remote sensing

MDPI

Communication

Using Canopy Measurements to Predict Soybean Seed Yield

Peder K. Schmitz * and Hans J. Kandel

Department of Plant Sciences, North Dakota State University, Fargo, ND 58105, USA; Hans.Kandel@ndsu.edu
* Correspondence: peder.schmitz@ndsu.edu or pederschmitz@gmail.com

Abstract: Predicting soybean [*Glycine max* (L.) Merr.] seed yield is of interest for crop producers to make important agronomic and economic decisions. Evaluating the soybean canopy across a range of common agronomic practices, using canopy measurements, provides a large inference for soybean producers. The individual and synergistic relationships between fractional green canopy cover (FGCC), photosynthetically active radiation (PAR) interception, and a normalized difference vegetative index (NDVI) measurements taken throughout the growing season to predict soybean seed yield in North Dakota, USA, were investigated in 12 environments. Canopy measurements were evaluated across early and late planting dates, 407,000 and 457,000 seeds ha^{-1} seeding rates, 0.5 and 0.8 relative maturities, and 30.5 and 61 cm row spacings. The single best yield predictor was an NDVI measurement at R5 (beginning of seed development) with a coefficient of determination of 0.65 followed by an FGCC measurement at R5 (R^2 = 0.52). Stepwise and Lasso multiple regression methods were used to select the best prediction models using the canopy measurements explaining 69% and 67% of the variation in yield, respectively. Including plant density, which can be easily measured by a producer, with an individual canopy measurement did not improve the explanation in yield. Using FGCC to estimate yield across the growing season explained a range of 49% to 56% of yield variation, and a single FGCC measurement at R5 (R^2 = 0.52) being the most efficient and practical method for a soybean producer to estimate yield.

Keywords: soybean yield prediction; fractional green canopy cover; NDVI; PAR; stepwise multiple regression; lasso multiple regression

check for updates

Citation: Schmitz, P.K.; Kandel, H.J. Using Canopy Measurements to Predict Soybean Seed Yield. *Remote Sens.* **2021**, *13*, 3260. https://doi.org/10.3390/rs13163260

Academic Editor: Shawn C. Kefauver

Received: 24 June 2021
Accepted: 13 August 2021
Published: 18 August 2021

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

Copyright: © 2021 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

Soybean is a major crop in the north-central USA region, with the states North Dakota, South Dakota, and Minnesota producing about 20% of the total soybean production [1]. Predicting the yield of crops such as soybean provides crucial information to producers, consultants, and economists for improving crop management decisions and subsequent profit. Early crop yield estimation using non-destructive measures can provide additional benefits to scientists and breeding programs furthering the identification of advantageous agronomic practices or high-yielding genotypes. Predicting soybean yield through a handheld or remote sensing canopy measurements is of high interest as the soybean canopy often reflects the progress and development of the crop. In addition, remote sensing images often suffer from various degradation, noise effects, or variability in image processing [2]. However, most methods used to quantify or estimate canopy cover are unpractical for soybean producers to use or equipment is expensive.

The normalized difference vegetative index (NDVI) is calculated from reflectance measurements in the red and near-infrared spectrums [3]. These reflectance measurements have been proven to indicate environmental and nutrient inadequacies [4–6]. A flexible Fourier transform model to predict yield using NDVI has been used previously but did not account for NDVI on a pixel basis [7]. Using NDVI to predict yield has provided variable results, Ma et al. [8] found NDVI had R^2 values between 0.65 and 0.80 for soybean yield during the first year of an experiment and then explained 45 and 70% of variation the subsequent year. Similar coefficients of determination of 0.63 and 0.65 were reported in

Wisconsin and Indiana, USA, by Mourtzinis et al. [9]. Several studies have shown NDVI explaining high amounts of variation for corn (*Zea mays* L.), cover crops, rice (*Oryza sativa* L.), and wheat (*Triticum aestivum*, L. emend. Thell.) [10–14].

Canopy cover is a useful proxy measurement for light interception potential and crop productivity. Maximum photosynthesis is achieved when plants maximize light interception and utilization of photosynthetic radiation [15,16]. Light interception can be quantified with methods such as quantum line sensors [17], approximated by fractional green canopy cover (FGCC) from pictures using the Canopeo app, as demonstrated by Patrignani and Ochsner [18], and leaf area index (LAI). Light interception measurements using quantum line sensors are considerably more time consuming, compared to measuring FGCC using the Canopeo app. In addition, precise LAI measurements require plant destruction in order not to overestimate the LAI in dense canopies [19]. The NDVI can also be used to approximate fractional canopy coverage but is not a useful substitute for above-ground biomass measurements [20]. Ma et al. [8] reported that plant density had no effect on the yield and NDVI relationship during the soybean reproductive growth stages. However, it would be useful to measure NDVI throughout the entire growing season to determine if early season plant density and NDVI measurements can improve seed yield prediction.

Estimating and predicting crop yields using canopy cover measurements is of interest to producers. Crop growth stage [21], row spacing [22], and canopy structure [23] can affect light interception, FGCC, and NDVI. Measurements such as NDVI and FGCC can predict yields for wheat [21], rice [24], and soybean [8]. Although previous research has evaluated the relationship between established stand, canopy measurements, and seed yield, this research evaluated combining measurements on an established stand, light interception, green canopy cover quantification, and NDVI to potentially allow for better yield prediction and provide a useful application in soybean production.

The objective of this experiment was to determine if measurements of canopy development could be used to predict soybean yield and if canopy measurements can predict yield, to determine the most accurate and most practical strategy for yield prediction.

2. Materials and Methods

Data were collected in the 2019 and 2020 cropping seasons at Casselton, Prosper, and Fargo, North Dakota, USA. Location and soil characteristics can be found in Table 1. Location and year were combined and are termed "environment", for a total of 12 environments. At the Fargo location, there were four experiments each year. Two experiments were on a tile-drained soil, and the other two were on a non-tile drained, and each soil drainage type had a 30.5 cm and a 61 cm row spacing experiment.

Table 1. Soil series, soil taxonomy, tillage system used, previous crop, and location coordinates for 12 environments in North Dakota, USA, in 2019 and 2020.

Location	Soil Series	Soil Taxonomy	Tillage	PC [1]	GPS
Casselton	Kindred	Fine-silty, mixed, superactive, frigid Typic Endoaquolls	CT	SB	46.882, −97.251
	Bearden	Fine-silty, mixed, superactive, frigid Aeric Calciaquolls			
Fargo	Fargo	Fine, smectitic, frigid Typic Epiaquerts	NT	W	46.932, −96.859
	Ryan	Fine, smectitic, frigid Typic Natraquerts			
Prosper	Bearden	Fine-silty, mixed, superactive, frigid Aeric Calciaquolls	CT	W	47.001, −97.112.
	Lindaas	Fine, smectitic, frigid Typic Argiaquolls			

[1] PC, previous crop; GPS, GPS coordinates. CT, conventional tillage; NT, no till; SB, soybean; W, hard red spring wheat.

Each experiment was a randomized complete block with a split-plot arrangement with four replicates. The whole plot was planting date and the sub-plots were a factorial combination the two of cultivars and seeding rate. Planting dates were at an optimal time in mid-May and a late planting date of two weeks thereafter. Seeding rates were 407,500 and 457,000 germinable seeds ha^{-1} based on rates of previous research results [25,26].

Cultivars used were AG 05X9 (0.5 maturity group) and AG 08X8 (0.8 maturity group) (Asgrow, Bayer Crop Science, Creve Coeur, MO, USA), which are adapted to the region [27]. Two row spacings (30.5 and 61 cm) were evaluated to represent the common range of row spacings used in North Dakota, USA [28]. A range of planting dates, seeding rates, relative maturities, and row spacing were used to create a large inference of common production practices in the North Dakota and northwestern Minnesota, USA, and Manitoba, Canada, region. The planting date, relative maturity, and seeding rate data were combined across all 12 environments. Early planting date treatments were seeded once soil temperatures reached 10 °C in early to mid-May but not earlier than five days prior to the last historical projected frost date, using a Hege 1000 no-till planter (Hege Company, Waldenberg, Germany), with 30.5 cm row spacing at Casselton and Prosper and 30.5 and 61 cm at the Fargo location. The second planting date was two to three weeks after the first planting, depending on field conditions. The plot size for the experimental unit was 1.52 m by 5.47 m. Soils test data are provided in Table 2. Fertility was not considered a limiting factor for yield [29], and experiments were kept weed-free using the herbicide Glyphosate [N-(phosphonomethyl) glycine] (Bayer Crop Science, St. Louis, MO, USA).

Table 2. Planting date and soil test results for soybean environments in 2019 and 2020.

Location	Planting Date							
	1	2	Depth	NO_3-N	P	K	pH	OM
	DOY [1]		cm	kg ha^{-1}	mg kg^{-1}			g kg^{-1}
				2019				
Casselton	137	154	0–15	16	8	368	7.4	5.2
			15–61	37	7	303	7.5	3.9
Fargo	137	154	0–15	8	15	495	7.8	5.8
			15–61	14	5	300	7.8	4.0
Prosper	136	149	0–15	35	20	232	7.9	3.4
			15–61	57	6	176	8.2	2.5
				2020				
Casselton	142	153	0–15	19	18	360	7.5	4.8
			15–61	18	7	279	7.8	4.5
Fargo	133	149	0–15	22	18	489	7.7	5.4
			15–61	26	6	353	8.0	4.0
Prosper	143	153	0–15	21	30	269	7.2	4.5
			15–61	24	17	216	7.4	3.2

[1] DOY, day of year; 15 May is the 135 DOY.

After planting, established plants were counted at the V2 (two trifoliolate stages [30]) by counting a 0.91 m length from the middle soybean rows. During the growing season, soil cover percent (Canopeo, Oklahoma State University, Stillwater, OK, USA) was recorded. Fractional green canopy cover photos were processed providing canopy coverage percentage [18]. Canopy pictures were taken approximately 1.5 m from the soil surface in the center of each plot using an iPad (Apple, Cupertino, CA, USA). MATLAB software (MathWorks, Inc., Natick, MA, USA) was used to calculate canopy cover by FGCC. Canopy photosynthetically active radiation (PAR) interception measurements were collected randomly in the front, middle, and back third of each experimental unit using an Accupar LP-80 (METER Group Inc, Pullman, WA, USA) with the sensor perpendicular to the plot at a height of 2 cm above the soil surface with the above canopy PAR measured at 1.5 m. The above and below canopy PAR was averaged for each experimental unit. The PAR interception was calculated by dividing the above canopy PAR by below canopy PAR, subtracting that value from one, and multiplying by 100. The NDVI was recorded at a height of 0.5 m above the canopy using a RapidSCAN CS-45 (Holland Scientific, Lincoln, NE, USA) with NDVI being averaged across the experimental unit. The standard NDVI was calculated using the 670 nm (red light) and 780 nm (near-infrared) wavelengths. Fractional green canopy coverage, Accupar, and RapidSCAN measurements were recorded when the soybean plants in the early planting were at the V2, V4, R1, R3, R5, or R7 (two trifoliolate, four trifoliolate, beginning of flowering, beginning of pod formation, beginning

of bean development, and pod and leaf yellowing, respectively) growth stage for a total of 18 different canopy measurements per experimental unit. The late-planted soybean samples were only slightly behind in the growth stage, and data were averaged across planting dates as soybean growth stages in a production field also vary slightly. Soybean seed yield was harvested after physiological maturity when the seed was harvestable using a Wintersteiger Classic plot combine (Wintersteiger Ag, Ried im Innkreis, Austria) and corrected to 13% moisture content.

Fractional green canopy cover, PAR interception, and NDVI measurements at each growth stage and the established plant density recorded at all environments were used in multiple linear regression analysis, for a total of 6688 data points. Measurements greater than 3 standard deviations of the mean in each environment for each measurement type and stage combination were removed from the data (165 data points). The multiple regression approach to predicting yield using stepwise and lasso regression was similar to Kumar et al. [31]. The Reg procedure in SAS (SAS Institute Inc., Cary, NC, USA) was used to analyze the relationship between each individual measurement and yield. Variable variance inflation factors (VIF) were reviewed to ensure VIF values were below 5 [32]. The Glmselect procedure in SAS was used for stepwise and lasso multiple linear regression methods. Models were compared using the lowest root mean square error (RMSE) and highest adjusted R^2 [31] and lowest Akaike information criterion (AIC) [33]. Stepwise regression, using a p-value entry-level selection criteria of 0.15 [34], was used to build a model to best predict soybean seed yield using the 18 canopy measurement variables for 6523 total data points from 768 total experimental units. The "validate" statement was used to randomly select 20% of the data, and adjusted R^2 was averaged over 50 iterations to validate the model for both regression methods.

3. Results and Discussion

Individual canopy measurements combined across environments moderately indicate that the variation in yield with FGCC and NDVI are better descriptors on average (Table 3). Through the R1 to R3 growth stages, soybean is actively producing more trifoliolates and producing seed, and FGCC was consistently (R^2 from 0.43 to 0.52) related to the yield at these stages. Canopy PAR interception was poorly related to seed yield throughout the season (R^2 from 0.01 to 0.30). The best time to record FGCC and NDVI is at R5. At R5, the PAR interception relationship with yield is considerably lower than at the other reproductive stages. This is likely due to most experimental units having similar PAR interception values regardless of the yield potential of the unit. Narrow row spacing (30.5 cm) typically improves PAR interception capacity and yield potential, compared to wide rows for soybean [28].

Table 3. Coefficients of determination (R^2) of soybean seed yield in relation to green canopy cover, light interception, and a vegetative index sampled at various growth stages for Casselton, Prosper, and Fargo, North Dakota, USA, in 2019 and 2020.

	FGCC [1]		PAR		NDVI	
Stage [2]	R^2 [3]	RMSE	R^2	RMSE	R^2	RMSE
V2	0.05	710	0.01	728	0.01	725
V4	0.21	646	0.21	647	0.19	653
R1	0.43	551	0.24	635	0.41	560
R3	0.49	519	0.30	608	0.05	708
R5	0.52	507	0.01	724	0.65	434
R7	0.16	668	0.23	637	0.01	728

[1] FGCC, fractional green canopy cover; PAR, photosynthetically active radiation interception; NDVI, normalized difference vegetative index. [2] V2, two trifoliolate; V4, four trifoliolate; R1, beginning of flowering; R3, beginning of pod formation; R5, beginning of bean development; R7, pod, and leaf yellowing. [3] coefficients of determination and root mean square error (RMSE) derived from SAS Proc Reg.

The R5 NDVI measurement was the best single observation (R^2 = 0.65) explaining yield differences. The relationship between NDVI and yield was expected to increase from planting until R6. The poor relationship between NDVI and yield (R^2 = 0.05) at R3 may have been due to experimental units absorbing comparable amounts of visible light at that stage. Ma et al. [8] found soybean NDVI and seed yield relationships improved from the R2 (full flowering) to R5 stages, which can discern between high- and low-yielding genotypes when measured at R5. The NDVI is strongly related to above soybean ground biomass [35], and soybean seed production potential increases as plant growth increases [36]. Board [37] found that total dry mass at R5, plant height, and length of the seed-filling period was highly correlated with seed yield (R^2 = 0.86). Christenson et al. [38] found no differences between canopy reflectance and yield estimation across growth stages although maturity was more accurately predicted during the seed-filling period. Therefore, the R5 NDVI results, which best describe yield in this study, are similar to those found by Ma et al. and Hoyos-Villegas and Fritschi [8,39]. However, genetic differences have a high impact on vegetative indices and vary depending on development and growing conditions [40,41].

The stepwise and Lasso regression model parameters were comparable with stepwise having a slight advantage with lesser deviation from the regression line (Table 4). The primary differences between the two methods are the variables used in the models with Lasso variable selection typically, minimizing overfitting of models, compared to stepwise regression [42]. Within the two models, the importance of NDVI at R1, R3, and R5 and FGCC at R3 are similar (Table 4). In this case, the variable combination produced by the stepwise (Adj. R^2 = 0.69) model is similar to the Lasso (Adj. R^2 = 0.67) model. These results are similar to linear multiple regression for soybean yield prediction using yield components (R^2 = 0.70) [43]. However, the Lasso variable selection provides a more practical use as only NDVI and FGCC measurements are necessary with a relatively negligible Adj. R^2 reduction. Previous soybean canopy measurements studies relate yield to canopy reflectance [8,9] and PAR interception [28]. However, incorporating NDVI, FGCC, and PAR interception across environments with early and late planting date, early and late relative maturity, 408,000 and 457,000 germinable seeds ha^{-1} seeding rate, and narrow (30.5 cm) and wide (61 cm) row spacing allows for a yield prediction model with a greater inference and, in this case, a greater explanation of yield, compared to a single canopy measurement.

Table 4. Stepwise and Lasso regression parameter comparison for soybean canopy measurements combined across 12 environments.

Parameter [1]	Stepwise Regression	Lasso Regression
Adj. R^2	0.68	0.66
Validated Adj. R^2	0.69	0.67
RMSE	411	425
AIC	3346	3362
Variables Used [2]	NDVI.R1 PAR.R1 NDVI.R3 FGCC.R3 NDVI.R5 PAR.R5	NDVI.R1 NDVI.R3 FGCC.R3 NDVI.R5 FGCC.R5

[1] Adj. R^2, adjusted R^2; RMSE, root mean square error; AIC, Akaike information criterion; FGCC, fractional green canopy cover; PAR, photosynthetically active radiation interception; NDVI, normalized difference vegetative index. [2] R1, beginning of flowering; R3, beginning of pod formation; R5, beginning of bean development.

Understanding the practical use of these models can provide researchers different estimates depending on which equation is used. For Example, Ma et al. [8] suggest measuring NDVI between R4 and R5 stages to screen and rank soybean genotypes. Although the models were validated, we demonstrate the applied usage of the models from Table 5 using sensor data from an experimental unit using the recommended farming practices

from the north-central USA region [44]. Using the stepwise model in Table 5 with the same measurement data from the same experimental unit with an actual yield of 3723 kg ha^{-1} and measurements of 0.72, 57.8, 0.82, 92.8, 0.84, and 85.7 for NDVI at R1, PAR at R1, NDVI at R3, FGCC at R3, NDVI at R5, and PAR at R5, respectively, the estimated yield is 3711 kg ha^{-1}. Using the Lasso model in Table 5 and values of 0.72, 0.82, 92.8, 0.84, and 93.3 for NDVI at R1, NDVI at R3, FGCC at R3, NDVI at R5, and FGCC at R5, respectively, the estimated yield is 3554 kg ha^{-1}. The yield predictions display how the stepwise model can provide higher yield values than the Lasso model although Lasso regression can have smaller prediction errors comparatively [31]. The behavior of these models is important to note to better understand the yield prediction from regression equations.

Table 5. Stepwise and Lasso regression equations to best predict soybean yield.

Method	Equation [1]
Stepwise	$\hat{Y} = 874 \times NDVI.R1 - 8 \times PAR.R1 + 1913 \times NDVI.R3 + 9 \times FGCC.R3 + 9357 \times NDVI.R5 - 13 \times PAR.R5 - 5604$
Lasso	$\hat{Y} = 40 \times NDVI.R1 + 562 \times NDVI.R3 + 7 \times FGCC.R3 + 8185 \times NDVI.R5 + 5 \times FGCC.R5 - 4921$

[1] FGCC, fractional green canopy cover expressed in percent; PAR, photosynthetically active radiation interception expressed in percent; NDVI, normalized difference vegetative index expressed as 0 to 1; $\hat{Y}$ is estimated yield kg ha^{-1}; R1, beginning of flowering; R3, beginning of pod formation; R5, beginning of bean development.

Combining established plant density with a canopy measurement could be a simple means to improve yield prediction. However, established plant density was not a strong predictor of yield within the range of plant densities we encountered in these experiments and did not improve R^2 values (Table 6), compared to canopy measurements alone (Table 3), similar to the study by Ma et al. [8]. Our FGCC results are comparable to Yu et al. [45], who reported a correlation coefficient of 0.56 between canopy cover and yield at R3. A single canopy measurement, especially FGCC at R3 or R5, is a more effective use of time to predict yield. Yu et al. [45] reported FGCC explained three times greater variation than the best vegetative index in Illinois, USA. Our results show that NDVI explains more yield variation than FGCC at R5; however, vegetative indices are more sensitive to genetic and environmental conditions and may not provide consistent results year to year [40,45].

Table 6. Coefficients of determination (R^2) of soybean seed yield in relation to established plant density and green canopy cover, light interception, or vegetative index sampled over soybean growth stages for Casselton, Prosper, and Fargo, North Dakota, USA, in 2019 and 2020.

	Established Plant Density					
	FGCC [1]		PAR		NDVI	
Stage [2]	Adj. R^2 [3]	RMSE	Adj. R^2	RMSE	Adj. R^2	RMSE
V2	0.05	711	0.01	729	0.01	726
V4	0.22	643	0.21	647	0.20	654
R1	0.44	548	0.24	636	0.42	557
R3	0.49	519	0.30	609	0.05	710
R5	0.52	508	0.01	725	0.65	435
R7	0.16	669	0.23	638	0.01	729

[1] FGCC, fractional green canopy cover; PAR, photosynthetically active radiation interception; NDVI, normalized difference vegetative index. [2] V2, two trifoliolate; V4, four trifoliolate; R1, beginning of flowering; R3, beginning of pod formation; R5, beginning of bean development; R7, pod and leaf yellowing. [3] Adjusted coefficients of determination (Adj. R^2) and RMSE from SAS Proc Reg.

The model that best predicts yield may not always be the most practical to use. Instruments for NDVI and PAR measurement can be expensive and may not be practical for every soybean producer or agronomist to own. However, FGCC can easily be obtained by using a cell phone with the Canopeo application in a few seconds. Table 3 provides R^2 values for the relationship between FGCC and yield at several growth stages, and Table 7 displays simple and multiple regression equations, which were most predictive

of yield. The relationship between FGCC and yield slightly improves when both the R3 and R5 measurements are included. The best yield prediction model using FGCC includes observations at V2, R1, R3, and R5 with the R5 observation having the greatest effect on the seed yield. It is important to note that the FGCC only explains at most 56% of the variation in yield encountered. Using a single FGCC measurement at the R5 growth stage is likely the most efficient way to collect data that give a reasonable prediction. However, given the only moderate R^2 values, FGCC should primarily be used to monitor soybean progress throughout the season rather than predicting yield per se.

Table 7. Regression parameters using fractional green canopy cover measured at various growth stages throughout the growing season to predict yield at Casselton, Prosper, and Fargo, North Dakota, USA, in 2019 and 2020.

FGCC [1]	Adj. R^2	RMSE	Equation
Growth Stage [2]			
R3	0.49	510	$\hat{Y} = 33.4 \times FGCC.R3 + 662.3$
R5	0.52	510	$\hat{Y} = 50.3 \times FGCC.R5 - 868.2$
R3 R5	0.54	479	$\hat{Y} = 18.8 \times FGCC.R3 + 29.4 \times FGCC.R5 - 603.2$
V2 R1 R3 R5	0.56	470	$\hat{Y} = -7 \times FGCC.V2 + 7 \times FGCC.R1 + 13 \times FGCC.R3 + 25 \times FGCC.R5 - 132$ [3]

[1] FGCC fractional green canopy cover, Adj. R^2, adjusted R^2; RMSE, root mean square error. [2] V2, two trifoliolate; R1, beginning of flowering; R3, beginning of pod formation, R5, beginning of bean development. [3] Equation derived using stepwise regression. $\hat{Y}$, yield in kg ha^{-1}; FGGC is expressed as percent cover.

Using aerial sensors to derive similar measurements would likely be less time consuming, less laborious, and more precise than using handheld sensors and, in turn, could improve the models. Our results fall within expected values of other soybean prediction modeling methods including spectral and canopy measurements range from 56 to 85% yield explanation [45–47]. Improving the soybean prediction model would benefit from a wider array of genotypes and additional years of data to expand inferences beyond the two-week planting window, relative maturities groups, seeding rates, and row spacings used in the study. Hyperspectral data and analyses similar to Hong et al. [48] and Yao et al. [49] for future soybean prediction studies may improve upon the results from this study.

4. Conclusions

Canopy cover data have been widely used in agricultural research to estimate or predict both biomass and seed yield. Results from this study suggest that a single canopy measurement prior to the soybean reproductive phase does not provide a high level of seed yield estimation. The multiple linear regression techniques used in this study suggest that most of the soybean seed yield can be explained by canopy measurements taken throughout the growing season, whereas a single NDVI measurement at the R5 stage is a single observation, which most closely predicts yield. Predicting yield at the R5 stage allows soybean producers to estimate production and provides a marketing advantage. Not all soybean producers or consultants have access to reflectance or light quantification instruments. Therefore, measuring the amount of green canopy cover at R5 is an easily accessible and reasonable measurement provided by a free application, explaining half of the yield variation. Further research is needed to evaluate if other measurements or using other technologies such as using aerial sensors that can improve the prediction of soybean yield.

Author Contributions: Literature review, conceptualization, methodology, research, formal. Statistical analysis, writing—original draft preparation, and writing—review and editing, P.K.S.; funding acquisition, project administration, conceptualization, investigation, supervision, and writing—review and editing, H.J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by the North Dakota Soybean Council and the North Dakota University Experiment Station.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Upon request from authors.

Acknowledgments: The authors thank Chad Deplazes for assisting with the management of the research areas and summer help for their assistance in the project.

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

References

1. NASS-USDA. Crop Production. 2020. Available online: https://www.nass.usda.gov/Quick_Stats/Lite/index.php (accessed on 16 August 2021).
2. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. An Augmented Linear Mixing Model to Address Spectral Variability for Hyperspectral Unmixing. *IEEE Trans. Image Process.* **2019**, *28*, 1923–1938. [CrossRef] [PubMed]
3. Rouse, J.W.; Haas, R.H.; Schnell, J.; Deering, D.W. Monitoring vegetation systems in the great plains with ERTS. *NASA Spec. Publ.* **1974**, *351*, 309–317.
4. Estep, L.; Terrie, G.; Davis, B. Crop stress detection using AVIRIS hyperspectral imagery and artificial neural networks. *Int. J. Remote Sens.* **2004**, *25*, 4999–5004. [CrossRef]
5. Thapa, S.; Rudd, J.C.; Xue, Q.; Bhandari, M.; Reddy, S.K.; Jessup, K.E.; Liu, S.; Devkota, R.N.; Baker, J.; Baker, S. Use of NDVI for characterizing winter wheat response to water stress in a semi-arid environment. *J. Crop. Improv.* **2019**, *33*, 633–648. [CrossRef]
6. Stoms, D.M.; Hargrove, W.W. Potential NDVI as a baseline for monitoring ecosystem functioning. *Int. J. Remote Sens.* **2000**, *21*, 401–407. [CrossRef]
7. Xu, C.; Katchova, A.L. Predicting Soybean Yield with NDVI Using a Flexible Fourier Transform Model. *J. Agric. Appl. Econ.* **2019**, *51*, 402–416. [CrossRef]
8. Ma, B.L.; Dwyer, L.M.; Costa, C.; Cober, E.R.; Morrison, M.J. Early prediction of soybean yield from canopy reflectance measurements. *Agron. J.* **2001**, *93*, 1227–1234. [CrossRef]
9. Mourtzinis, S.; Rowntree, S.C.; Suhre, J.J.; Weidenbenner, N.H.; Wilson, E.W.; Davis, V.M.; Naeve, S.L. The use of reflectance data for in-season soybean yield prediction. *Agron. J.* **2014**, *106*, 115–1168. [CrossRef]
10. Vannoppen, A.; Gobin, A.; Kotova, L.; Top, S.; De Cruz, L.; Viksna, A.; Aniskevich, S.; Bobylev, L.; Buntemeyer, L.; Caluwaerts, S.; et al. Wheat yield estimation from NDVI and regional climate models in Latvia. *Remote Sens.* **2020**, *12*, 2206. [CrossRef]
11. Teal, R.K.; Tubana, B.; Girma, K.; Freeman, K.W.; Arnall, D.B.; Walsh, O.; Raun, W.R. In-Season Prediction of Corn Grain Yield Potential Using Normalized Difference Vegetation Index. *Agron. J.* **2006**, *98*, 1488–1494. [CrossRef]
12. Genovese, G.; Vignolles, C.; Nègre, T.; Passera, G. A methodology for a combined use of normalised difference vegetation index and CORINE land cover data for crop yield monitoring and forecasting. A case study on Spain. *Agronomie* **2001**, *21*, 91–111. [CrossRef]
13. Harrell, D.L.; Tubaña, B.S.; Walker, T.W.; Phillips, S.B. Estimating Rice Grain Yield Potential Using Normalized Difference Vegetation Index; Estimating Rice Grain Yield Potential Using Normalized Difference Vegetation Index. *Agron. J.* **2011**, *103*, 1717–1723. [CrossRef]
14. Vannoppen, A.; Gobin, A. Estimating Farm Wheat Yields from NDVI and Meteorological Data. *Agronomy* **2021**, *11*, 946. [CrossRef]
15. Lee, C.D. Reducing Row Widths to Increase Yield: Why It Does Not Always Work. *Crop. Manag.* **2006**, *5*, 1–7. [CrossRef]
16. Wells, R. Soybean Growth Response to Plant Density: Relationships among Canopy Photosynthesis, Leaf Area, and Light Interception. *Crop. Sci.* **1991**, *31*, 755–761. [CrossRef]
17. Egli, D.B. Mechanisms responsible for soybean yield response to equidistant planting patterns. *Agron. J.* **1994**, *86*, 1046–1049. [CrossRef]
18. Patrignani, A.; Ochsner, T.E. Canopeo: A powerful new tool for measuring fractional green canopy cover. *Agron. J.* **2015**, *107*, 2312–2320. [CrossRef]
19. Wilhelm, W.W.; Ruwe, K.; Schlemmer, M.R. Comparison of three leaf area index meters in a corn canopy. *Crop. Sci.* **2000**, *40*, 1179–1183. [CrossRef]
20. Perry, E.M.; Fitzgerald, G.J.; Poole, N.; Craig, S.; Whitlock, A. Ndvi from Active Optical Sensors as a Measure of Canopy Cover and Biomass. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2012**, 317–319. [CrossRef]
21. Goodwin, A.W.; Lindsey, L.E.; Harrison, S.K.; Paul, P.A. Estimating wheat yield with normalized difference vegetation index and fractional green canopy cover. *Crop. Forage Turfgrass Manag.* **2018**, *4*, 1–6. [CrossRef]
22. Singer, J.W. Soybean light interception and yield response to row spacing and biomass removal. *Crop. Sci.* **2001**, *41*, 424–429. [CrossRef]
23. Gardner, F.P.; Auma, E.O. Canopy structure, light interception, and yield and market quality of peanut genotypes as influenced by planting pattern and planting date. *F. Crop. Res.* **1989**, *20*, 13–29. [CrossRef]
24. Chang, K.W.; Shen, Y.; Lo, J.C. Predicting rice yield using canopy reflectance measured at booting stage. *Agron. J.* **2005**, *97*, 872–878. [CrossRef]

25. Schmitz, P.K.; Stanley, J.D.; Kandel, H.J. Row Spacing and Seeding Rate Effect on Soybean Seed Yield in North Dakota. *Crop. Forage Turfgrass Manag.* **2020**, *6*, e20010. [CrossRef]
26. Stanley, J.D. Yield-Limiting Factors in North Dakota Soybean Fields. Master's Thesis, North Dakota State University, Fargo, ND, USA, 2017.
27. Mourtzinis, S.; Conley, S.P. Delineating soybean maturity groups across the US. *Agron. J.* **2017**, *109*, 1397–1403. [CrossRef]
28. Andrade, F.H.; Calviño, P.; Cirilo, A.; Barbieri, P. Yield Responses to Narrow Rows Depend on Increased Radiation Interception. *Agron. J.* **2002**, *94*, 975–980. [CrossRef]
29. Kandel, H.; Endres, G. *Soybean Production Field Guide for North Dakota*; A1172 (revised); North Dakota State University: Fargo, ND, USA, 2019.
30. Fehr, W.R.; Caviness, C.E.; Burmood, D.T.; Pennington, J.S. Stage of Development Descriptions for Soybeans, Glycine Max (L.) Merrill1. *Crop. Sci.* **1971**, *11*, 929–931. [CrossRef]
31. Kumar, S.; Attri, S.D.; Singh, K.K. Comparison of lasso and stepwise regression technique for wheat yield prediction. *J. Agrometeorol.* **2019**, *21*, 188–192.
32. Burnham, K.P.; Anderson, D.R. Information and likelihood theory: A basis for model selection and inference. In *e: A Practical Information-Theoretic Approach*; Springer: New York, NY, USA, 2004.
33. Lollato, R.P.; Diaz, D.A.R.; DeWolf, E.; Knapp, M.; Peterson, D.E.; Fritz, A.K. Agronomic practics for reducing wheat yield gaps: A quantitative appraisal of progressive producers. *Crop. Sci.* **2019**, *59*, 333–350. [CrossRef]
34. Derksen, S.; Keselman, H.J. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *Br. J. Math. Stat. Psychol.* **1992**, *45*, 265–282. [CrossRef]
35. Ma, B.L.; Morrison, M.J.; Dwyer, L.M. Canopy light reflectance and field greenness to assess nitrogen fertilization and yield of maize. *Agron. J.* **1996**, *88*, 915–920. [CrossRef]
36. Vega, C.R.; Andrade, F.H.; Sadras, V.O.; Uhart, S.A.; Valentinuz, O.R. Seed number as a function of growth. A comparative study in soybean, sunflower, and maize. *Crop. Sci.* **2001**, *41*, 748–754. [CrossRef]
37. Board, J. Light interception efficiency and light quality affect yield compensation of soybean at low plant populations. *Crop. Sci.* **2000**, *40*, 1285–1294. [CrossRef]
38. Christenson, B.S.; Schapaugh, W.T.; An, N.; Price, K.P.; Prasad, V.; Fritz, A.K. Predicting soybean relative maturity and seed yield using canopy reflectance. *Crop. Sci.* **2016**, *56*, 625–643. [CrossRef]
39. Hoyos-Villegas, V.; Fritschi, F.B. Relationships among vegetation indices derived from aerial photographs and soybean growth and yield. *Crop. Sci.* **2013**, *53*, 2631–2642. [CrossRef]
40. Aparicio, N.; Villegas, D.; Casadesus, J.; Araus, J.L.; Royo, C. Spectral vegetation indices as nondestructive tools for determining durum wheat yield. *Agron. J.* **2000**, *92*, 83–91. [CrossRef]
41. Zaman-Allah, M.; Vergara, O.; Araus, J.L.; Tarekegne, A.; Magorokosho, C.; Zarco-Tejada, P.J.; Hornero, A.; Albà, A.H.; Das, B.; Craufurd, P.; et al. Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods* **2015**, *11*, 35. [CrossRef] [PubMed]
42. Tibshirani, R.J. The LASSO method for variable selection in the cox model. *Stat. Med.* **1997**, *16*, 385–395. [CrossRef]
43. Wei, M.C.F.; Molin, J.P. Soybean yield estimation and its components: A linear regression approach. *Agriculture* **2020**, *10*, 384. [CrossRef]
44. Schmitz, P.K.; Kandel, H.J. Individual and Combined Effects of Planting Date, Seeding Rate, Relative Maturity, and Row Spacing on Soybean Yield. *Agronomy* **2021**, *11*, 605. [CrossRef]
45. Yu, N.; Li, L.; Schmitz, N.; Tian, L.F.; Greenberg, J.A.; Diers, B.W. Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. *Remote Sens. Environ.* **2016**, *187*, 91–101. [CrossRef]
46. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Hartling, S.; Esposito, F.; Fritschi, F.B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* **2020**, *237*, 111599. [CrossRef]
47. Zhang, X.; Zhao, J.; Yang, G.; Liu, J.; Cao, J.; Li, C.; Zhao, X.; Gai, J. Establishment of plot-yield prediction models in soybean breeding programs using UAV-based hyperspectral remote sensing. *Remote Sens.* **2019**, *11*, 2752. [CrossRef]
48. Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Plaza, A.; Chanussot, J. Graph Convolutional Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5966–5978. [CrossRef]
49. Yao, J.; Meng, D.; Zhao, Q.; Cao, W.; Xu, Z. Nonconvex-Sparsity and Nonlocal-Smoothness-Based Blind Hyperspectral Unmixing. *IEEE Trans. Image Process.* **2019**, *28*, 2991–3006. [CrossRef] [PubMed]