

Article

Use of UAS Multispectral Imagery at Different Physiological Stages for Yield Prediction and Input Resource Optimization in Corn

Razieh Barzin ¹, Rohit Pathak ¹ , Hossein Lotfi ², Jac Varco ³  and Ganesh C. Bora ^{1,*}

¹ Department of Agricultural and Biological Engineering, Mississippi State University, Starkville, MS 39762, USA; rb2146@msstate.edu (R.B.); rp1265@msstate.edu (R.P.)

² Department of Geosciences, Mississippi State University, Starkville, MS 39762, USA; hl847@msstate.edu

³ Department of Plant and Soil Sciences, Mississippi State University, Starkville, MS 39762, USA; jjv3@msstate.edu

* Correspondence: gcbora@abe.msstate.edu; Tel.: +1-662-325-3155

Received: 12 June 2020; Accepted: 16 July 2020; Published: 26 July 2020



Abstract: Changes in spatial and temporal variability in yield estimation are detectable through plant biophysical characteristics observed at different phenological development stages of corn. A multispectral red-edge sensor mounted on an Unmanned Aerial Systems (UAS) can provide spatial and temporal information with high resolution. Spectral analysis of UAS acquired spatiotemporal images can be used to develop a statistical model to predict yield based on different phenological stages. Identifying critical vegetation indices (VIs) and significant spectral information could lead to increased yield prediction accuracy. The objective of this study was to develop a yield prediction model at specific phenological stages using spectral data obtained from a corn field. The available spectral bands (red, blue, green, near infrared (NIR), and red-edge) were used to analyze 26 different VIs. The spectral information was collected from a cornfield at Mississippi State University using a MicaSense multispectral red-edge sensor, mounted on a UAS. In this research, a new empirical method used to reduce the effects of bare soil pixels in acquired images was introduced. The experimental design was a randomized complete block that consisted of 16 blocks with 12 rows of corn planted in each block. Four treatments of nitrogen (N) including 0, 90, 180, and 270 kg/ha were applied randomly. Random forest was utilized as a feature selection method to choose the best combination of variables for different stages. Multiple linear regression and gradient boosting decision trees were used to develop yield prediction models for each specific phenological stage by utilizing the most effective variables at each stage. At the V3 (3 leaves with visible leaf collar) and V4-5 (4-5 leaves with visible leaf collar) stages, the Optimized Soil Adjusted Vegetation Index (OSAVI) and Simplified Canopy Chlorophyll Content Index (SCCCI) were the single dominant variables in the yield predicting models, respectively. A combination of the Green Atmospherically Resistant Index (GARI), Normalized Difference Red-Edge (NDRE), and green Normalized Difference Vegetation Index (GNDVI) at V6-7, SCCCI, and Soil-Adjusted Vegetation Index (SAVI) at V10,11, and SCCCI, Green Leaf Index (GLI), and Visible Atmospherically Resistant Index (VARI_{green}) at tasseling stage (VT) were the best indices for predicting grain yield of corn. The prediction models at V10 and VT had the greatest accuracy with a coefficient of determination of 0.90 and 0.93, respectively. Moreover, the SCCCI as a combined index seemed to be the most proper index for predicting yield at most of the phenological stages. As corn development progressed, the models predicted final grain yield more accurately.

Keywords: Gradient Boosting; Multispectral red-edge sensor; Random forest; UAS imagery; Vegetation index; Yield prediction.

1. Introduction

Estimation of corn yield during the crop-growing season is essential for efficient management of corn at strategic phenological stages. Agricultural surveys and field sampling of standing crops are supposed to be reliable approaches to estimate corn production. However, the spatiotemporal variability of biophysical characteristics of the crops due to inconsistency in soil nutrients and water availability, as well as other environmental parameters affecting plant growth presenting challenges in estimating yield on a large spatial scale accurately. The Normalized Difference Vegetation Index (NDVI) [1] can be used to quantify biomass production [2] by measuring the difference between near-infrared (NIR) and red wavelengths and is widely used in agricultural crop studies [3–5]. In this study, several vegetation indices (VIs) such as Normalized Difference Red-Edge (NDRE), Optimized Soil Adjusted Vegetation Index (OSAVI), Simplified Canopy Chlorophyll Content Index (SCCCI), and Visible Atmospherically Resistant Index (VARI_{green}), were used to determine their importance in predicting corn grain yield.

Remote-sensing technologies have been used across a wide range of applications in agriculture to detect and monitor the biophysical characteristics of plants. The spectral information collected by pixels is used to compute VIs, which are algorithms derived from the spectral transformation of reflectance at two or more specified wavelengths and are used to evaluate vegetative cover or biomass and plant growth or health status. Differencing, rationing, rationing sums and differences, and linear combinations of different spectral wavelengths are standard methods used to calculate different VIs. One of the advantages of using a remotely sensed VI products is that they are computed in a uniform manner and comparable during time and location [6].

Unmanned aerial systems (UAS) with dedicated sensors and communication packages are gaining popularity across a range of disciplines. As per requirements, the system can be designed for specific missions or applications. The UAS equipped with the RedEdge™ multispectral camera can be used to detect spatial and temporal variability in biophysical characteristics of corn, such as spectral reflectance for the specified wavelengths, which can be used to compute multiple VIs. Satellite imagery is routinely used to estimate yield of different crops [7–9]. Ongoing past examinations show that the red-edge waveband is useful for estimating the chlorophyll content and N status of plants. NDVI-Red-edge is increasingly profitable and helpful for later stages when contrasted with the early V6 stage for in-season N application [10]. Despite several applications of satellite imagery in agricultural studies, a lack of consistency and reliability of satellite products due to noise or errors and atmospheric and cloud interferences are unavoidable issues. Clouds, shadows, and atmospheric scattering of light can be a significant obstacle to obtaining high-quality data which can make it difficult to detect and analyze land-surface features due to interference [11]. A UAS can acquire data from low altitude; where, interference by clouds is not an obstacle between the sensor and land surface [12,13], but shadows created by them can still be an issue. Another advantage of a UAS is they can provide greater spatial resolution, and flights can be scheduled for key periods of designated phenological stages considering the weather conditions.

In comparison to satellite images, which are blurred and in which objects are difficult to distinguish, the data collected by the UAS has a high spatial resolution that provides a reasonable basis to identify and analyze crop development. Satellite images are affected by the scattering and absorbing effects of atmospheric gases and aerosols. From an altitude of 60 m, a camera mounted on a UAS can collect more detailed and important local landscape information at around 4 cm spatial resolution. The spatial resolution can be improved by using a more accurate camera sensor or decreasing the UAS flight altitude [14]. Therefore, a UAS can provide images with smaller pixel sizes, and it is possible to acquire spectral images as far as required for research objectives more frequently. The revisit time is more crucial for the plant growth assessment. Since the plant responses and requirements are different at their different phenological growth stages. The UAS image acquisition process, as well as spatial resolution, are flexible and more economical than point measurements [15].

Crop yield prediction can provide information for improving crop management, food production monitoring, economic trading, and global food security. The emergence of new statistical learning models such as ensemble methods based on decision trees can estimate yield before harvesting. The decision tree approach is increasingly being used for different purposes such as corn optimal fertilizer estimation [16] and corn yield estimation [17]. Gradient boosting machines (GBMs) are ensemble learning models to empower the weaker models such as decision trees by combining the results from them. GBMs are widely used in a broad range of practical applications [18] and have demonstrated remarkable success for regression and classification applications.

The objectives of this study were to track five different spectral bands obtained through sensors mounted on the UAS at five different phenological stages of corn and use 26 calculated VIs at each specific stage of growth. Feature selection approaches were applied to reduce the number of predictors. Consequently, relationships between the spectral bands and 26 VIs (as predictors), and corn yield (as a response variable) were investigated to determine more correlated covariates with the response variables. Finally, machine-learning techniques were hired to developed models for corn yield prediction at each phenological stage.

2. Materials and Methods

The study was undertaken on an experimental plot at Mississippi State University.

2.1. Study Area

The study area was located at the W.B. Andrews Agriculture Systems Research Farm at Mississippi State, MS, USA (33°28′13.5″N, 88°45′48.0″W) (Figure 1). The total area of the field was 0.8 ha mapped as a Marietta fine sandy loam (fine-loamy, mixed, siliceous, thermic, Aquic Fluventic Eutrochept). The imagery data were collected during the corn growing season for the years 2017, 2018, and 2019.

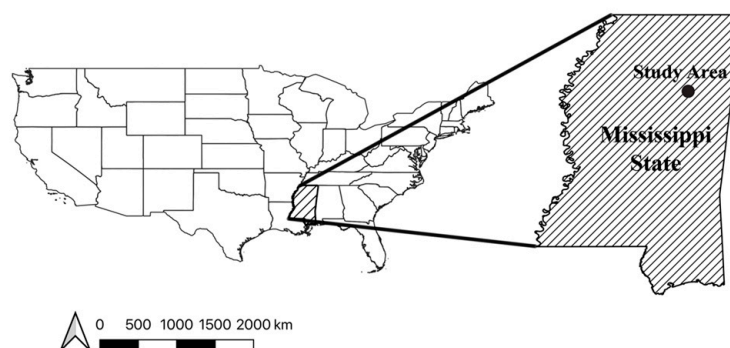


Figure 1. Geolocation of the study area monitored during the growing season.

Growing season precipitation totals measured at the experimental field varying between 58 cm in 2017, 42 cm in 2018, and 76 cm in 2019, as shown in Figure 2. The data were retrieved from the Mississippi Delta weather information monitored by the Delta Agricultural Weather Center at the Delta Research and Extension Center, which is located at a distance of 1 km from the research field. The precipitation in 2019 was the greatest of the three years, whereas 2018 was the lowest year. Because of the low precipitation, signs of water stress were observed in the plants. The water deficiency issue was addressed through furrow-irrigation in early June 2018. The mean temperature was almost similar throughout all three growing seasons, which was 23 degrees Celsius.

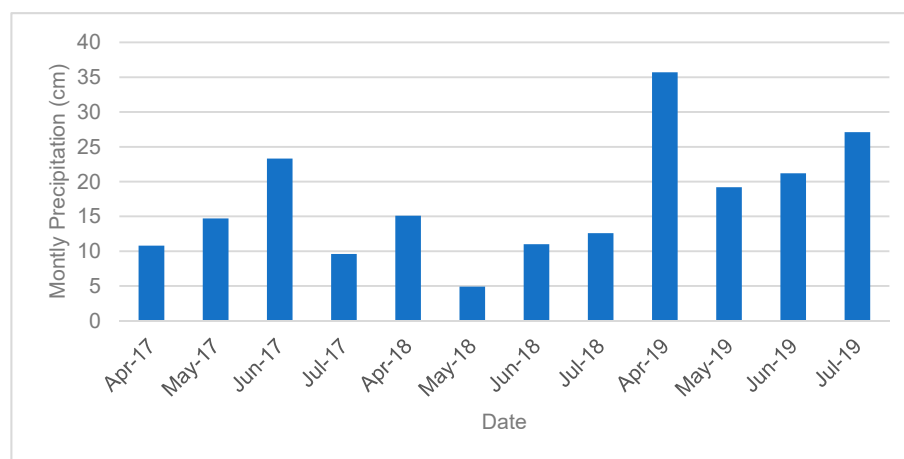


Figure 2. Monthly precipitation chart for the growing seasons: 2017, 2018, and 2019.

2.2. Experimental Design

The experimental field was divided into 16 plots, including 12 rows of corn, which were planted at a row spacing of 97 cm, and plot length was 38 m, and there was a 3 m alleyway in between each plot. The experimental design was a randomized complete block. Corn (DeKalb Brand-DKC67-72 variety) was planted on 13 April 2017, 19 April 2018, and 23 April 2019. There were four treatments of N, including 0, 90, 180, and 270 kg/ha, applied randomly with four replicates. Figure 3, illustrates the spatial distribution of each treatment and associated replication in the study field. The goal of the spatially varied N application was to identify the optimal N requirements for the corn crop and to address the spatial variability of the soil. Treatments were randomly assigned to the plots and have been repeated each year on the same experimental units.

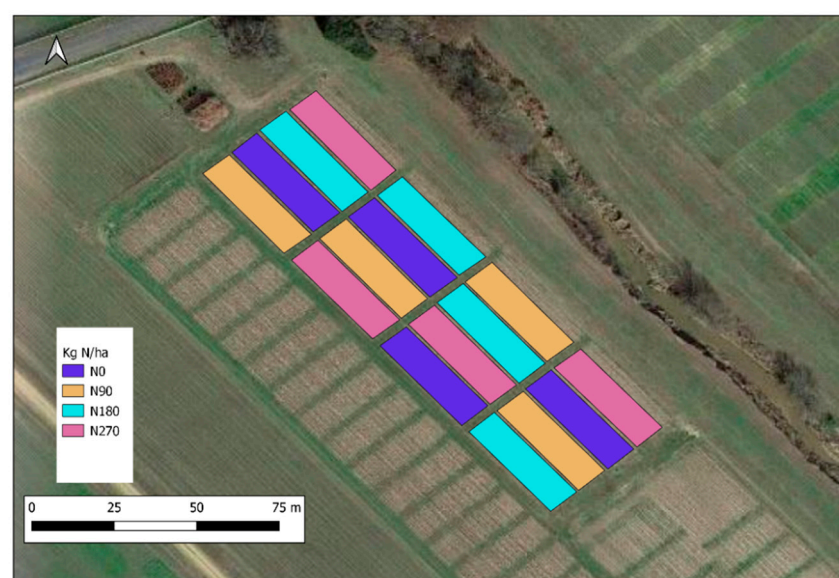


Figure 3. Nitrogen treatments and four replicates in 2017-2019 at the Research Farm, Mississippi State, US.

The first N application was made just after emergence in each year at V2-3 (2–3 leaves with visible leaf collar), followed by a second application at V6-7 (6–7 leaves with visible leaf collar). Figure 4 illustrates N applications, planting/harvesting, and flight dates during the different phenological stages of corn for 2017, 2018, and 2019. Fertilizer N was side dressed as liquid urea ammonium nitrate (UAN) (32-0-0) with an applicator equipped with colters, and liquid knives spaced 23 cm from one side of each corn row and 7.62 cm deep. Limited irrigation was supplied through furrow

irrigation to prevent drought stress. Strip tillage was utilized for these years, although plots were disked, and beds were formed following the 2017 growing season. Following the 2017 corn harvest, a Persian clover (*Trifolium resupinatum* L.) cover crop was planted at a percent live seeding rate of 6.74 kg seed ha⁻¹ across the whole experimental area using a no-till grain drill. Plots were fertilized based on soil test results and received uniform applications of P-K-Mg-S before planting for all parcels. The fertilizer blend consisted of two parts muriate of potash (0-0-60), one part concentrated super phosphate (0-46-0), and one part sulfate of potash-magnesia (0-0-22-11Mg-22S) and was applied at a material rate of 224 kg ha⁻¹. Weeds and pests were controlled based on Mississippi State University Extension recommendations. The field under study has been used for corn cultivation since 2012 with the same fertilize N rates applied each year. Corn grain was harvested using a two-row plot combine, which collected the yield from rows 2 and 3 and rows 10 and 11 (Figure 5).

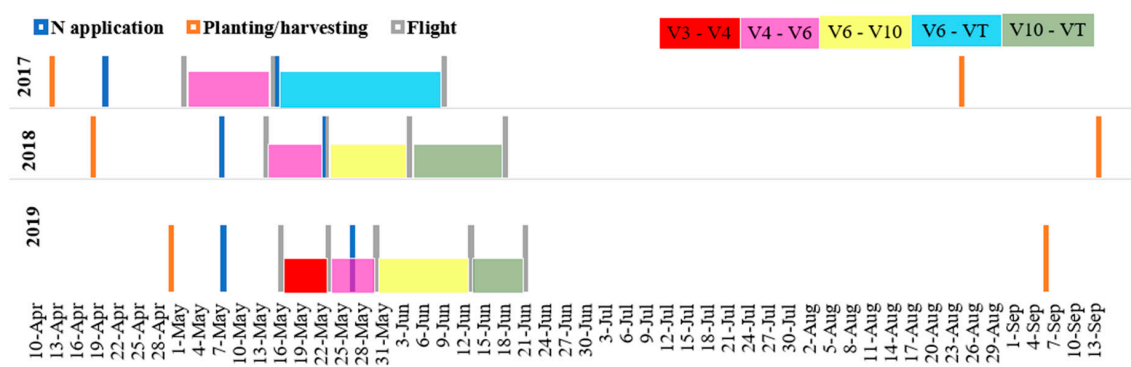


Figure 4. N applications, planting/harvesting, and flight dates during the different phenological stages of corn for 2017, 2018, and 2019.

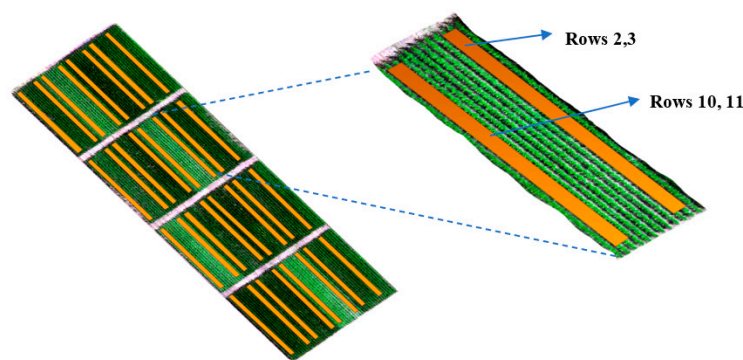


Figure 5. The spatial location of rows 2, 3, 10, and 11.

2.3. Data Collection

The MicaSense RedEdge™ multispectral band sensor mounted on a UAS was used to capture images in five different spectral bands simultaneously. The unit weight was 150 grams with a dimension of 12.1 cm × 6.6 cm × 4.6 cm. The UAS was flown at an altitude of 60 m in 2017 and 2018, whereas in 2019 it was flown at an altitude of 30 m. Decreasing the altitude from 60 m to 30 m provided better images with approximately four times greater resolution. The enhanced resolution was beneficial in separation of soil and vegetation. The sensor was mounted on the bottom of the UAS with a viewing angle not exceeding 10 degrees from nadir. The multispectral sensor measured the wavelength at five different spectral bands, including blue (475 nm center, 32 nm bandwidth), green (560 nm center, 27 nm bandwidth), red (668 nm center, 16 nm bandwidth), red-edge (717 nm center, 12 nm bandwidth), and near-infrared (842 nm center, 57 nm bandwidth). All five bands were collected simultaneously at a speed of one capture per second. Optimal image acquisition time is within plus or minus two and a half hours of local solar noon [4,19–22], therefore, all the flights were performed around 10:30 am under

cloud-free conditions. The length of the flight was around 10 minutes for the 0.8 ha field area; therefore, environmental conditions such as solar radiation, temperature, and humidity were nearly constant during the data acquisition process. The UAS images were acquired with a horizontal overlap of at least 75%. Images were stitched and mosaicked with the Pix4D mapper software (Pix4D SA, Lausanne, Switzerland) to obtain unique and compiled images for the study area. Superimposed images obtained through the stacking of images were disoriented during the first flight. This may be due to the errant movement of the camera. To address this issue a co-registration process was adopted. A calibrated reflectance panel (CRP) was used for the radiometric calibration of the acquired images. The CRP offers calibration information associated with the acquired images across the visible and near-infrared images. Images of the CRP that had been taken before and after the flight were used to convert raw pixel values into reflectance. The initial processing of the raw images was done at the Geosystems Research Institute (GRI) at Mississippi State University.

2.3.1. Vegetation Indices

Vegetation indices are mathematical combinations of wavelength-specific spectral reflectance developed to detect and monitor vegetation's phenological conditions remotely. For vegetation, reflectance by itself is low in both the blue and red bands of the spectrum due to maximum chlorophyll absorption in those bands while reflectance has a peak in the green band. Because of the cellular structures of leaves, the reflectance is much more significant in the NIR bands compared to visible bands. In this study, several VIs were derived from a 5-band multispectral sensor. Multispectral bands are visually and numerically similar; on the other hand, they are often highly correlated [23,24]. To avoid the issue associated with VI calculation, raw data was converted to percentage reflectance to signify the quantitative data. The name of the indices and associated spectral bands are listed in Table 1. The choices of indices by the researchers may vary according to their need but for biomass content, most indices involving red, infrared, and red-edge bands were preferred. These bands are supposed to explain even the subtle changes in biomass content.

Table 1. Mathematical representation of vegetation indices and ratios calculated from spectral reflectance.

Vegetation Indices (VI)	Name	Formula	Study Groups (Reference)
1	DVI	Difference Vegetation Index	$\text{NIR} - \text{Reds}$ [25]
2	GDVI	Green Difference Vegetation Index	$\text{NIR} - \text{Green}$ [26]
3	RDVI	Renormalized Difference Vegetation Index	$(\text{NIR} - \text{Red}) / \sqrt{\text{NIR} + \text{Red}}$ [27]
4	TDVI	Transformed Difference Vegetation Index	$1.5 (\text{NIR} - \text{Red}) / \sqrt{\text{NIR}^2 + \text{Red} + 0.5}$ [28]
5	NDVI	Normalized Difference Vegetation Index	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$ [1,29]
6	GNDVI	Green Normalized Difference Vegetation Index	$(\text{NIR} - \text{Green}) / (\text{NIR} + \text{Green})$ [30]
7	NDRE	Normalized Difference Red-edge	$(\text{NIR} - \text{Red-edge}) / (\text{NIR} + \text{Red-edge})$ [31,32]
8	SCCCI	Simplified Canopy Chlorophyll Content Index	$\text{NDRE} / \text{NDVI}$ [32]
9	EVI	Enhanced Vegetation Index	$2.5 * (\text{NIR} - \text{Red}) / (\text{NIR} + 6\text{Red} - 7.5\text{Blue} + 1)$ [33]
10	TVI	Triangular Vegetation Index	$0.5 [120 (\text{NIR} - \text{Green})] - 200 (\text{Red} - \text{Green})$ [34]
11	VARI _{green}	Visible Atmospherically Resistant Index	$(\text{Green} - \text{Red}) / (\text{Green} + \text{Red} - \text{Blue})$ [20]
12	GARI	Green Atmospherically Resistant Index	$\text{NIR} - \text{Green} - (1.7 (\text{Blue} - \text{Red})) / (\text{NIR} + \text{Green} - (1.7 (\text{Blue} - \text{Red})))$ [35]
13	GCI	Green Chlorophyll Index	$(\text{NIR} / \text{Green}) - 1$ [36]
14	GLI	Green Leaf Index	$(\text{Green} - \text{Red} - \text{Blue}) / (2\text{Green} + \text{Red} + \text{Blue})$ [37]
15	TGI	Triangular Greenness Index	$(\text{Red} - \text{Blue}) (\text{Red} - \text{Green}) - (\text{Red} - \text{Green}) (\text{Red} - \text{Blue}) / 2$ [38]
16	NLI	Non-Linear Index	$(\text{NIR}^2 - \text{Red}) / (\text{NIR}^2 + \text{Red})$ [39]
17	MNLI	Modified Non-Linear Index	$(\text{NIR}^2 - \text{Red}) * (1 + 0.5) / (\text{NIR}^2 + \text{Red} + 0.5)$ [40,41]
18	SAVI	Soil-Adjusted Vegetation Index	$1.5 * (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + 0.5)$ [42]
19	GSAVI	Green Soil-Adjusted Vegetation Index	$1.5 * (\text{NIR} - \text{Green}) / (\text{NIR} + \text{Green} + 0.5)$ [43]
20	OSAVI	Optimized Soil-Adjusted Vegetation Index	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + 0.16)$ [42]

Table 1. Cont.

	Vegetation Indices (VI)	Name	Formula	Study Groups (Reference)
21	GOSAVI	Green Optimized Soil-Adjusted Vegetation Index	$(\text{NIR} - \text{Green})/(\text{NIR} + \text{Green} + 0.16)$	[43]
22	MSAVI2	Modified Soil-Adjusted Vegetation Index 2	$(2\text{NIR} + 1 - \sqrt{(2\text{NIR} + 1)^2 - 8(\text{NIR} - \text{Red})})/2$	[44]
23	MSR	Modified Simple Ratio	$(\text{NIR}/\text{Red}) - 1/\sqrt{\text{NIR}/\text{Red} + 1}$	[5]
24	GRVI	Green Ratio Vegetation Index	NIR/Green	[25]
25	WDRVI	Wide Dynamic Range Vegetation Index	$(0.1 \text{ NIR} - \text{Red})/(0.1 \text{ NIR} + \text{red})$	[19]
26	SR	Simple Ratio	NIR/Red	[45]

Where Red, Green, Blue, Red-edge, and NIR are the reflectance values at 0.67 μm , 0.56 μm , 0.45, 0.72 μm , and 0.84 μm respectively.

2.3.2. Masking Soil Pixels

To estimate the spatial average of VIs for each corn row, it was essential to mask bare soil pixels located between corn rows. After the VIs calculation, the bare soil pixels were removed since these pixels do not provide further information in the yield estimation modeling. Eliminating these pixels reduced the image processing time and attributed to better estimate the spatial average of VIs for each row. Moreover, reflectance data from the corn rows contain information associated with the corn leaves and the scattered wavelength from the background soil within the leaves. The background soil reflectance potentially decreases the effectiveness of the leaves in VI values [46]. The occurrence of such a phenomenon is explicitly noticed when the leaves are in the primary phenological stages. To reduce this effect, different VIs such as the Soil-Adjusted Vegetation Index (SAVI), Optimized Soil-Adjusted Vegetation Index (OSAVI), Green Soil-Adjusted Vegetation Index (GSAVI), Green Optimized Soil-Adjusted Vegetation Index (GOSAVI), and Modified Soil-Adjusted Vegetation Index 2 (MSAVI2) have been used. These VIs take care of the contribution of the soil reflectance in VIs calculation, specifically in the leaf edge pixels which may have soil and vegetation information together, therefore, the pure soil pixels were removed. As a result, an empirical equation (Equation 1) was used to mask the unshaded and shaded bare soil pixels.

$$G_{index} = 2 * \text{Green} - \text{Red} - \text{Blue} \quad (1)$$

The G_{index} values greater than 0.06 were selected as vegetation pixels based on trial and error. Although NDVI has been used to remove soil pixels [47,48], it does not detect and remove shaded pixels. Therefore, the proposed G_{index} filter can remove all shaded and unshaded bare soil pixels precisely.

2.3.3. Harvesting Process

Corn grain was harvested by a two-row plot combine for the whole plot length. Rows 2 and 3 and rows 10 and 11 of each plot were combined. Some of the plots suffered extensive raccoon damage in 2018; hence, ears were harvested by hand from uniformly standing undamaged rows. Hand harvesting was performed by pulling ears from two 6.1-m row lengths of each harvest row pairs (rows 2, 3, and rows 10, 11). The regions of damage were skipped during the hand harvesting. All grain yield data were adjusted to 15.5% moisture content. Since yield data were collected for rows 2–3 and rows 10–11, VIs derived from pixels reciprocating the rows were calculated. The bare soil pixels between rows were eliminated by applying Equation 1, and then the spatial average of pixels was taken for each row. Figure 5 shows the spatial location of the rows within each treatment.

2.4. Outlier Detection

After calculating VIs for each phenological stage, outliers for each VI and at each stage were removed from the data set by utilizing a well-known z-score ($z = (x - \bar{x})/\sigma_x$) [49,50]. Here, the observations with z-score greater than 2.5 were considered as outlier data. The threshold number flexible between 2.5

and 3 were used to remove outliers [51]. A smaller threshold number of results in a greater selection of outliers. All in all, approximately 3–8% of the data for each growth stage were removed as outliers.

2.5. Feature Selection

The process of identifying the most important features is called “feature selection”. The random forest method is one of the most popular machine-learning methods used in data science workflows. This method is a combination of tree predictors used commonly as a tool for classification, regression, and ranking of candidate predictors [52,53]. In this research, the most important variables for each phenological stage were identified by the random forest feature selection method. The random forest method uses a training dataset and creates multiple subsets of the data randomly. Then, trees (samples) are used to create a ranking of classifiers and perform a vote for each predicted result. Finally, prediction results are selected which have the most votes [53]. Random forest is considered a highly accurate and robust method [52] because of the number of decision trees participating in the process. This method has acceptable predictive performance, low overfitting, and simple interpretability.

2.6. Statistical Analysis

Many of the statistical parametric tests such as correlation analysis, regression modeling, and analysis of variance (ANOVA) assume the data follow Gaussian distribution. In this research, the density plot and the Shapiro–Wilk’s test [54–57] were used to evaluate whether the data follow Gaussian distribution or not. Although some statisticians suggest that in case of the large sample size ($n > 30$), we can ignore the distribution of the data and use parametric tests [55,58], the observed yield data were not large enough (32 samples size for each year) and were not following the normal distribution; therefore, two approaches were used to make yield prediction models: (1) the data were normalized and then multiple regression models were fitted using the important features, and (2) a gradient boosting decision tree model was hired as a non-parametric method to estimate corn yield. Gradient boosting machines (GBMs) is a method of converting weak learners into strong learners like the random forest, however, in GBM the k th tree is trained from the first $k-1$ trees and updated the residual for the i th example of the difference between prediction and observations [59]. In other words, the predictors were sequentially trained and tried to correct the predecessors. One of the advantages of using GBM is that this method is highly customizable to the specific necessity of the application, such as being learned with regard to various loss functions [18].

For the multiple regression models, some of the input variables were not associated with the response variables that triggered excessive complexity in the final model. Therefore, possible combinations of the essential variables were used to fit different multiple regression models. Linear model selection was used to determine the number of significant variables that improve the model by maximizing the adjusted R^2 , minimizing Bayesian information criterion (BIC), and minimizing the cross-validated prediction error (C_p) [60]. Furthermore, cross-validation (CV) [61] was used as a backup method to ensure the predictors were correctly determined. Cross-validation is a method used in the selection of models to test the ability of different models in their accuracy in the prediction of results. In this research, the data were split into two subsets. Eighty percent of the data were used as training samples or the model-building set, and 20% of the data were used for prediction or as a validation set. Each variable was included in the model, and then the average of the cross-validation error was estimated. Overall, after removing outliers, the random forest and cross-validation methods were used to find the number of influential variables for predicting the corn yield. Random forest feature selection illustrates the importance of variables at each stage. Different variables were selected for each phenological stage of corn.

In this research, the software programs ArcMap 10.7.1 (Environmental Systems Research Institute, Inc. (ESRI), Redlands, CA, USA) [62], QGIS v.3.12.0 (Böschacherstrasse 10a CH-8624 Grüt (Gossau ZH), Zurich, Switzerland) [63], and R version 3.6 (R Core Team: Vienna, Austria) [64] were used to manipulate and analyze the data.

3. Results and Discussion

Graphical (density plot) and numerical (Shapiro–Wilk’s test) assessment of the normality of the data illustrated that the data does not follow a normal distribution (Figure 6). It can be observed that the corn yield data distribution shape does not match the normal distribution (dashed lines). Since the normality test is sensitive to sample size, therefore, it is important to combine visual inspection and significance tests in order to make the right decision. The Shapiro–Wilk’s test confirmed the same result and therefore, the data was normalized. The correlation between yield and 31 independent variables is shown in Figure 7 at V4-5 and VT stages.

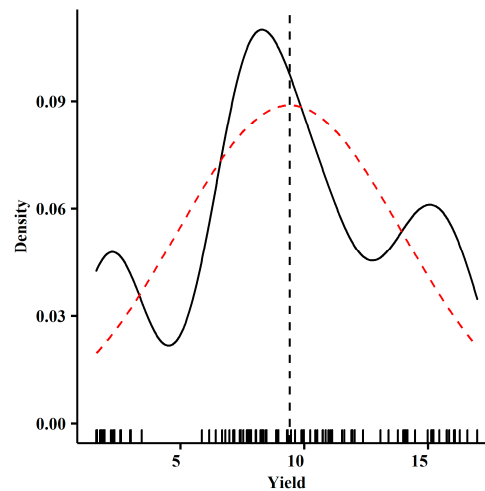


Figure 6. Yield density plot (solid line) and corresponding normal distribution (dash line).

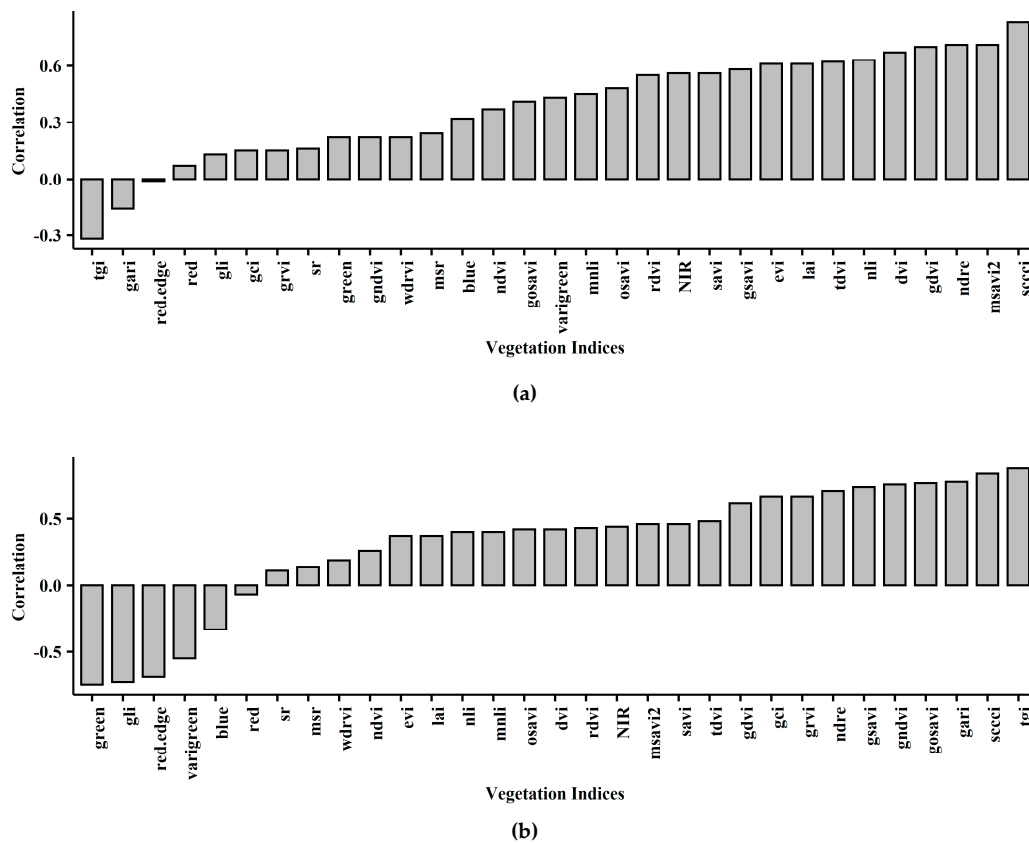


Figure 7. Correlation between corn yield and 31 variables (5 spectral bands and 26 vegetation indices (VIs): (a) at V4-5 (4-5 leaves with visible leaf collar); (b) at tasseling stage (VT).

The taller the data bar, the greater is the correlation between each variable and yield. As shown in Figure 7, the SCCCI, NDRE, MSAVI2, and Green Difference Vegetation Index (GDVI) at V4-5 stage and Triangular Greenness Index (TGI), SCCCI, Green Atmospherically Resistant Index (GARI), and GOSAVI at VT were more correlated with yield as a response variable. However, all of these variables were not included in the final model due to their interaction between the independent variables.

Random forest selected different features for each corn phenological stage. For instance, the SCCCI, NDRE, and MSAVI2 were the most striking features to predict the yield at the V4-5 stage (Figure 8a). The TGI, Green Leaf Index (GLI), $VARI_{green}$, and SCCCI were the most significant VI's to estimate yield at the VT stage (Figure 8b).

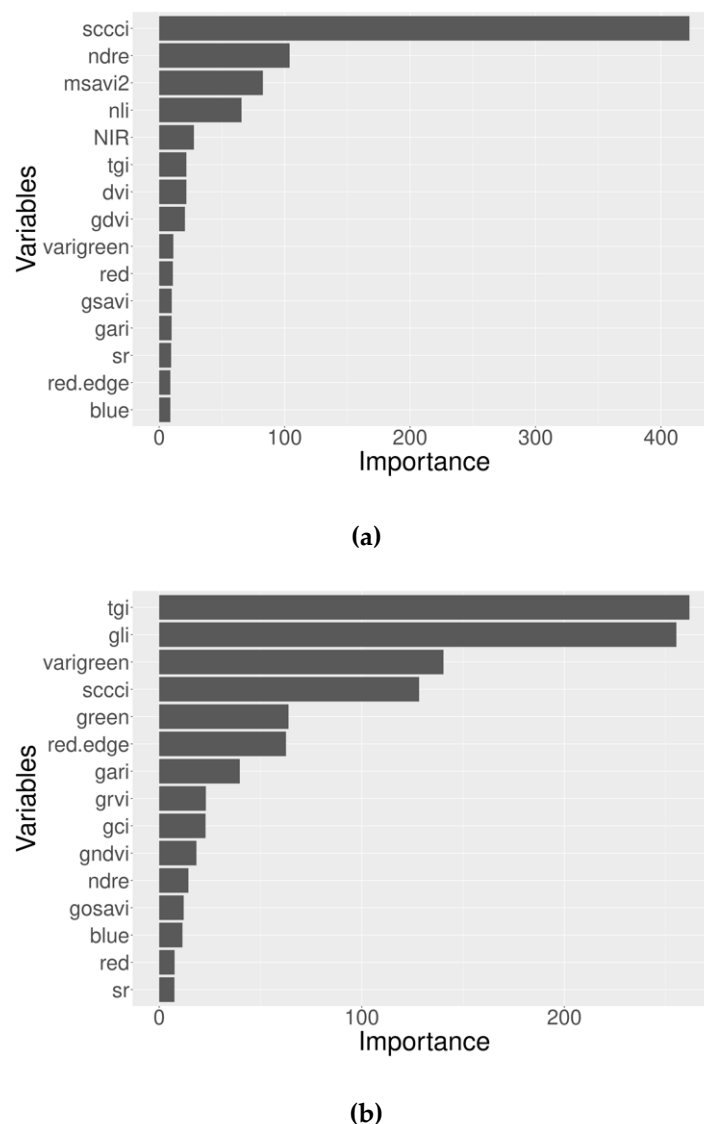


Figure 8. Feature importance by Random forest feature selection method: (a) at V4-5; (b) at VT.

Regarding the model selection method, for the VT stage, three predictors had a significant impact on increasing yield prediction accuracy (Figure 9). The three variables lead to almost the greatest adjusted- R^2 and the lowest BIC and Cp. Although adding a 4th variable increased the adjusted- R^2 or decreased the BIC and Cp, these improvements were not significant. Therefore, three predictors were used in the yield prediction model for the VT stage, which explained approximately 95% of the corn yield variation.

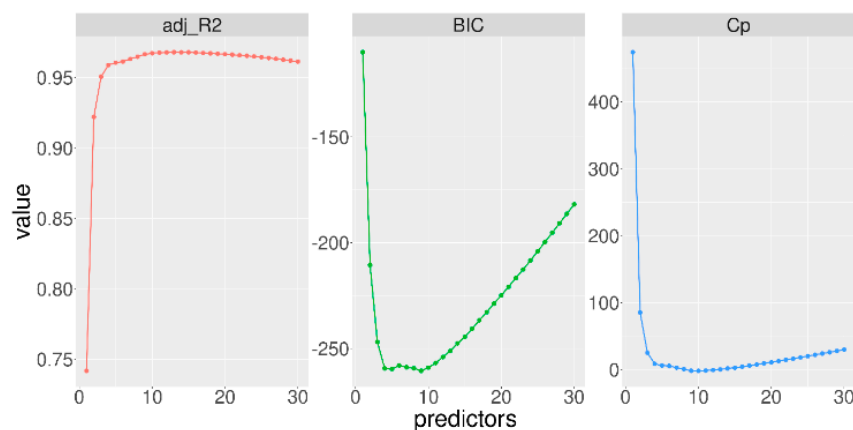


Figure 9. The linear model selection at the VT stage.

Moreover, the mean CV error confirmed that the same number of variables were needed for the final model (Figure 10). As a result, the best subset selection on the full dataset with the lowest mean square error (MSE) was the 3-variable model used to predict grain yield at the VT stage.

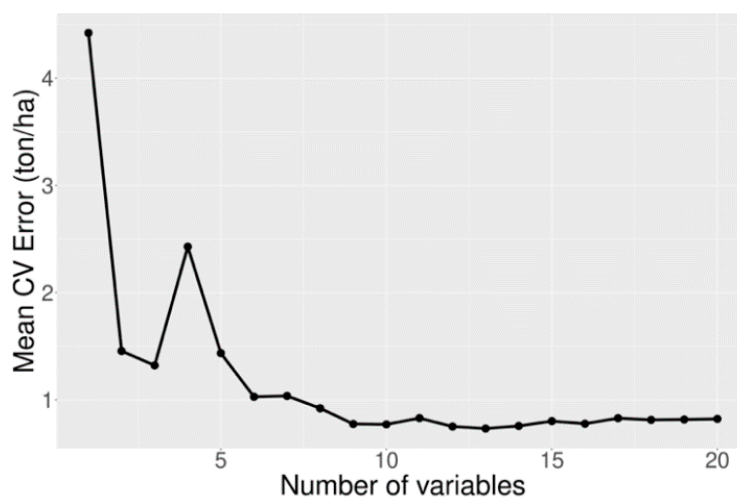


Figure 10. Mean cross-validation error versus the number of variables.

To achieve a mathematical yield prediction algorithm, multiple linear regression models were fitted for each phenological stage (Table 2).

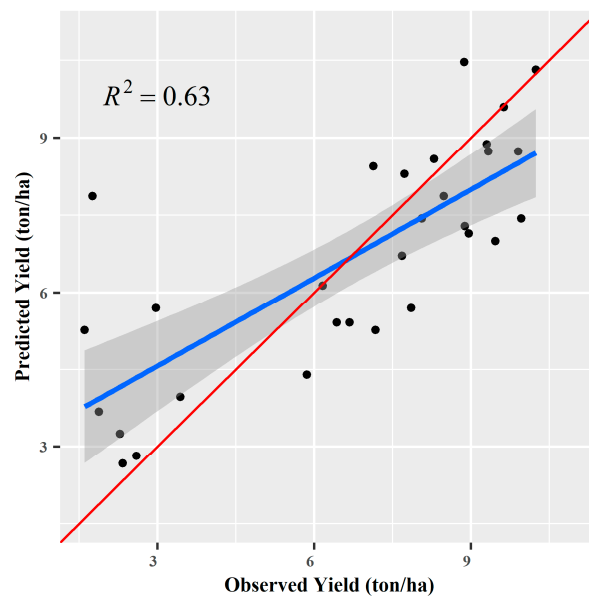
Table 2. Regression models and performance for each model to predict yield at different phenological stages.

Phenological Stage	Yield Prediction Models	R ² -adj
V3	Yield = − 23 + 144.4 OSAVI	0.63
V4-5	Yield = − 13.36 + 45.48 SCCCI	0.69
V6-7	Yield = − 161 + 590.3 GARI + 151.7 NDRE − 456.9 GNDVI	0.70
V10-11	Yield = − 22.64 + 68.93 SCCCI − 19.13 SAVI	0.90
VT	Yield = − 10.96 + 26.07 SCCCI − 68.25 GLI + 13.25 VARI _{green}	0.93

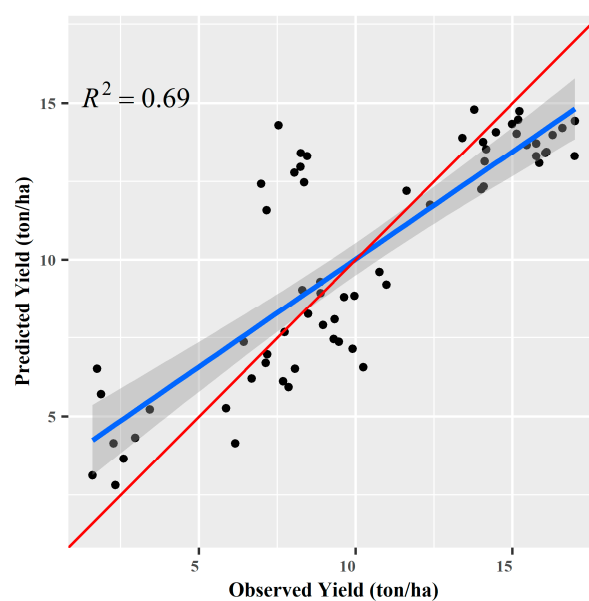
Although TGI was the most important feature at VT (Figure 8b), it was not statistically significant among other selected variables. Similarly, all these processing methods were applied for each of the five growth stages in order to predict grain yield. Since plant leaf area and metabolism are different at

phenological stages, the relationships between yield and spectral bands/VI are likely to differ; therefore, a model for each stage was developed. The coefficients of determination (R^2) for different models at each phenological stage are shown in Table 2. All the variables used in these models are statistically significant (at the $\alpha = 95\%$ significance level).

Furthermore, the measured yield data were compared to the fitted models to evaluate the performance of the algorithms. Figure 12a–e illustrate the scatter plot of observed yield values versus predicted yield values at different stages using multiple linear regression. In these figures, the blue line is the 1:1 line, which has a slope of 1, and the red line is the modeled prediction regression line. The regression line indicates how a response variable changes as predictors change. The increasing similarity in the slope and intercepts of the regression line to the 1:1 indicates a more significant model predictive capacity.



(a)



(b)

Figure 11. Cont.

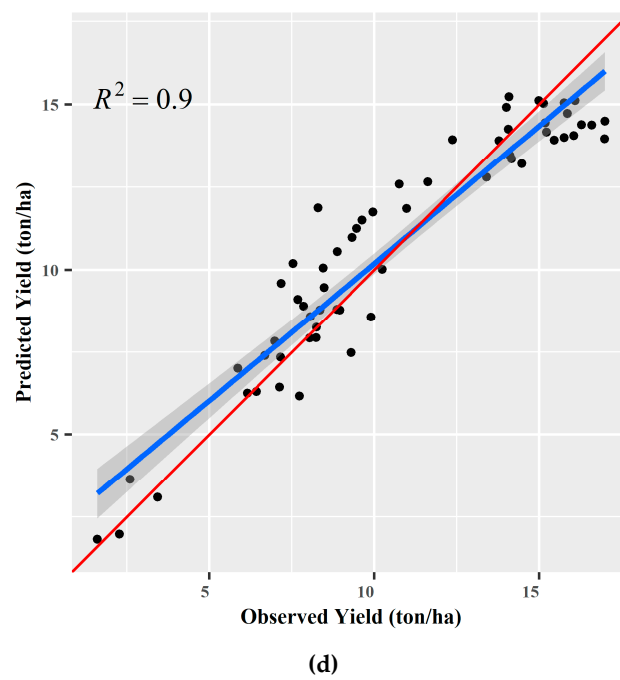
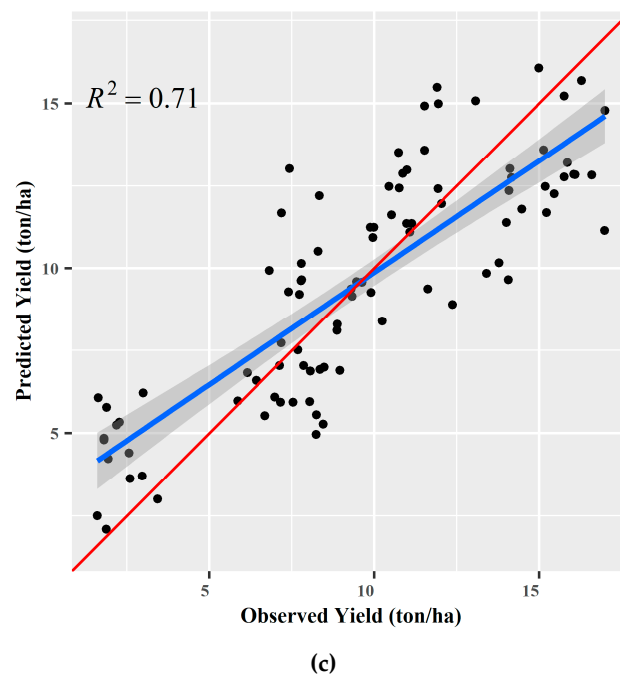


Figure 11. Cont.

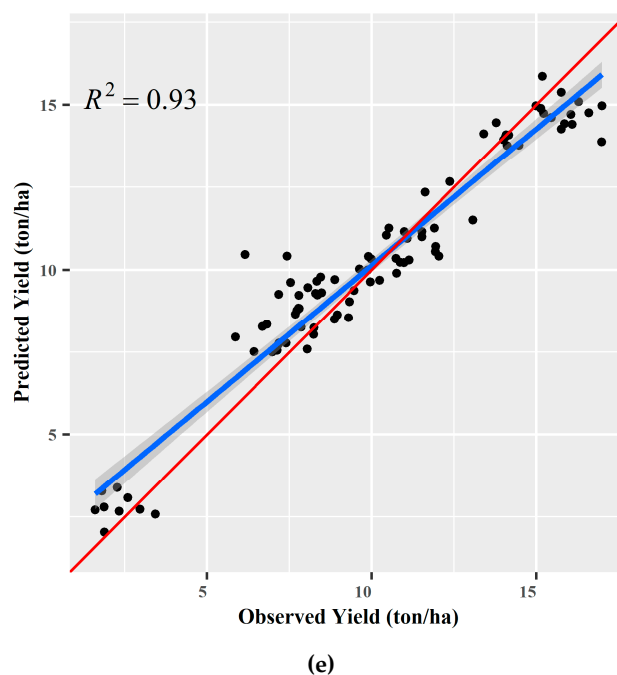


Figure 11. Scatter plot of observed versus predicted yield model of corn using multiple linear regression: (a) V3 (3 leaves with visible leaf collar); (b) V4-5; (c) V6-7; (d) V10; (e) VT stages.

The yield model at V3 with one single variable, OSAVI, resulted in the simplest yield estimation model, which has the advantage of simplicity and ease of calculation. As Rondeaux (1996) concluded, OSAVI excels in regions with sparse vegetation where the soil is visible through the plants [42,65] and, thus, utilization of a VI that corrects for soil interference such as OSAVI worked the best when corn leaf area was at the lowest and bare soil was the greatest among all the sampling times. This model predicted corn grain yield using only one variable at the V3 stage, and the model explained the 63% variation in the corn yield. At the V4-5 stage, the SCCC predicted grain yield with the highest level of accuracy. The variable selection method indicated that including more variables did not significantly improve the efficiency of the model at V3 and V4-5 stages. Adding a 2nd variable (VI) only increased the R^2 to 0.7 which was not significant. At the V6-7 stage, the GARI, NDRE, and Green Normalized Difference Vegetation Index (GNDVI) together resulted in the best prediction accuracy of grain yield. Although the NDVI is a commonly used index to predict yield, this index saturates when the leaf area index is greater than 1.5 [66]. The green NDVI (GNDVI) was most strongly related to grain yield because it has a broader dynamic range than NDVI and combines the green and NIR wavelengths, which are more strongly associated with leaf chlorophyll, N content, and grain yield than the red wavelength [22,35,67]. The SAVI is a VI related to biomass to predict yield in highly vegetated areas [66]. At the V10-11 stage, SAVI was one of the most effective variables in predicting yield with an R^2 of 0.90. The SCCC, GLI, and $VARI_{green}$ were found to provide the best predictive accuracy at the VT stage ($R^2 = 0.93$). Moderate to high R^2 values (0.63, 0.69, and 0.70) at V3, V4-5, and V6-7 were observed respectively, and high R^2 values (0.90 and 0.93) at V10 and VT were obtained, respectively. As a result, at V3 and V4-5, the models with a single index, and after V6-7, the predictive models with multiple indices produced the most solid relationships between observed and predicted grain yields. Similar results were reported for lettuce yield prediction [68]. To conclude, the VIs used in the predictive models (Table 2) were exchangeable with the next importance VIs (Figure 8); however, this replacement led to a reduction in the accuracy of the model by about 5%.

Assessment of the gradient-boosting models used for prediction of corn yield for five phenological stages suggested that VIs derived from MicaSense observations can predict corn yield with relatively higher accuracy (Figure 12a–e). The R^2 ranging from 0.84 at V3 to 0.97 at VT. Although GBM for

the V3 stage has relatively high R^2 , the root means square error is high (RMSE = 2.1 ton/ha). Compared with multiple regression models, ensemble learning models performed better estimation. For instance, at the V4 to V5 stages, gradient boosting consistently performed better in the prediction of corn yield during both the cross-validation and validation phase. The advantage of the non-parametric ensemble learning models over the regression-based model could be associated with the existence of non-linear relationships between the yield and VIs that random forest-based algorithms can integrate during model development. Additionally, the flexibility in the hyperparameters configuration of the boosting methods can result in higher performance in yield prediction.

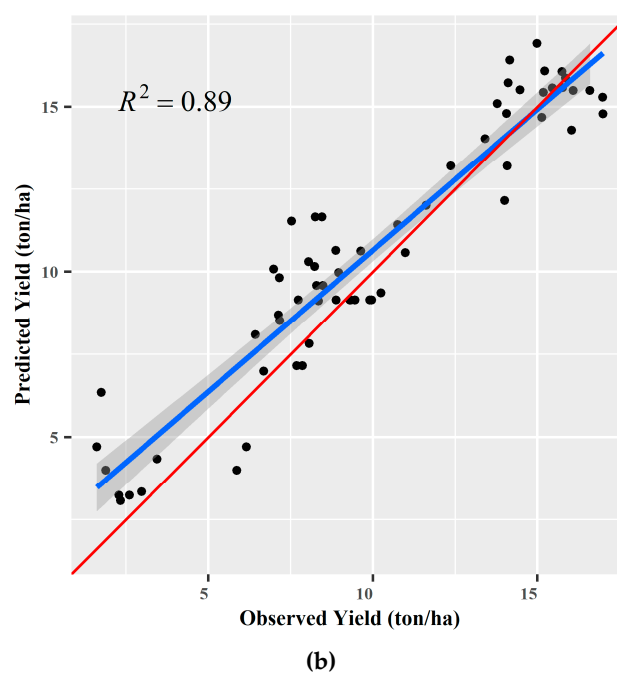
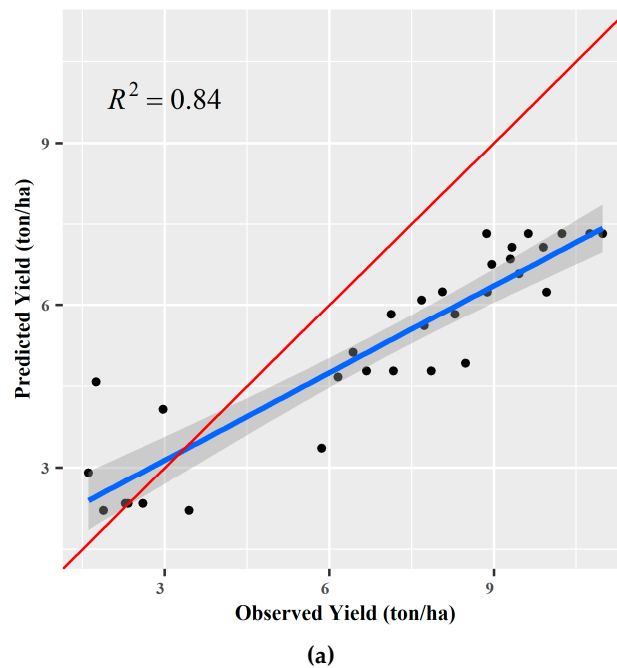


Figure 12. Cont.

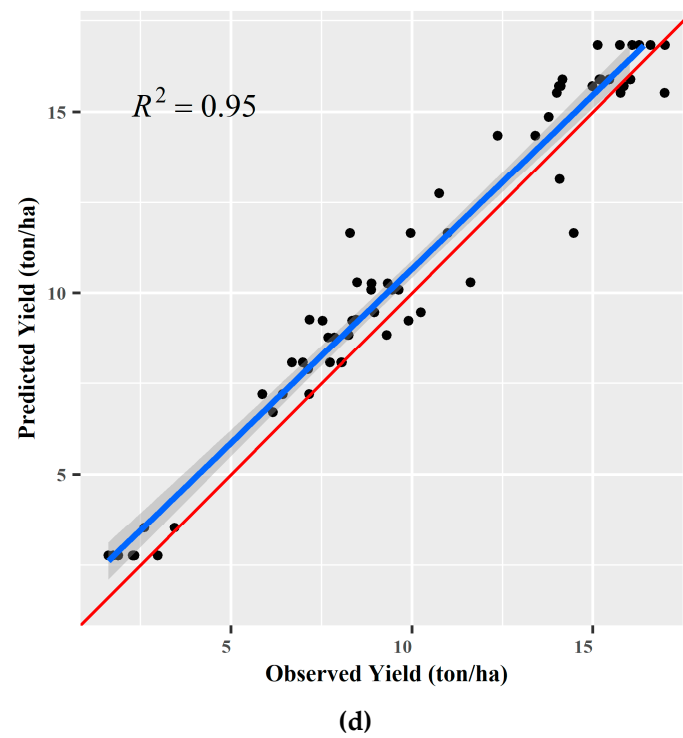
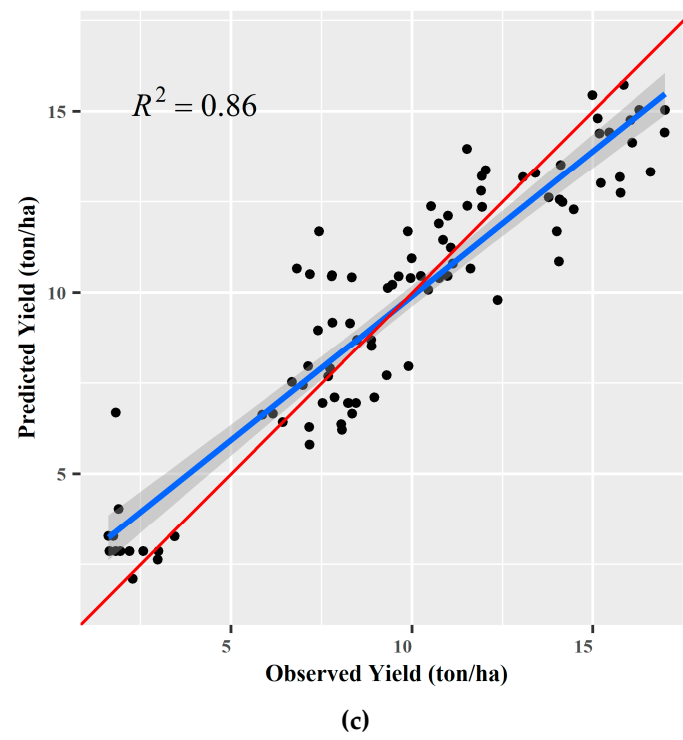


Figure 12. Cont.

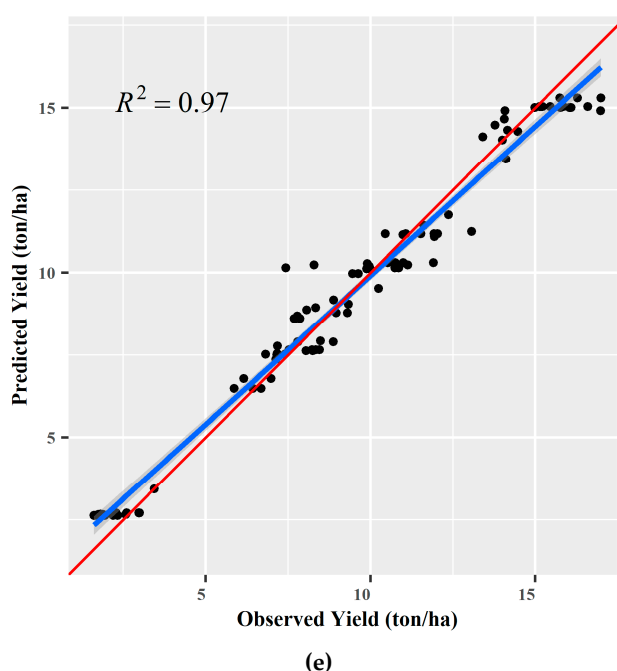


Figure 12. Scatter plot of observed versus predicted yield model of corn using gradient boosting machines: (a) V3; (b) V4-5; (c) V6-7; (d) V10; (e) VT stages.

Two critical results were observed from this study: first, as shown in Table 2, the SCCCI as a combined index seems to be the most appropriate index for predicting yield [32]. Second, as corn development progressed, the GBM and regression models predicted final grain yield more accurately, indicating that there is some relationship between the VI and the biomass content of the corn. It appears that the variation in fertilizer N rates provided a good basis for differentiating growth and, ultimately, grain yield, which provided adequate sensitivity for model building and testing.

4. Conclusions

Spatial mapping of corn yield is required for proper field management which is needed for improving crop productivity. This study utilized five distinct wavelengths and 26 calculated VIs as input variables to develop regression-based and tree-based learning models at different corn phenological stages to predict corn grain yield at V3, V4-5, V6-7, V10-11, and VT phenological stages. The influence of the variables was found to vary with the phenological stages. In general, the VI that contributed to the majority of the models was SCCCI, suggesting the importance of red-edge-based VIs during yield estimation. At V3 and V4-5 stages, OSAVI and SCCCI were the single dominant features in the yield-predicting models, respectively. The most suitable GBM models with the greatest R^2 values of 0.97 and 0.95 resulted at the V10 and VT stages, respectively. Similarly, the highest R^2 values were obtained at the same stages using regression-based models. When the models' performances were compared for individual stages in both regression-based and tree-based models, however, the accuracies were higher as corn development progressed. One of the goals of this research was to find the models for each stage with minimum error (maximum R^2) by using the appropriate number of predictors. The methodology used in this research can be extended to predict yield for other crops or in other regions as well, where yield prediction is mainly reliant on weather and climatic conditions.

The accuracy of the models in this research might have been affected by different things, including a smaller number of yield samples collected for each year and the use of limited machine-learning algorithms. In this study, we observed considerable improvement in yield prediction with the use of ensemble learning model than the linear regression algorithm. For corn yield prediction, spectral information, preprocessing, and preprocessing algorithms were important.

In conclusion, results of this research demonstrated the use of the smallest number of predictive variables that are statistically significant which resulted in an improved explanation for corn yield prediction. Contributing to the accuracy in the predictive capacity of these models included the following: preprocessing of data, including removal of soil pixels, deletion of 3–8% of outliers before conducting the statistical analysis, evaluating for appropriate variables, and selecting appropriate machine-learning model.

Author Contributions: J.V. designed the experiments, provided the yield data, and revised the paper; R.B. and H.L. processed the imagery and wrote computer codes; R.B. performed the analysis, set up the corn yield prediction models, and wrote the manuscript; R.P. assisted in data collection, G.C.B. is the principal investigator, and supervised the project. All authors contributed, reviewed, and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study is funded by the Mississippi Corn Promotion Board (MCPB).

Acknowledgments: Multispectral data were collected and preprocessed by the Geosystems Research Institute (GRI), Mississippi State University.

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

References

1. Rouse, J.W.; Hass, R.H.; Schell, J.A.; Deering, D.W.; Harlan, J.C. *Monitoring the Vernal Advancement and Retrogradation (GreenWave Effect) of Natural Vegetation [Great Plains Corridor]*; NASA: Washington, DC, USA, 1974.
2. Santamar, L. Modeling Biomass Production in Seasonal Wetlands Using MODIS NDVI Land Surface Phenology. *Remote Sens.* **2017**, *9*, 392. [[CrossRef](#)]
3. Hogrefe, K.R.; Patil, V.P.; Ruthrauff, D.R.; Meixell, B.W.; Budde, M.E.; Hupp, J.W.; Ward, D.H. Normalized Difference Vegetation Index as an Estimator for Abundance and Quality of Avian Herbivore Forage in Arctic Alaska. *Remote Sens.* **2017**, *9*, 1234. [[CrossRef](#)]
4. Bronson, K.F.; Booker, J.D.; Keeling, J.W.; Boman, R.K.; Wheeler, T.A.; Lascano, R.J.; Nichols, R.L. Cotton Canopy Reflectance at Landscape Scale as Affected by Nitrogen Fertilization. *Agron. J.* **2005**, *97*, 654–660. [[CrossRef](#)]
5. Chen, J.M. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Can. J. Remote Sens.* **1996**, *22*, 229–242. [[CrossRef](#)]
6. Jackson, R.D.; Huete, A.R. Interpreting Vegetation Indices. *Prev. Vet. Med.* **1991**, *11*, 185–200. [[CrossRef](#)]
7. Báez-González, A.D.; Chen, P.Y.; Tiscareño-López, M.; Srinivasan, R. Using Satellite and Field Data with Crop Growth Modeling to Monitor and Estimate Corn Yield in Mexico. *Crop Sci.* **2002**, *42*, 1943–1949. [[CrossRef](#)]
8. Shiu, Y.S.; Chuang, Y.C. Yield Estimation of Paddy Rice Based on Satellite Imagery: Comparison of Global and Local Regression Models. *Remote Sens.* **2019**, *11*, 111. [[CrossRef](#)]
9. Silvestro, P.C.; Pignatti, S.; Pascucci, S.; Yang, H.; Li, Z.; Yang, G.; Huang, W.; Casa, R. Estimating Wheat Yield in China at the Field and District Scale from the Assimilation of Satellite Data into the Aquacrop and Simple Algorithm for Yield (SAFY) Models. *Remote Sens.* **2017**, *9*, 509. [[CrossRef](#)]
10. Sharma, L.K.; Bu, H.; Denton, A.; Franzen, D.W. Active-Optical Sensors Using Red NDVI Compared to Red Edge NDVI for Prediction of Corn Grain Yield in North Dakota, U.S.A. *Sensors* **2015**, *15*, 27832–27853. [[CrossRef](#)]
11. Wang, T.; Shi, J.; Letu, H.; Ma, Y.; Li, X.; Zheng, Y. Detection and Removal of Clouds and Associated Shadows in Satellite Imagery Based on Simulated Radiance Fields. *J. Geophys. Res. Atmos.* **2019**, *124*, 7207–7225. [[CrossRef](#)]
12. Bondi, E.; Salvaggio, C.; Montanaro, M.; Gerace, A.D. Calibration of UAS Imagery Inside and Outside of Shadows for Improved Vegetation Index Computation. In *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping, Proceedings of the SPIE, Baltimore, MD, USA, 17–26 May 2016*; SPIE: Bellingham, WA, USA, 2016; Volume 9866. [[CrossRef](#)]
13. Zhang, C.; Kovacs, J.M. The Application of Small Unmanned Aerial Systems for Precision Agriculture: A Review. *Precis. Agric.* **2012**, *13*, 693–712. [[CrossRef](#)]

14. Iizuka, K.; Itoh, M.; Shiodera, S.; Matsubara, T.; Dohar, M.; Watanabe, K. Advantages of Unmanned Aerial Vehicle (UAV) Photogrammetry for Landscape Analysis Compared with Satellite Data: A Case Study of Postmining Sites in Indonesia. *Cogent Geosci.* **2018**, *4*, 1498180. [[CrossRef](#)]
15. Gnädinger, F.; Schmidhalter, U. Digital Counts of Maize Plants by Unmanned Aerial Vehicles (UAVs). *Remote Sens.* **2017**, *9*, 544. [[CrossRef](#)]
16. Qin, Z.; Myers, D.B.; Ransom, C.J.; Kitchen, N.R.; Liang, S.Z.; Camberato, J.J.; Carter, P.R.; Ferguson, R.B.; Fernandez, F.G.; Franzen, D.W.; et al. Application of Machine Learning Methodologies for Predicting Corn Economic Optimal Nitrogen Rate. *Agron. J.* **2018**, *110*, 2596–2607. [[CrossRef](#)]
17. Khanal, S.; Fulton, J.; Klopfenstein, A.; Douridas, N.; Shearer, S. Integration of High Resolution Remotely Sensed Data and Machine Learning Techniques for Spatial Prediction of Soil Properties and Corn Yield. *Comput. Electron. Agric.* **2018**, *153*, 213–225. [[CrossRef](#)]
18. Natekin, A.; Knoll, A. Gradient Boosting Machines, a Tutorial. *Front. Neurobot.* **2013**, *7*. [[CrossRef](#)]
19. Gitelson, A.A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *J. Plant Physiol.* **2004**, *161*, 165–173. [[CrossRef](#)]
20. Gitelson, A.A.; Kaufman, Y.J.; Stark, R.; Rundquist, D. Novel Algorithms for Remote Estimation of Vegetation Fraction. *Remote Sens. Environ.* **2002**, *80*, 76–87. [[CrossRef](#)]
21. Erdle, K.; Mistele, B.; Schmidhalter, U. Field Crops Research Comparison of Active and Passive Spectral Sensors in Discriminating Biomass Parameters and Nitrogen Status in Wheat Cultivars. *Field Crops Res.* **2011**, *124*, 74–84. [[CrossRef](#)]
22. Shanahan, J.F.; Schepers, J.S.; Francis, D.D.; Varvel, G.E.; Wilhelm, W.W.; Tringe, J.M.; Schlemmer, M.R.; Major, D.J. Use of Remote-Sensing Imagery to Estimate Corn Grain Yield. *Agron. J.* **2001**, *93*, 583–589. [[CrossRef](#)]
23. Barzin, R.; Shirvani, A.; Lotfi, H. Estimation of Daily Average Downward Shortwave Radiation from MODIS Data Using Principal Components Regression Method: Fars Province Case Study. *Int. Agrophys.* **2017**, *31*, 23–34. [[CrossRef](#)]
24. González-audicana, M.; Saleta, J.L.; Catalán, R.G.; García, R. Fusion of Multispectral and Panchromatic Images Using Improved IHS and PCA Mergers Based on Wavelet Decomposition. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1291–1299. [[CrossRef](#)]
25. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
26. Wu, W. The Generalized Difference Vegetation Index (GDVI) for Dryland Characterization. *Remote Sens.* **2014**, *6*, 1211–1233. [[CrossRef](#)]
27. Roujean, J.L.; Breon, F.M. Estimating PAR Absorbed by Vegetation from Bidirectional Reflectance Measurements. *Remote Sens. Environ.* **1995**, *51*, 375–384. [[CrossRef](#)]
28. Bannari, A.; Asalhi, H.; Teillet, P.M. Transformed Difference Vegetation Index (TDVI) for Vegetation Cover Mapping. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 24–28 June 2002; Volume 5, pp. 3053–3055. [[CrossRef](#)]
29. Pathak, R.; Barzin, R.; Bora, G.C. Data-Driven Precision Agricultural Applications Using Field Sensors and Unmanned Aerial Vehicle. *Int. J. Precis. Agric. Aviat.* **2018**, *1*, 19–23. [[CrossRef](#)]
30. Gitelson, A.A.; Merzlyak, M.N. Remote Estimation of Chlorophyll Content in Higher Plant Leaves. *Int. J. Remote Sens.* **1998**, *18*, 2691–2697. [[CrossRef](#)]
31. Gitelson, A.A.; Merzlyak, M.N. Quantitative Estimation of Chlorophyll-a Using Reflectance Spectra: Experiments with Autumn Chestnut and Maple Leaves. *J. Photochem. Photobiol.* **1994**, *22*, 247–252. [[CrossRef](#)]
32. Raper, T.B.; Varco, J.J. Canopy-Scale Wavelength and Vegetative Index Sensitivities to Cotton Growth Parameters and Nitrogen Status. *Precis. Agric.* **2015**, *16*, 62–76. [[CrossRef](#)]
33. Matsushita, B.; Yang, W.; Chen, J.; Onda, Y.; Qiu, G. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-Density Cypress Forest. *Sensors* **2007**, *7*, 2636–2651. [[CrossRef](#)]
34. 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](#)]

35. 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](#)]
36. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [[CrossRef](#)]
37. Louhaichi, M.; Borman, M.M.; Johnson, D.E. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. *Geocarto Int.* **2001**, *16*, 65–70. [[CrossRef](#)]
38. Raymond Hunt, E.; 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](#)]
39. Vesco, L.; Gianelle, D. Using the MIR Bands in Vegetation Indices for the Estimation of Grassland Biophysical Parameters from Satellite Remote Sensing in the Alps Region of Trentino (Italy). *Adv. Space Res.* **2008**, *41*, 1764–1772. [[CrossRef](#)]
40. Gong, P.; Pu, R.; Biging, G.S.; Larrieu, M.R. Estimation of Forest Leaf Area Index Using Vegetation Indices Derived from Hyperion Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1355–1362. [[CrossRef](#)]
41. Feng, W.; Wu, Y.; He, L.; Ren, X.; Wang, Y.; Hou, G.; Wang, Y.; Liu, W.; Guo, T. An Optimized Non-Linear Vegetation Index for Estimating Leaf Area Index in Winter Wheat. *Precis. Agric.* **2019**, *20*, 1157–1176. [[CrossRef](#)]
42. Rondeaux, G.; Steven, M.; Baret, F. Optimization of Soil-Adjusted Vegetation Indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [[CrossRef](#)]
43. Sripada, R.P.; Schmidt, J.P.; Dellinger, A.E.; Beegle, D.B. Evaluating Multiple Indices from a Canopy Reflectance Sensor to Estimate Corn N Requirements. *Agron. J.* **2008**, *100*, 1553–1561. [[CrossRef](#)]
44. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modify Soil Adjust Vegetation Index. *Remote Sens. Environ.* **1994**, *126*, 119–126. [[CrossRef](#)]
45. Fraser, R.H.; Latifovic, R. Mapping Insect-Induced Tree Defoliation and Mortality Using Coarse Spatial Resolution Satellite Imagery. *Int. J. Remote Sens.* **2005**, *26*, 193–200. [[CrossRef](#)]
46. Morales, A.; Nielsen, R.; Camberato, J. Effects of Removing Background Soil and Shadow Reflectance Pixels from RGB and NIR-Based Vegetative Index Maps. In Proceedings of the Purdue GIS Day, Purdue University, West Lafayette, IN, USA, 7 November 2019.
47. Sader, S.A.; Winne, J.C. RGB-NDVI Colour Composites for Visualizing Forest Change Dynamics. *Int. J. Remote Sens.* **1992**, *13*, 3055–3067. [[CrossRef](#)]
48. Gallo, B.C.; Demattê, J.A.M.; Rizzo, R.; Safanelli, J.L.; Mendes, W.d.S.; Lepsch, I.F.; Sato, M.V.; Romero, D.J.; Lacerda, M.P.C. Multi-Temporal Satellite Images on Topsoil Attribute Quantification and the Relationship with Soil Classes and Geology. *Remote Sens.* **2018**, *10*, 1571. [[CrossRef](#)]
49. Torres, J.M.; Nieto, P.J.G.; Alejano, L.; Reyes, A.N. Detection of Outliers in Gas Emissions from Urban Areas Using Functional Data Analysis. *J. Hazard. Mater.* **2011**, *186*, 144–149. [[CrossRef](#)]
50. Schubert, E.; Kriegel, H. Generalized Outlier Detection with Flexible Kernel Density Estimates. In Proceedings of the 2014 SIAM International Conference on Data Mining, Philadelphia, PA, USA, 24–26 April 2014; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2014; pp. 542–550.
51. Nurunnabi, A.; West, G.; Belton, D. Outlier Detection and Robust Normal-Curvature Estimation in Mobile Laser Scanning 3D Point Cloud Data. *Pattern Recognit.* **2015**, *48*, 1404–1419. [[CrossRef](#)]
52. Janitza, S.; Tutz, G.; Boulesteix, A.L. Random Forest for Ordinal Responses: Prediction and Variable Selection. *Comput. Stat. Data Anal.* **2016**, *96*, 57–73. [[CrossRef](#)]
53. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
54. Kox, M.A.R.; Lüke, C.; Fritz, C.; van den Elzen, E.; van Alen, T.; Op den Camp, H.J.M.; Lamers, L.P.M.; Jetten, M.S.M.; Ettwig, K.F. Effects of Nitrogen Fertilization on Diazotrophic Activity of Microorganisms Associated with Sphagnum Magellanicum. *Plant Soil* **2016**, *406*, 83–100. [[CrossRef](#)]
55. Akbarzadeh Baghban, A.; Younespour, S.; Jambarsang, S.; Yousefi, M.; Zayeri, F.; Azizi Jalilian, F. How to Test Normality Distribution for a Variable: A Real Example and a Simulation Study. *J. Paramed. Sci.* **2013**, *4*, 2008–4978.
56. Vellidis, G.; Tucker, M.; Perry, C.; Reckford, D.; Butts, C.; Henry, H.; Liakos, V.; Hill, R.W.; Edwards, W. A Soil Moisture Sensor-Based Variable Rate Irrigation Scheduling System. In *Precision Agriculture '13*; Wageningen Academic Publishers: Wageningen, The Netherlands, 2013.

57. Razali, N.M.; Wah, Y.B. Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *J. Stat. Model. Anal.* **2011**, *2*, 21–33. [CrossRef]
58. Ghasemi, A.; Zahediasl, S. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *Int. J. Endocrinol. Metab.* **2012**, *10*, 486–489. [CrossRef] [PubMed]
59. Friedman, J.H. Stochastic Gradient Boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]
60. Hastie, T.; Tibshirani, R.; Jerome, F. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*; Springer Science & Business Media: Berlin, Germany, 2004.
61. Kutner, M.H.; Nachtsheim, C.J.; Neter, J.; Li, W. *Applied Linear Statistical Models*, 5th ed.; McGraw-Hill Irwin: New York, NY, USA, 2005.
62. Environmental Systems Research Institute ESRI. ArcGIS Desktop: Release 10. Redlands, CA, USA, Using ArcMap. 2019. Available online: <https://qgis.org/en/site/> (accessed on 22 July 2020).
63. QGIS Development Team: 2020. QGIS.org (2020). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available online: <https://www.esri.com/news/arcnews/spring12articles/introducing-arcgis-101.html> (accessed on 22 July 2020).
64. R Core Team. *A Language and Environment for Statistical Computing: R Foundation for Statistical Computing*; R Core Team: Vienna, Austria, 2019.
65. Hatfield, J.L.; Prueger, J.H. Value of Using Different Vegetative Indices to Quantify Agricultural Crop Characteristics at Different Growth Stages under Varying Management Practices. *Remote Sens.* **2010**, *2*, 562–578. [CrossRef]
66. Trotter, T.F.; Frazier, P.S.; Trotter, M.G.; Lamb, D.W. Objective Biomass Assessment Using an Active Plant Sensor (Crop Circle), Preliminary Experiences on a Variety of Agricultural Landscapes. In Proceedings of the Ninth International Conference on Precision Agriculture, Denver, CO, USA, 20–23 July 2008; Khosla, R., Ed.; Colorado State University: Fort Collins, CO, USA, 2008.
67. Rattanakaew, T. *Utilization of Canopy Reflectance to Predict Yield Response of Corn and Cotton to Varying Nitrogen Rates*; Mississippi State University: Starkville, MS, USA, 2015.
68. Kizil, Ü.; Genç, L.; İnalpulat, M.; Şapolyo, D.; Mırık, M. Lettuce (*Lactuca Sativa* L.) Yield Prediction under Water Stress Using Artificial Neural Network (ANN) Model and Vegetation Indices. *Žemdirbystė= Agriculture* **2012**, *99*, 409–418.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).