ha)	Difference (kg/ha)	% Difference
Farm A	3180.72	3226.46	−45.74	−1.44
Farm B	1180.33	1190.31	−9.98	−0.85
Farm C	2400.56	2461.79	−61.23	−2.55
Farm D	140.75	156.88	−16.13	−11.46

6. Discussion

Crop yield prediction using satellite imagery has become a popular method for providing near-real-time prediction of crop yield for varying degrees of farm size and crop types, with our findings being consistent with the results of these findings [10,45,91,92]. Overall, the RF5 model, which was based on the five most important variables, outperformed the SLR and the MLR models in predicting the groundnut yield in smallholder farms in the sub-tropical climate zone. Other studies in different climatic zones have found similar results [10,35,44]. The RF and the MLR models produced accurate results during the modelling and validation stages, but the RF5 was more robust, with a lower RMSE and a higher R^2 (Tables 4 and 5). The random forest regressions generally more accurately predicted the crop yields, even with the small sample sizes, as has been found in several studies [93,94], which likely explains why the RF5 model produced superior results. The evidence suggests that there are trade-offs between the model accuracy and the time of growth when the yield prediction is made [16,95]. As such, we predicted the groundnut yields for two growth stages. The predicted yields in the two growth stages showed that the R5/beginning seed/development of seed growth stage emerged as the best time to predict the groundnut yield in the study location. This finding is consistent with the work of [96], who used OptRX and GreenSeeker sensors to establish that the best period to measure the groundnut yield variability in Brazil was between 45 days and 75 days after planting, which coincides with the R5 growth stage. This finding provides important information that can be used by the agricultural extension officers who deal with smallholder farmers to non-destructively predict the yield ahead of harvest and to help them to plan for household food needs and market dynamics. There was no relationship found between the yield and the variables on Site 1. This may be explained by the lower number of time-series PlanetScope images (three images) that were combined in order to compute the VIs for the analysis, compared to the six images on Site 2. There is evidence that combining time-series satellite data can produce more accurate analysis results [97,98].

The variable importance analysis has identified the TNDVI, the NDVI, the RVI, the MSAVI2, and the SAVI as the most important variables for predicting the groundnut yield. When all of the variables were added to the RF model, the R^2 was lower, and the model was not statistically significant at the 95% confidence level. Using the ranking function in RF reduces the redundancy in the number of variables, which likely explains the failure of the RF All model to produce statistically significant results. This observation demonstrates the utility of the RF model as a feature-selection method to identify the best variables for accurate prediction [99]. The difference in response of the VIs to the groundnut yields likely explains how different environmental factors, such as soil structure and soil slope, affect the performance of the spectral indices, which in turn may determine the relationship between

the VIs and the observed groundnut yield. For instance, [100] have shown that the RVI is the best index for measuring the vegetation biophysical parameters in low-density vegetation areas, such as with groundnut plants. The RF model has identified the important context-specific VIs that were relevant to this study and the general approach that is described in this manuscript can be applied to other locations in Malawi if a reasonable amount of yield data is available for building a reliable crop prediction model. This approach could be a handy decision support tool for agricultural decisions in smallholder contexts on how to manage and plan for risks in the farming system. It is also worth noting that the data for Site 1 produced a weak and statistically insignificant correlation between the yield and the VIs, therefore, they were not used for further analysis. The number of cloud-free images that were combined in order to generate the VI was lower (three images) compared to those that were combined for Site 2 (six images). Multitemporal images perform better in determining crop phenology [101,102], therefore, it is likely that the many images that were used for Site 2 more accurately captured the full season spectrum of the VI, which produced higher and more significant correlations with the yield data.

The results from applying the RF5 model onto an independent set of data generated varying degrees of accuracies for the four farms. One possible explanation of the discrepancy between the field reported yield data by the farmers and the RF5 predicted yield data could be that the model's errors resulted in the overestimation, as has been reported in prior studies [34,103]. It is also likely that the reported yield may have been underestimated due to post-harvest losses. Harvesting groundnuts in the study context was carried out by pulling the plants out or digging them up with a hoe. Either way, some pods are likely to have remained in the ground. Additional losses are also often generated by the hand plucking of the groundnuts by the farmers. Besides, harvesting crops in the study context is often carried out by employing communal or extended family labor. The laborers are often paid using part of the harvested groundnuts rather than with money. These losses could have affected the final collected yield and may partly explain the discrepancies between the model's estimated yields and the field reported yield. Studies by [104] in Malawi have revealed that farm-level post-harvest losses averaged between 5% and 12% of the farmer's total harvest, which is consistent with the findings of this study.

The study results demonstrate the potential for predicting groundnut yield in smallholder contexts using high resolution satellite-derived VIs. Groundnuts are highly nutritious and in a country such as Malawi where nutrition insecurity is high [11], understanding the interannual yield patterns is important for planning for household nutrition needs and farm-level management. Furthermore, being able to predict the crop yields can provide early warning information of either yield deficits or excesses, which supports advanced planning. Even though the models have been developed based on a single site in rural Malawi, the approach that has been described here can be applied to other smallholder farms in the sub-tropical region and other areas in SSA. The sample size for this study is relatively small compared to other studies using similar methods in other contexts. This could have led to false positives or false negatives, which may affect the generalizability of the results [104]. Given the small farm sizes and the spatial resolution of the satellite data, however, the sample that was used is representative of the crop conditions on the farm and the results are, therefore, applicable in such resource-poor contexts. Remote sensing analysis does not account for the post-harvest losses resulting from the mishandling of the harvested yield. Therefore, the predicted yield may slightly deviate from reality due to the losses that may occur; however, in the resource-poor smallholder contexts such results provide important information for planning purposes when trying to address food insecurity.

7. Conclusions

In this study, we used different methods in order to make a within-season groundnut yield prediction in smallholder contexts in SSA. The results suggest that groundnut yields can be better predicted at the R5 stage of growth using vegetation indices that are

computed from multitemporal, multispectral remote sensing data and a random forest regression modelling approach. These findings have practical implications for both public and private sector policies. First, The Ministry of Agriculture of the Republic of Malawi, in partnership with local organizations, such as the Soils, Food, and Healthy Communities (SFHC) organization, can build the capacity of extension officers to predict the yield to a reasonable degree of accuracy in advance of harvesting using publicly available satellite data. Second, our findings demonstrate the practicality of using high-resolution remote sensing data and machine learning methods in order to predict yield in smallholder agricultural systems, which can increase the use of precision agricultural practices among smallholder farmers in northern Malawi. Employing precision agricultural practices, such as crop yield prediction, would aid smallholder farmers and other agri-food system stakeholders to plan and mitigate low yields and crop failures. In addition, integrating this paper's results with climate and household health data could provide vital information on household nutrition and food needs and enable the policymakers to address the food insecurity in rural Malawi better.

Despite the usefulness of the findings, future studies could improve the generalizability of the current results by incorporating weather data and soil analysis in order to better understand the behavior of the predictors. Moreover, images with higher spectral resolution containing spectral bands, such as the red-edge and the shortwave infrared, could be used in order to explore a wider range of VIs that can better predict the yield of groundnuts and other crops. In addition, in the study context, farmlands tend to be small. As such, using UAVs for more detailed spatio-temporal analysis in order to compare them with the satellite data models would improve the generalizability of the models. Another limitation is the use of a relatively small sample (i.e., two smallholder farms) in this study, which may limit the generalizability of the yield estimation approaches. However, given that ecological and climatic factors in northern Malawi do not significantly vary from the sampled farms, the study findings are generalizable across northern Malawi.

Author Contributions: Conceptualization, D.K., I.L., J.W. and R.B.K.; Data curation, D.K. and K.M.; Formal analysis, D.K.; Funding acquisition, I.L., J.W. and R.B.K.; Investigation, D.K. and K.M.; Methodology, D.K., K.M. and J.W.; Project administration, E.L. and L.D.; Software, D.K., K.M. and J.W.; Supervision, I.L., J.W. and R.B.K.; Validation, K.M.; Visualization, K.M.; Writing—original draft, D.K., I.L. and J.W.; Writing—review and editing, K.M., E.L. and L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded as part of the FARMS4Biodiversity project through the 2017–2018 Belmont Forum and BiodivERsA joint call for research proposals, under the BiodivScen ERA-Net COFUND program, and with the funding organizations German Federal Ministry of Education and Research (BMBF Förderkennzeichen 01LC11804A), Research Council of Norway (no: 295442), National Science Foundation (NSF project 1852587) and the Natural Sciences and Engineering Research Council of Canada (NSERC Grant#: 523660-2018).

Data Availability Statement: The data used in not available.

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

References

1. USDA. *World Agricultural Production*; USDA: Washington, DC, USA, 2020.
2. Nautiyal, P.C. Groundnut: Post-Harvest Operations. *Res. Cent. Groundn. ICAR* **2002**, *23*, 2013.
3. Sako, D.; Traoré, M.; Doumbia, F.; Diallo, F.; Fané, M.; Kapran, I. Kolokani Groundnut Innovation Platform Activities and Achievements through TL III Project in Mali. In *Enhancing Smallholder Farmers' Access to Seed of Improved Legume Varieties through Multi-Stakeholder Platforms*; Springer: Singapore, 2021; pp. 51–64.
4. Aransiola, E.F.; Ehinmitola, E.O.; Adebimpe, A.I.; Shittu, T.D.; Solomon, B.O. Prospects of Biodiesel Feedstock as an Effective Ecofuel Source and Their Challenges. In *Advances in Eco-Fuels for a Sustainable Environment*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 53–87.
5. Corporate, C.A.I. *Malawi Groundnut Outlook*; Lilongwe, Malawi, 2016. Available online: https://mitc.mw/trade/index.php/groundnuts-export-product (accessed on 15 September 2022).

6. Abady, S.; Shimelis, H.; Janila, P.; Mashilo, J. Groundnut (*Arachis hypogaea* L.) Improvement in Sub-Saharan Africa: A Review. *Acta Agric. Scand. Sect. B-Soil Plant Sci.* **2019**, *69*, 528–545.
7. AICC. *Harmonized Groundnut Production Manual for Malawi*; Legumes Development Trust: Lilongwe, Malawi, 2014.
8. Holzman, M.E.; Rivas, R.; Piccolo, M.C. Estimating Soil Moisture and the Relationship with Crop Yield Using Surface Temperature and Vegetation Index. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *28*, 181–192. [CrossRef]
9. Schwalbert, R.A.; Amado, T.; Corassa, G.; Pott, L.P.; Prasad, P.V.V.; Ciampitti, I.A. Satellite-Based Soybean Yield Forecast: Integrating Machine Learning and Weather Data for Improving Crop Yield Prediction in Southern Brazil. *Agric. For. Meteorol.* **2020**, *284*, 107886. [CrossRef]
10. WFP. *Prevention of Undernutrition: 2019-Malawi Factsheets*; WFP: Rome, Italy, 2019.
11. Gao, F.; Anderson, M.; Daughtry, C.; Johnson, D. Assessing the Variability of Corn and Soybean Yields in Central Iowa Using High Spatiotemporal Resolution Multi-Satellite Imagery. *Remote Sens.* **2018**, *10*, 1489. [CrossRef]
12. Thenkabail, P.S.; Knox, J.W.; Ozdogan, M.; Gumma, M.K.; Congalton, R.G.; Wu, Z.; Milesi, C.; Finkral, A.; Marshall, M.; Mariotto, I. Assessing Future Risks to Agricultural Productivity, Water Resources and Food Security: How Can Remote Sensing Help? *Photogramm. Eng. Remote Sens.* **2012**, *78*, 773–782.
13. da Silva, E.E.; Baio, F.H.R.; Teodoro, L.P.R.; da Silva Junior, C.A.; Borges, R.S.; Teodoro, P. UAV-Multispectral and Vegetation Indices in Soybean Grain Yield Prediction Based on in Situ Observation. *Remote Sens. Appl. Soc. Environ.* **2020**, *18*, 100318. [CrossRef]
14. Vergara-Díaz, O.; Zaman-Allah, M.A.; Masuka, B.; Hornero, A.; Zarco-Tejada, P.; Prasanna, B.M.; Cairns, J.E.; Araus, J.L. A Novel Remote Sensing Approach for Prediction of Maize Yield under Different Conditions of Nitrogen Fertilization. *Front. Plant Sci.* **2016**, *7*, 666. [CrossRef] [PubMed]
15. Bolton, D.K.; Friedl, M.A. Forecasting Crop Yield Using Remotely Sensed Vegetation Indices and Crop Phenology Metrics. *Agric. For. Meteorol.* **2013**, *173*, 74–84. [CrossRef]
16. Swain, K.C.; Thomson, S.J.; Jayasuriya, H.P.W. Adoption of an Unmanned Helicopter for Low-Altitude Remote Sensing to Estimate Yield and Total Biomass of a Rice Crop. *Trans. ASABE* **2010**, *53*, 21–27. [CrossRef]
17. Wang, J.; Zhang, J.; Bai, Y.; Zhang, S.; Yang, S.; Yao, F. Integrating Remote Sensing-Based Process Model with Environmental Zonation Scheme to Estimate Rice Yield Gap in Northeast China. *Field Crops Res.* **2020**, *246*, 107682. [CrossRef]
18. Stepanov, A.; Dubrovin, K.; Sorokin, A.; Aseeva, T. Predicting Soybean Yield at the Regional Scale Using Remote Sensing and Climatic Data. *Remote Sens.* **2020**, *12*, 1936. [CrossRef]
19. Gómez, D.; Salvador, P.; Sanz, J.; Casanova, J.L. Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data. *Remote Sens.* **2019**, *11*, 1745. [CrossRef]
20. Hunt, E.R.; Horneck, D.A.; Spinelli, C.B.; Turner, R.W.; Bruce, A.E.; Gadler, D.J.; Brungardt, J.J.; Hamm, P.B. Monitoring Nitrogen Status of Potatoes Using Small Unmanned Aerial Vehicles. *Precis. Agric.* **2018**, *19*, 314–333. [CrossRef]
21. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* **2017**, *2017*, 1353691. [CrossRef]
22. Gitelson, A.A.; Peng, Y.; Arkebauer, T.J.; Suyker, A.E. Productivity, Absorbed Photosynthetically Active Radiation, and Light Use Efficiency in Crops: Implications for Remote Sensing of Crop Primary Production. *J. Plant Physiol.* **2015**, *177*, 100–109. [CrossRef]
23. 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]
24. Zhao, Y.; Potgieter, A.B.; Zhang, M.; Wu, B.; Hammer, G.L. Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. *Remote Sens.* **2020**, *12*, 1024. [CrossRef]
25. Baio, F.H.R.; Neves, D.C.; da Silva Campos, C.N.; Teodoro, P.E. Relationship between Cotton Productivity and Variability of NDVI Obtained by Landsat Images. *Biosci. J.* **2018**, *34*, 197–205. [CrossRef]
26. 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]
27. Chen, J.M.; Black, T.A. Defining Leaf Area Index for Non-flat Leaves. *Plant Cell Environ.* **1992**, *15*, 421–429. [CrossRef]
28. Buermann, W.; Wang, Y.; Dong, J.; Zhou, L.; Zeng, X.; Dickinson, R.E.; Potter, C.S.; Myneni, R.B. Analysis of a Multiyear Global Vegetation Leaf Area Index Data Set. *J. Geophys. Res. Atmos.* **2002**, *107*, ACL-14. [CrossRef]
29. Ke, L.; Zhou, Q.; Wu, W.; Tian, X.; Tang, H. Estimating the Crop Leaf Area Index Using Hyperspectral Remote Sensing. *J. Integr. Agric.* **2016**, *15*, 475–491.
30. Fu, Z.; Jiang, J.; Gao, Y.; Krienke, B.; Wang, M.; Zhong, K.; Cao, Q.; Tian, Y.; Zhu, Y.; Cao, W. Wheat Growth Monitoring and Yield Estimation Based on Multi-Rotor Unmanned Aerial Vehicle. *Remote Sens.* **2020**, *12*, 508. [CrossRef]
31. Hou, H.; Wei, M.; Noor, M.A.; Tang, L.; Li, C.; Ding, Z.; Ming, Z. Quantitative Design of Yield Components to Simulate Yield Formation for Maize in China. *J. Integr. Agric.* **2020**, *19*, 668–679. [CrossRef]
32. Paul, G.C.; Saha, S.; Hembram, T.K. Application of Phenology-Based Algorithm and Linear Regression Model for Estimating Rice Cultivated Areas and Yield Using Remote Sensing Data in Bansloi River Basin, Eastern India. *Remote Sens. Appl. Soc. Environ.* **2020**, *19*, 100367.
33. Khaki, S.; Wang, L.; Archontoulis, S. V A Cnn-Rnn Framework for Crop Yield Prediction. *Front. Plant Sci.* **2020**, *10*, 1750. [CrossRef]

34. 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]
35. Khan, M.S.; Semwal, M.; Sharma, A.; Verma, R.K. An Artificial Neural Network Model for Estimating Mentha Crop Biomass Yield Using Landsat 8 OLI. *Precis. Agric.* **2020**, *21*, 18–33. [CrossRef]
36. Prasad, N.R.; Patel, N.R.; Danodia, A. Crop Yield Prediction in Cotton for Regional Level Using Random Forest Approach. *Spat. Inf. Res.* **2020**, *29*, 195–206. [CrossRef]
37. Martín-Ortega, P.; García-Montero, L.G.; Sibelet, N. Temporal Patterns in Illumination Conditions and Its Effect on Vegetation Indices Using Landsat on Google Earth Engine. *Remote Sens.* **2020**, *12*, 211. [CrossRef]
38. Olmos-Trujillo, E.; González-Trinidad, J.; Júnez-Ferreira, H.; Pacheco-Guerrero, A.; Bautista-Capetillo, C.; Avila-Sandoval, C.; Galván-Tejada, E. Spatio-Temporal Response of Vegetation Indices to Rainfall and Temperature in A Semiarid Region. *Sustainability* **2020**, *12*, 1939. [CrossRef]
39. Semeraro, T.; Luvisi, A.; Lillo, A.O.; Aretano, R.; Buccolieri, R.; Marwan, N. Recurrence Analysis of Vegetation Indices for Highlighting the Ecosystem Response to Drought Events: An Application to the Amazon Forest. *Remote Sens.* **2020**, *12*, 907. [CrossRef]
40. Yang, Q.; Zheng, F.; Jia, X.; Liu, P.; Dong, S.; Zhang, J.; Zhao, B. The Combined Application of Organic and Inorganic Fertilizers Increases Soil Organic Matter and Improves Soil Microenvironment in Wheat-Maize Field. *J. Soils Sediments* **2020**, *20*, 2395–2404. [CrossRef]
41. de Sousa, M.A.; de Oliveira, M.M.; Damin, V.; de Brito Ferreira, E.P. Productivity and Economics of Inoculated Common Bean as Affected by Nitrogen Application at Different Phenological Phases. *J. Soil Sci. Plant Nutr.* **2020**, *20*, 1848–1858. [CrossRef]
42. Lowder, S.K.; Skoet, J.; Raney, T. The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Dev.* **2016**, *87*, 16–29. [CrossRef]
43. Gyamerah, S.A.; Ngare, P.; Ikpe, D. Probabilistic Forecasting of Crop Yields via Quantile Random Forest and Epanechnikov Kernel Function. *Agric. For. Meteorol.* **2020**, *280*, 107808. [CrossRef]
44. Karst, I.G.; Mank, I.; Traoré, I.; Sorgho, R.; Stückemann, K.-J.; Simboro, S.; Sié, A.; Franke, J.; Sauerborn, R. Estimating Yields of Household Fields in Rural Subsistence Farming Systems to Study Food Security in Burkina Faso. *Remote Sens.* **2020**, *12*, 1717. [CrossRef]
45. Gama, A.C.; Mapemba, L.D.; Masikat, P.; Tui, S.H.-K.; Crespo, O.; Bandason, E. *Modeling Potential Impacts of Future Climate Change in Mzimba District, Malawi, 2040–2070: An Integrated Biophysical and Economic Modeling Approach*; Intl Food Policy Res Inst: Washington, DC, USA, 2014; Volume 8.
46. Snapp, S.S. Soil Nutrient Status of Smallholder Farms in Malawi. *Commun. Soil Sci. Plant Anal.* **1998**, *29*, 2571–2588. [CrossRef]
47. Mzimba District Department of Planning. *Mzimba District Socioeconomic Profile*; Mzimba District Assembly: Mzimba, Malawi, 2008.
48. Lunduka, R.; Ricker-Gilbert, J.; Fisher, M. What Are the Farm-level Impacts of Malawi's Farm Input Subsidy Program? A Critical Review. *Agric. Econ.* **2013**, *44*, 563–579. [CrossRef]
49. Hall-Spencer, J.M. Agriculture Production as a Major Driver of the Earth System Exceeding Planetary Boundaries. *Ecol. Soc.* **2017**, *22*, 8.
50. Bi, L.; Yao, S.; Zhang, B. Impacts of Long-Term Chemical and Organic Fertilization on Soil Puddlability in Subtropical China. *Soil Tillage Res.* **2015**, *152*, 94–103. [CrossRef]
51. van Leeuwen, W.J.D.; Huete, A.R.; Laing, T.W. MODIS Vegetation Index Compositing Approach: A Prototype with AVHRR Data. *Remote Sens. Environ.* **1999**, *69*, 264–280. [CrossRef]
52. Fang, H.; Li, W.; Wei, S.; Jiang, C. Seasonal Variation of Leaf Area Index (LAI) over Paddy Rice Fields in NE China: Intercomparison of Destructive Sampling, LAI-2200, Digital Hemispherical Photography (DHP), and AccuPAR Methods. *Agric. For. Meteorol.* **2014**, *198*, 126–141. [CrossRef]
53. Mokhtari, A.; Noory, H.; Vazifedoust, M. Improving Crop Yield Estimation by Assimilating LAI and Inputting Satellite-Based Surface Incoming Solar Radiation into SWAP Model. *Agric. For. Meteorol.* **2018**, *250*, 159–170. [CrossRef]
54. Weiss, M.; Baret, F. *CAN_EYE V6.4.91 User Manual*; INRA Science & Impact: Avignon, France, 2017.
55. Mougin, E.; Demarez, V.; Diawara, M.; Hiernaux, P.; Soumaguel, N.; Berg, A. Estimation of LAI, FAPAR and FCover of Sahel Rangelands (Gourma, Mali). *Agric. For. Meteorol.* **2014**, *198*, 155–167. [CrossRef]
56. Jonckheere, I.; Nackaerts, K.; Muys, B.; Coppin, P. Assessment of Automatic Gap Fraction Estimation of Forests from Digital Hemispherical Photography. *Agric. For. Meteorol.* **2005**, *132*, 96–114. [CrossRef]
57. Weiss, M.; Baret, F.; Smith, G.J.; Jonckheere, I.; Coppin, P. Review of Methods for in Situ Leaf Area Index (LAI) Determination: Part II. Estimation of LAI, Errors and Sampling. *Agric. For. Meteorol.* **2004**, *121*, 37–53. [CrossRef]
58. Nilson, T. A Theoretical Analysis of the Frequency of Gaps in Plant Stands. *Agric. Meteorol.* **1971**, *8*, 25–38. [CrossRef]
59. Planet Labs Inc. Planet Imagery and Archive, 2020. Available online: https://www.planet.com/products/planet-imagery/ (accessed on 6 June 2021).
60. Kotchenova, S.Y.; Vermote, E.F. Validation of a Vector Version of the 6S Radiative Transfer Code for Atmospheric Correction of Satellite Data. Part II. Homogeneous Lambertian and Anisotropic Surfaces. *Appl. Opt.* **2007**, *46*, 4455–4464. [CrossRef]
61. Guerini Filho, M.; Kuplich, T.M.; Quadros, F.L.F. De Estimating Natural Grassland Biomass by Vegetation Indices Using Sentinel 2 Remote Sensing Data. *Int. J. Remote Sens.* **2020**, *41*, 2861–2876. [CrossRef]

62. Kpienbaareh, D.; Kansanga, M.; Luginaah, I. Examining the Potential of Open Source Remote Sensing for Building Effective Decision Support Systems for Precision Agriculture in Resource-Poor Settings. *GeoJournal* **2018**, *84*, 1481–1497. [CrossRef]
63. Kpienbaareh, D.; Sun, X.; Wang, J.; Luginaah, I.; Kerr, R.B.; Lupafya, E.; Dakishoni, L. Crop Type and Land Cover Mapping in Northern Malawi Using the Integration of Sentinel-1, Sentinel-2, and PlanetScope Satellite Data. *Remote Sens.* **2021**, *13*, 700. [CrossRef]
64. Todd, S.W.; Hoffer, R.M. Responses of Spectral Indices to Variations in Vegetation Cover and Soil Background. *Photogramm. Eng. Remote Sens.* **1998**, *64*, 915–922.
65. Hashimoto, A.; Segah, H.; Yulianti, N.; Naruse, N.; Takahashi, Y. A New Indicator of Forest Fire Risk for Indonesia Based on Peat Soil Reflectance Spectra Measurements. *Int. J. Remote Sens.* **2021**, *42*, 1917–1927. [CrossRef]
66. Lucas, B.D.; Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Vancouver, BC, Canada, 24–28 August 1981.
67. 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]
68. Tucker, C.J. Red and Photographic Infrared Linear Combinations Monitoring Vegetation. *J. Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
69. Huete, A. Huete, AR A Soil-Adjusted Vegetation Index (SAVI). Remote Sensing of Environment. *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]
70. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]
71. Pearson, R.L.; Miller, L.D. Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Shortgrass Prairie. *Remote Sens. Environ.* **1972**, *1355*, 37–50.
72. Senseman, G.M.; Bagley, C.F.; Tweddale, S.A. Correlation of Rangeland Cover Measures to Satellite-imagery-derived Vegetation Indices. *Geocarto Int.* **1996**, *11*, 29–38. [CrossRef]
73. Clevers, J. The Derivation of a Simplified Reflectance Model for the Estimation of Leaf Area Index. *Remote Sens. Environ.* **1988**, *25*, 53–69. [CrossRef]
74. Panda, S.S.; Ames, D.P.; Panigrahi, S. Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. *Remote Sens.* **2010**, *2*, 673–696. [CrossRef]
75. Satir, O.; Berberoglu, S. Crop Yield Prediction under Soil Salinity Using Satellite Derived Vegetation Indices. *Field Crops Res.* **2016**, *192*, 134–143. [CrossRef]
76. Malone, S.; Ames Herbert, D., Jr.; Holshouser, D.L. Relationship between Leaf Area Index and Yield in Double-Crop and Full-Season Soybean Systems. *J. Econ. Entomol.* **2002**, *95*, 945–951. [CrossRef]
77. Mendoza-Pérez, C.; Ramírez-Ayala, C.; Ojeda-Bustamante, W.; Flores-Magdaleno, H. Estimation of Leaf Area Index and Yield of Greenhouse-Grown Poblano Pepper; 2017. Available online: https://www.researchgate.net/publication/318018386_Estimation_of_leaf_area_index_and_yield_of_greenhouse-grown_poblano_pepper (accessed on 15 September 2022).
78. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* **2006**, *9*, 181–199. [CrossRef]
79. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
80. Verikas, A.; Gelzinis, A.; Bacauskiene, M. Mining Data with Random Forests: A Survey and Results of New Tests. *Pattern Recognit.* **2011**, *44*, 330–349. [CrossRef]
81. Belgiu, M.; Drăguţ, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
82. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitao, P.J. Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Ecography* **2013**, *36*, 27–46. [CrossRef]
83. Team, R.C. R: A Language and Environment for Statistical Computing; MSOR Connections; 2013. Available online: https://www.semanticscholar.org/paper/R%3A-A-language-and-environment-for-statistical-Team/659408b243cec55de8d0a3bc51b81173007aa89b (accessed on 15 September 2022).
84. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In Proceedings of the International Joint Conference on Articial Intelligence (IJCAI), Montreal, QC, Canada, 20 August 1995; Volume 14, pp. 1137–1145.
85. Peduzzi, A.; Wynne, R.H.; Fox, T.R.; Nelson, R.F.; Thomas, V.A. Estimating Leaf Area Index in Intensively Managed Pine Plantations Using Airborne Laser Scanner Data. *For. Ecol. Manag.* **2012**, *270*, 54–65. [CrossRef]
86. Brovelli, M.A.; Crespi, M.; Fratarcangeli, F.; Giannone, F.; Realini, E. Accuracy Assessment of High Resolution Satellite Imagery Orientation by Leave-One-out Method. *ISPRS J. Photogramm. Remote Sens.* **2008**, *63*, 427–440. [CrossRef]
87. Yue, J.; Feng, H.; Jin, X.; Yuan, H.; Li, Z.; Zhou, C.; Yang, G.; Tian, Q. A Comparison of Crop Parameters Estimation Using Images from UAV-Mounted Snapshot Hyperspectral Sensor and High-Definition Digital Camera. *Remote Sens.* **2018**, *10*, 1138. [CrossRef]
88. Boote, K.J. Growth Stages of Peanut (*Arachis hypogaea* L.). *Peanut Sci.* **1982**, *9*, 35–40. [CrossRef]
89. Munger, P.; Bleiholder, H.; Hack, H.; Hess, M.; Stauss, R.; van Den Boom, T.; Weber, E. Phenological Growth Stages of the Peanut Plant (*Arachis Hypogaea* L.): Codification and Description According to the BBCH Scale 1. *J. Agron. Crop Sci.* **1998**, *180*, 101–107. [CrossRef]

90. Azzari, G.; Jain, M.; Lobell, D.B. Towards Fine Resolution Global Maps of Crop Yields: Testing Multiple Methods and Satellites in Three Countries. *Remote Sens. Environ.* **2017**, *202*, 129–141. [CrossRef]
91. Lobell, D.B.; Thau, D.; Seifert, C.; Engle, E.; Little, B. A Scalable Satellite-Based Crop Yield Mapper. *Remote Sens. Environ.* **2015**, *164*, 324–333. [CrossRef]
92. Liang, L.; Qin, Z.; Zhao, S.; Di, L.; Zhang, C.; Deng, M.; Lin, H.; Zhang, L.; Wang, L.; Liu, Z. Estimating Crop Chlorophyll Content with Hyperspectral Vegetation Indices and the Hybrid Inversion Method. *Int. J. Remote Sens.* **2016**, *37*, 2923–2949. [CrossRef]
93. Tillack, A.; Clasen, A.; Kleinschmit, B.; Förster, M. Estimation of the Seasonal Leaf Area Index in an Alluvial Forest Using High-Resolution Satellite-Based Vegetation Indices. *Remote Sens. Environ.* **2014**, *141*, 52–63. [CrossRef]
94. You, J.; Li, X.; Low, M.; Lobell, D.; Ermon, S. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
95. Carneiro, F.M.; Furlani, C.E.A.; Zerbato, C.; de Menezes, P.C.; da, S. Gírio, L.A. Correlations among Vegetation Indices and Peanut Traits during Different Crop Development Stages. *Eng. Agríc.* **2019**, *39*, 33–40. [CrossRef]
96. Waldner, F.; Lambert, M.-J.; Li, W.; Weiss, M.; Demarez, V.; Morin, D.; Marais-Sicre, C.; Hagolle, O.; Baret, F.; Defourny, P. Land Cover and Crop Type Classification along the Season Based on Biophysical Variables Retrieved from Multi-Sensor High-Resolution Time Series. *Remote Sens.* **2015**, *7*, 10400–10424. [CrossRef]
97. Djamai, N.; Zhong, D.; Fernandes, R.; Zhou, F. Evaluation of Vegetation Biophysical Variables Time Series Derived from Synthetic Sentinel-2 Images. *Remote Sens.* **2019**, *11*, 1547. [CrossRef]
98. Lawrence, R.L.; Wood, S.D.; Sheley, R.L. Mapping Invasive Plants Using Hyperspectral Imagery and Breiman Cutler Classifications (RandomForest). *Remote Sens. Environ.* **2006**, *100*, 356–362. [CrossRef]
99. Broge, N.H.; Leblanc, E. Comparing Prediction Power and Stability of Broadband and Hyperspectral Vegetation Indices for Estimation of Green Leaf Area Index and Canopy Chlorophyll Density. *Remote Sens. Environ.* **2001**, *76*, 156–172. [CrossRef]
100. Tian, H.; Huang, N.; Niu, Z.; Qin, Y.; Pei, J.; Wang, J. Mapping Winter Crops in China with Multi-Source Satellite Imagery and Phenology-Based Algorithm. *Remote Sens.* **2019**, *11*, 820. [CrossRef]
101. Pan, Z.; Huang, J.; Zhou, Q.; Wang, L.; Cheng, Y.; Zhang, H.; Blackburn, G.A.; Yan, J.; Liu, J. Mapping Crop Phenology Using NDVI Time-Series Derived from HJ-1 A/B Data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 188–197. [CrossRef]
102. Wei, M.C.F.; Maldaner, L.F.; Ottoni, P.M.N.; Molin, J.P. Carrot Yield Mapping: A Precision Agriculture Approach Based on Machine Learning. *AI* **2020**, *1*, 15. [CrossRef]
103. Ambler, K.; de Brauw, A.; Godlonton, S. Measuring Postharvest Losses at the Farm Level in Malawi. *Aust. J. Agric. Resour. Econ.* **2018**, *62*, 139–160. [CrossRef]
104. Jenkins, D.G.; Quintana-Ascencio, P.F. A Solution to Minimum Sample Size for Regressions. *PLoS ONE* **2020**, *15*, e0229345. [CrossRef] [PubMed]