Assessing on-farm cover crop biomass accumulation using a mixed modeling approach

Kushal KC ¹, Nora M. Bello ², Steve W. Culman ³, and Sami Khanal ¹,*

- ¹ Department of Food, Agricultural, and Biological Engineering, The Ohio State University, Columbus, OH 43210, USA; kc.7@buckeyemail.osu.edu
- ² U.S. Department of Agriculture, Agricultural Research Service; Nora.Bello@usda.gov
- ³ Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA; steven.culman@wsu.edu

*Correspondence: khanal.3@osu.edu

A paper from the Proceedings of the 16th International Conference on Precision Agriculture 21-24 July 2024 Manhattan, Kansas, United States

Abstract.

Improved understanding of cover crop performance in a large geographic area is crucial for fostering widespread adoption of cover crops as a conservation practice. Conventional regression methods fail to analyze on-farm agricultural data spanning various locations and multiple timeframes effectively. This research proposes to employ a linear mixed modeling approach to assess remote sensing (RS), biophysical characteristics and management practices as potential predictors of cereal rye biomass across multiple sites during a growing season. Field measurements, including cereal rye biomass and biophysical crop characteristics were collected alongside multispectral images from an unmanned aerial system (UAS) in 13 fields across Northwest Ohio in the spring of 2021. A baseline linear mixed model was specified to evaluate biomass accumulation across fields as a function of management practices defined by planting timing and methods. Residuals from the fitted model were extracted and subjected to stepwise selection for additional explanatory variables, including 13 vegetation indices (VIs) derived from remotely sensed images and two biophysical characteristics, namely crop height and canopy cover percentage (CCpercent). An extended model expanded the Baseline model by further incorporating selected explanatory variables into the linear predictor, specifically blue green ratio (BGratio), simple ratio rededge (SRrededge) and CCpercent as well as their interactions with management groups. Results showed that the early planted cereal rye using drilling approach consistently had the greatest biomass accumulation throughout the spring, suggesting differences in overall growth induced by different farm management practices. The extended model had better model fit than the baseline model based on Akaike Information Criteria (AIC), AIC corrected (AICC), and Bayesian Information Criteria (BIC) statistics. This was further validated by significant

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 16th International Conference on Precision Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2024). Title of paper. In Proceedings of the 16th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

differences observed in a likelihood ratio test. The total variance in the extended model was reduced by up to 61%, underscoring the importance of RS and biophysical characteristics in enhancing precision and predictability of the model. In summary, the linear mixed model accompanied by RS-derived VI and biophysical characteristics better captured variabilities in onfarm cereal rye biomass across multiple sites along a growing season.

Keywords.

Cover crop, Cereal rye; Cover crop biomass estimation; Unmanned Aerial System (UAS); Linear mixed modeling; Remote Sensing

1. Introduction

In recent years, winter cover crops have been proposed as mitigation strategies for both environmental and agronomic benefits, namely prevention of soil erosion, control of weeds, reduction of soil compaction, and improved soil carbon sequestration (Daryanto et al., 2018; Finney et al., 2016). Indeed, cover crops are increasingly recognized as important tools for improving soil health, fostering climate-resilient farms, and mitigating adverse environmental impacts, as evidenced by the implementation in federal and state-level incentive programs (Duke et al., 2022; Wallander et al., 2021).

However, the current rate of cover crop adoption in Midwest remains relatively low, with only 7% of total croplands planted in cover crops in 2021 (Zhou et al., 2022). Factors limiting the adoption of cover crops include logistical challenges in establishing them and managing their biomass at the end of the season, as well as agronomic concerns such as crop yield penalties in subsequent cash crops due to slow/delayed emergence due to cover crop residues on the ground (Plastina et al., 2020; Burnett et al., 2018; Roesch-Mcnally et al., 2018). Additionally, social and behavioral factors further compound these challenges (CTIC, 2016; Myers & Watts, 2015; Singer, 2008).

Knowledge of factors influencing the cover crop growth and its variability within a field can help farmers in several ways. For instance, understanding cover crop biomass variability allows farmers to: 1) identify areas with higher or lower nutrient content, enabling targeted fertilization and nutrient management for the subsequent cash crops; 2) identify areas prone to weed pressure and plan appropriate weed control measures; 3) determine potential pest and disease dynamics; and 4) manage input costs and increase overall farm profitability by tailoring management practices to the specific needs of different areas within a field (Ruis et al., 2019).

Winter cover crops are typically planted in the fall, either alongside standing cash crops or immediately after cash crops harvest, leaving a narrow time window for germination and establishment. Delayed planting of cover crops can reduce their spring biomass production. Furthermore, the choice of planting method, such as aerial broadcasting or drilling, influences seed-to-soil contact and subsequently impacts overall biomass growth. Previous studies have investigated the effects of management practices on cover crop growth through controlled plot-scale experiments (Balkcom et al., 2023; Boyd et al., 2009; Haramoto, 2019; St Aime et al., 2022). However, the relevance of these findings to larger areas under commercial production remains unclear, largely due to variable soil and topographic conditions (Duiker, 2014; Hayden et al., 2015; Moore & Mirsky, 2020). On-farm data poses unique challenges for collection due to the cumbersome and costly nature of the conventional manual approach to collecting ground-truth data. Similarly, analysis of on-farm data requires simultaneous consideration of multiple sources of variability (e.g. between- and within-field variability).

With continuous development and advancement in sensor technologies, remote sensing (RS), particularly unmanned aerial system (UAS), is becoming a cost-effective approach for collecting field-level data of high resolution that could help in understanding within-field variability in crop growth. Specifically, vegetation indices (VIs), computed from RS-based images, have been used as proxies of crop growth (Avneri et al., 2023; Rosle et al., 2019; Teshome et al., 2023; Zhang et al., 2021). Previous studies used linear and non-linear regression-type approaches to model cover crop biomass and its nitrogen (N) content as a function of direct and indirect traits of cover crop growth, including VIs and biophysical measurements (Brennan & Smith, 2023; Marcillo et al., 2020; Moore & Mirsky, 2020; Muñoz et al., 2010; Murrell et al., 2017; Prabhakara et al., 2015; Roth & Streit, 2018). For instance, Prabhakara et al. (2015) explored the linear relationship between winter cover crop biomass and RS image-derived VIs, growing degree days, and percent ground cover and showed correlations of up to 0.84 between 10 VIs and cover crop biomass. Another study by Marcillo et al. (2020) compared least absolute shrinkage and selection operator (LASSO), Ridge, and random forest (RF) for predicting both cereal rye biomass and its N content as a function of RS image derived features.

Despite wide use of conventional linear regression and machine learning methods in agricultural research, they have limitations when used for the analysis of on-farm data. Most prominent amongst said limitations is the inability of traditional regression and machine learning to accommodate hierarchical data structure (e.g. multiple observations across times from multiple fields across sites) and ensuing correlation patterns that often emerge from temporal and spatial dependencies among data (Yang, 2010), as well as heteroskedasticity. Linear mixed models can seamlessly incorporate hierarchical data structure and multiple sources of variability in a dataset through specification of fixed and random effects for systematic and non-systematic sources of variability, respectively. While fixed effects represent the factor of interest or explanatory covariates on the response variable, random effects incorporate hierarchical levels of data architecture reflective of the design structure or data generation process. Through this, linear mixed models allow us to naturally delineate different levels of data organization and corresponding inferential scope (Milliken & Johnson, 2009; Stroup, 2013).

The overarching goal of this study was to assess the differences in spring growth of winter cereal rye between management practices defined by planting timing and method in large production areas. To this end, we will utilize a linear mixed modeling approach that accommodates the hierarchical structure of the data. Enhancements of model fit to data will be further explored by considering direct (i.e. biophysical parameters) and indirect (i.e. RS-based) measurements as explanatory variables.

2. Materials and Methods

2.1 Study Sites and Sampling Protocol

The study focused on 13 fields located in northwest Ohio during the 2020 to 2021 cover crop season (Figure 1). Fields were planted with cereal rye between September 30 and November 17, following soybean harvest. Planting methods for cereal rye seeds included drilling, aerial, and broadcasting in 9, 2, and 2 fields, respectively. Fields planted during September to October 15 were grouped as early planting while the fields planted after October 16 were grouped as late planting. Planting methods and timing were grouped together to generalize the diverse management practices across each field.

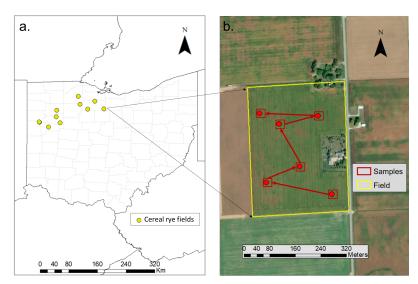


Figure 1: Study area reflecting (a) locations of cereal rye fields across northwestern Ohio, and (b) randomly selected six sampling locations within one of the fields.

In each field, 2-3 sub-field zones were first identified based on the dominance of various soil types by referencing the soil map from the soil survey geographic database (SSURGO) (USDA NRCS, 2022). A total of six sampling locations were randomly selected from the sub-field zones, with each zone containing 1-3 locations, following a zig-zag pattern as recommended by the OSU e-

fields data collection protocol (OSUE, 2021) to ensure representativeness of the area.

2.2 Field Data Collection

Composite samples of cereal rye biomass were collected in the spring of 2021 between March and May using a $(0.50)^2$ m² quadrat (Figure 2). At each of the six locations within a field, samples were repeatedly collected for three times (T1, T2, and T3) during the cereal rye growing season, where T1 is early March, T2 is late March to early April, and T3 is late April to early May.



Figure 2: Data collection process in each field reflecting (a) 0.50 * 0.50 m2 quadrant used for biomass data collection, (b) crop height measurement using a tape, (c) measurement of canopy cover percentage using canopeo app in a smartphone, and (d) destructive sampling of cereal rye within a quadrat.

Collected biomass samples were subsequently oven-dried at 55°C and weighed for dry biomass content; we hereafter refer to this as cereal rye biomass. In addition, crop height and canopy cover percentage (CCpercent) were also collected from sampling locations in each field. Average crop height in a quadrat was manually measured using a meter-scale tape (Figure 2b). CCpercent was measured using a mobile app-Canopeo (Patrignani & Ochsner, 2015). Canopeo uses color values in the red-green-blue (RGB) channel to first classify all the image pixels into green and non-green and then calculates the fraction of green pixels to produce a binary image showing the green cover percentage.

2.3 Remote Sensing Variables

Alongside field data collection, DJI Phantom 4 multispectral UAS was flown at approximately 90 m of height over the field to capture data in five multispectral bands (blue (B), green (G), red (R), red edge (RE), near-infrared (NIR)) at a ground sampling distance (GSD) of around 4-6 cm per pixel. The images were captured in a lawnmower fashion, maintaining an 80% front overlap and a 70% side overlap. The UAS was equipped with a real-time kinematics (RTK) module, ensuring centimeter-level precision in image positioning. Red markers were strategically placed at sampling locations to precisely locate sampled biomass in the images. Raw multispectral images collected by UAS were processed using Pix4D, an image processing software, which employs structure-from-motion (SFM) techniques (Pix4D, 2022) to stitch hundreds of raw images together to generate a single orthomosaic per field. Post-processing steps included identifying the sampling locations in the orthomosiac image, drawing an area corresponding to the size of the quadrat, and computing average pixel values across five bands – R, G, B, RE, and NIR within the quadrat, representing the spectral properties of collected biomass samples.

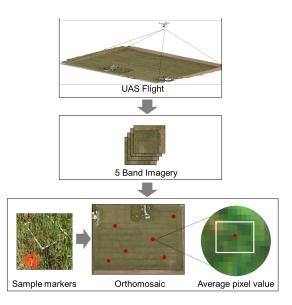


Figure 3: UAS image data collection, processing, and data extraction workflow.

Based on the average pixel values at each sampling location for each of the five spectral bands, 13 VIs were computed to relate with biomass. The VIs included blue-green ratio (BGratio), redgreen ratio (RGratio), normalized difference vegetation index (NDVI), green normalized difference vegetation index (GNDVI), simple ratio (SR), enhanced vegetation index (EVI), normalized green red difference index (NGRDI), excess green (ExG), triangular vegetation index (TVI), soil adjusted vegetation index (SAVI), normalized difference red edge Index (NDRE), simple ratio red edge (SRrededge), and red-edge triangular vegetation index (RTVI) (Table 1).

Table 1: Vegetation indices (VIs) computed based on the combinations of five multispectral bands i.e. red (R), blue (B),

green (G), red edge (RE), and near-infrared (NIR).

Indices	Equation	Source		
RGratio	$\frac{R}{G}$	(Gamon & Surfus, 1999)		
BGratio	$\frac{B}{G}$	(Sellaro et al., 2010)		
NDVI	$\frac{(NIR - R)}{(NIR + R)}$	(Rouse et al., 1974)		
GNDVI	$\frac{(NIR - G)}{(NIR + G)}$	(Moges et al., 2005)		
SR	$\frac{NIR}{R}$	(Jordan, 1969)		
EVI	$\frac{2.5(NIR - R)}{(NIR + 6xR - 7.5xB + 1)}$	(Huete et al., 2002)		
NGRDI	$\frac{(G-R)}{(G+R)}$	(Tucker & Sellers, 1986)		
ExG	(2xG - R + B)	(Woebbecke et al., 1995)		

TVI	0.5 * [120 * (NIR - G) - 200 * (R - G)]	(Broge & Leblanc, 2001)	
SAVI	(1+0.5)*(NIR-R)/(NIR-G)	(Huete, 1988)	
NDRE	$\frac{(NIR - RE)}{(NIR + RE)}$	(Gitelson & Merzlyak, 1994)	
SRrededge	$\frac{NIR}{RE}$	(Cao et al., 2016)	
RTVI	100 * (NIR - RE) - 10 * (NIR - G)	(Chen et al., 2010)	

2.4 Remote Sensing Variables

2.4.1 Preliminary Data Descriptive and Preprocessing

The data used in this study consisted of a total of 228 samples collected in 13 different fields across three distinct periods. A list of potential explanatory variables for biomass included 13 RS-based VIs, derived from UAS images, and two biophysical characteristics - CCpercent and crop height. Due to the skewed distribution of cereal rye biomass, a natural logarithm transformation was applied to normalize the data and mitigate skewness in the response during modeling. Scatterplots were used for preliminary screening of the relationship between the explanatory variables and cereal rye biomass in the transformed scale. Log transformation of explanatory variables including SR, EVI, TVI, and RTVI was implemented to facilitate linear relationship for modeling purposes.

2.4.2 Specification of Statistical Models

To analyze the data, two linear mixed models, referred to as Baseline and Extended, each incorporating varying combinations of variables, were developed in three subsequent steps. Both models were fitted to the log-transformed cereal rye biomass.

Step 1: Baseline Model

Model I was developed to reflect the data collection process across multiple fields and periods. The model consists of fixed effects of (1) groups representing four combinations of planting timing and planting method of cereal rye in fall, (2) time for biomass data collection in spring, and (3) their two-way interaction as linear predictors. Meanwhile, random effects include (1) fields nested within the group, (2) fields nested within the group and crossed with time of biomass data collection to identify repeated measures within each field, and (3) random residual to identify subsampling at the data measurement level. The model was specified as:

$$log (Y_{ijkl}) = \eta + \alpha_j + F_{i(j)} + \tau_k + \alpha \tau_{jk} + F \tau_{i(j)k} + e_{ijkl}$$
 where $F_{i(j)} \sim NIID(0, \sigma_F^2) \perp F \tau_{i(j)k} \sim NIID(0, \sigma_F^2) \perp e_{ijkl} \sim NIID(0, \sigma_e^2)$

 $i=1,\ ...,\ 13\ j=1,\ ...,4\ k=1,\ ...,\ 3\ l=1,\ ...,\ 6.$ NIID represents Normally, Identically, and Independently Distributed.

 Y_{ijkl} = Observed cereal rye biomass of the I_{th} cereal rye sample from the i_{th} field at the k_{th} time assigned to the j_{th} management group.

 η = Intercept for the response variable.

 α_i = Differential effect of the i_{th} management group.

 $F_{i(i)}$ = Differential effect of the i_{th} field assigned to (nested within) j_{th} management group.

 τ_k = Differential effect of the k_{th} timepoint.

 αT_{ik} = Differential effect of the i_{th} management group at the k_{th} timepoint.

 $F_{T_{i(j)k}}$ = Differential effect of the i_{th} field nested within the j_{th} management group at timepoint k.

 e_{ijkl} = Leftover residual noise for the l_{th} sample from the i_{th} field assigned to the j_{th} management group and measured at the k_{th} time.

The random effect factors consist of three components, reflecting sources of random variability in data, namely σ^2_F (i.e. between-field variance), σ^2_{FT} (i.e. variance between multiple time points measured within a field), and σ^2_e (i.e. within-field variance or residual). Total variance (σ^2_T) is given as sum of three variance components (Equation 1).

$$\sigma_{\rm T}^2 = \sigma_{\rm F}^2 + \sigma_{\rm F\tau}^2 + \sigma_{\rm \tau}^2 \tag{1}$$

Step 2: Selection of additional explanatory variables

After fitting the Baseline Model, residuals were computed as the difference between observed and predicted values, $e_{ijkl} = Y_{ijkl} - \hat{Y}_{ijkl}$, which reflect leftover noise in the data after accounting for differences between the management group and the sampling scheme (i.e. multiple samples collected from each field at each time point) throughout the growing season. Residuals were then subjected to model selection using a stepwise approach to identify additional suitable explanatory variables that might help explain leftover noise when incorporated into the linear predictor. The potential explanatory variables included 13 VIs, crop height, and CCpercent. Based on this process, CCpercent, BGratio, and SRrededge were selected from 15 candidates for inclusion in the final model.

Step 3: Extended Model

Using three covariates selected from stepwise selection process, the Baseline Model was further extended to include fixed effects of (1) group, (2) time, (3) their two-way interaction, (4) three covariates and (4) their two-way interaction with group. It also included random effects of (1) fields nested within the group, (2) fields nested within the group and crossed with a time of biomass data collection, and (3) random residual.

2.4.3 Model Comparison

For model comparison, the Baseline and Extended Models were fitted using maximum likelihood estimation, and fit statistics, including Akaike Information Criteria (AIC) (Akaike, 1974), AIC corrected (AICC) (Burnham & Anderson, 2002), and Bayesian Information Criteria (BIC) (Schwarz, 1978), of these two models were compared. A likelihood ratio test (LRT) statistic was also used to assess the statistical significance between the Baseline and Extended models.

Both models were expanded to accommodate heterogeneous residual variance across fields, with variance components estimated using the restricted maximum likelihood (REML) method. Kenward Roger's method was used to estimate degrees of freedom and adjust standard errors. Model assumptions were evaluated using externally studentized residuals. Estimated least-square means for cereal rye biomass and the corresponding 95% confidence intervals were then back-transformed to the original data scale. Comparisons between management groups throughout the growing season were based on the Baseline Model using Bonferroni adjustments to prevent Type I error inflation. A Bonferroni-adjusted P-value of less or equal to 0.05 was considered statistically significant, while a P-value between 0.05 and 0.10 was considered marginally significant. The relationship between covariates and cereal rye biomass was analyzed using the Extended Model. All statistical analyses were performed using the GLIMMIX procedure in SAS (Version 9.4; SAS Institute, Cary, NC, USA).

3. Results

3.1 Baseline Model

The baseline model showed no evidence of a two-way interaction between management groups and timing for data collection on cereal rye biomass (P = 0.16). However, there were evidence of the main effects of management groups (P = 0.01) and time of data collection (P < 0.001) on cereal rye biomass. Throughout all data collection periods in spring, the Early-Drilled group had consistently greater cereal rye biomass than the Late-Aerial, Late-Broadcast, and Late-Drilled groups, which showed no evidence for significant differences from each other. In early spring, the Early-Drilled group had significantly more biomass than the Late-Aerial (P = 0.002) and Late-Drilled (P = 0.04) groups. In mid-spring, the significant difference was observed only between the Early-Drilled and Late-Aerial groups (P = 0.02). By late spring, the Early-Drilled had significantly more biomass than the Late-Aerial (P = 0.02) and Late-Broadcast (P = 0.05) groups.

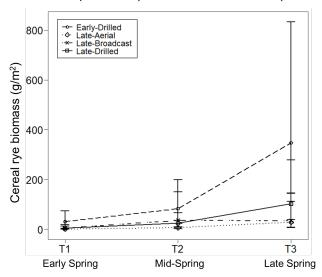


Figure 4: Cereal rye biomass (least square mean estimate ± 95% confidence interval) for four different management groups in spring of 2021.

3.2 Extended Model - Model Enhancement with Remote Sensing and Biophysical **Predictors**

Comparison based on Fit Statistics 3.2.1

Extended

Compared to the Baseline Model, the Extended Model including additional explanatory covariates including RS-derived VIs and biophysical characteristics improved fit to data, which was evidenced by maximum-likelihood-based fit statistics. Specifically, the Extended Model showed a decrease of 62 BIC units, 69 AIC units, and 63 AICC units relative to the Baseline Model, thereby demonstrating that the inclusion of heterogeneous slopes for CCpercent, BGratio, and SRrededge help explain variability in cereal rye biomass. Similarly, the chi-square statistic (i.e., 93.09) based on the likelihood ratio test exceeded the critical value at 95% quantile of the reference distribution with 12 degrees of freedom (i.e., 21.03), indicating that the Extended Model provides a better fit than the Baseline Model.

Table 4: Model fit statistics and likelihood ratio test of the Baseline and Extended models.								
Model	-2Log-likelihood	AIC	AICC	віс	X ² test statistic	χ ² p-value (DF = 12, 95%)		
Baseline	506.22	536.22	538.49	544.70				
	440.40	407.40	475.00	400.00	93.09	21.03		

Note: Chi-square test statistic is computed based on difference of log-likelihood between the models while chi-square at critical value is computed based on the degrees of freedom and 95% significance level at chi-square distribution.

3.2.2 Comparison based on Variance Estimates

The importance of covariates in controlling non-systematic variance was evaluated by comparing residual and total variance between Extended and Baseline Models (Figure 5). Since residual variance varied from one field to another for all 13 fields, total variance was unique to each field. For the Baseline Model, variance estimates showed a between-field variance of 0.57, variance between multiple time points of 0.23, and residual variance ranging from 0.01 to 1.68. For the Extended Model, between-field variance was reduced to 0.23, variance between multiple time points remained at 0.23, and residual variance ranged from 0.007 to 1.5. Similarly, total variance ranged from 0.58 to 2.25 without covariates and from 0.57 to 2.06 with covariates.

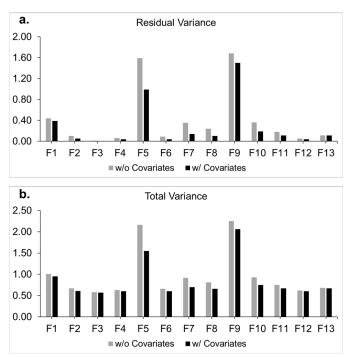


Figure 5: (a) Residual variance and (b) total variance across 13 fields for the models without (w/o) (i.e., baseline model) and with (w/) including the RS-based (BGratio and SRrededge) and biophysical (CCpercent) covariates (i.e. extended model).

Inclusion of covariates in the model resulted in a decrease in both total variance and residual variance across all fields, indicating a reduction in unexplained variance. The highest reduction in total variance and residual variance were 60 and 61%, respectively, for field F5. By incorporating covariates as a systematic component, the Extended Model accounted for some of the nonsystematic variability that was previously attributed to random factors.

4. Discussion

In this study, we focused on assessing the variability in springtime cereal rye growth relative to management practices, based on multitemporal data collected on farmers' fields, in a linear mixed modeling framework. We also evaluated the efficacy of RS-derived and biophysical covariates in explaining the variation present in the data and improving model precision. Of the 13 covariates, CCpercent, BGratio, and SRrededge were selected based on a formal stepwise variable selection process and were found to account unexplained variability in the cereal rye biomass data better than other covariates.

RS-based VIs frequently serves as reliable proxies for crop health, offering insights into environmental stresses like nutrients, heat, and water that may or may not be linked with management practices. For instance, BGratio, which considers crop reflectance in blue and green Proceedings of the 16th International Conference on Precision Agriculture

wavelengths of electromagnetic radiation, can capture variability in crop growth dynamics specifically related to canopy development. In crop canopies, blue light is absorbed by the photosynthetic pigments, notably chlorophyll, facilitating photosynthesis and subsequent carbon assimilation crucial for biomass accumulation. Meanwhile, green light is a strong indicator of the abundance of chlorophyll. These photosynthetic pigments are sensitive to different environmental stressors and hence vary across the crop growth stages. Consistent with prior studies, we found red-edge-based VIs such as SRrededge to be strongly associated with crop biomass (Kanke et al., 2016; M. A. Cho & Atzberger, 2008; Mutanga & Skidmore, 2004).

Consistent with our findings, previous studies have recognized canopy cover percentage as an important variable in explaining variations in crop biomass (Flombaum & Sala, 2007; Goslee, 2020; Montès et al., 2004). However, canopy cover percentage data were collected at a ground level and hence presents logistical challenges when it comes to collecting data from multiple fields. In contrast, SRrededge, which was also found to well explain leftover variability in the biomass data, can be effectively obtained using UAS at high resolution and with broader coverage.

The inclusion of RS-based and biophysical covariates in the linear mixed model provided evidence for improved model performance reflected through AIC, AICC, and BIC metrics compared to the model without these covariates. Based on likelihood ratio test, we found a significant differences between the models with and without covariates. With the inclusion of covariates, there was decrease in total variance due to reduction in leftover or unexplained variation which is key to improving precision of estimates for fixed effects and random effects. This further elucidates the role of RS-based and biophysical characteristics in explaining the non-systematic variation in cereal rye biomass and improving model precision for better predictability and generalizability.

Consistent with prior works (Duiker, 2014; Feyereisen et al., 2006; Staver & Brinsfield, 1998), we observed consistently higher springtime biomass accumulation for cereal rye planted early in the fall using a drilled approach. We also observed differences in cereal rye biomass between Early-Drilled and Late-Drilled groups particularly during early spring, indicating the potential impact of planting time regardless of planting methods. From 2003 to 2005, Duiker (2014) found over 1.5 Mg/ha of cereal rye biomass in early May when planted in early October, dropping below this threshold for mid-October plantings. Similarly, Feyereisen et al. (2006) noted higher cumulative biomass in southwestern Minnesota when cereal rye was sown early (September 15) compared to late October plantings. Haramoto (2019) observed superior establishment with the drilling method (75%) over broadcasting (22%) in a two-year study following corn, though delayed drilling led to reduced growth and nitrogen uptake. The reason for relatively lower biomass accumulation in late planted groups could be due to slowdown in phenological development of cereal rye when the establishment date is delayed from the early part to late fall (Farsad et al., 2011).

Broadcasting and aerial seeding are popular choices among growers due to their economical and logistical advantages. Aerial seeding is an efficient approach of planting cover crops on standing cash crops over a large area. This method is especially valuable in regions where harvesting of summer cash crops is delayed and there is a limited time window for cover crops' planting. However, success in cover crop emergence with aerial seeding depends on adequate soil moisture and late summer/early fall precipitation (Wilson et al., 2013). Broadcasting, similar to aerial seeding but ground based, can result in improved cover crop emergence with tillage and early fall seeding, capitalizing on warmer temperatures and potential rainfall (Brooker et al., 2020). However, without additional tillage, seeding into dry soil risks delayed establishment until sufficient rainfall is received, which improves soil moisture (Hillel, 2003). Based on analyses of precipitation data from PRISM, we observed a dry spell for about 10 days from November 1 to November 10 in three field locations planted late (i.e. late October to early November) with aerial and broadcasting method. This could be the reason the Late-Aerial and Late-Broadcast groups consistently showed lower springtime biomass, particularly during mid- and late spring.

The multi-location multi-temporal agronomic data on winter cover crop collected across fields in

a larger production area inherently are hierarchical in nature. It is thus important to carefully analyze such data to identify the data generation process as well as structure of the data. Most of the prior studies of similar kind often relied either on standard statistical methods or advanced machine learning techniques while assuming mutually independent observations and nonexistent data structure. This undermines the critical statistical assumptions while ignoring the correlation patterns in the data due to hierarchical structure and often leads to false inference and conclusion (Stroup, 2013). In our study, we incorporated a linear mixed modeling framework that possess unique ability to encode multilevel hierarchical data structure into the analysis. We specified fixed effects concerning winter cover crop management groups, time of data collection and various RS-based and biophysical explanatory covariates. Meanwhile, random effects were used to reflect multiple sources of non-systematic variation not explained by the fixed effects. This allowed us to incorporate the data generation process in the modeling while assessing whether differences in cereal rye biomass accumulation are consistent across different field and growing conditions over a larger production region in the northwestern, Ohio.

While this study offers insights on factors useful in assessing variability in cereal rye biomass in on-farm conditions, our study's scope is constrained as it only considers fields previously planted with soybeans, limiting generalizability to fields with different prior crops like corn. Future research should incorporate diverse prior crop data to enhance generalizability. Furthermore, our study treats biomass sampling as a repeated measure across three time periods. However, the destructive nature of sampling may introduce bias as data samples were not collected from the same spot but from the surrounding locations. It is a challenge for repeatedly measuring the same area while conducting destructive sampling of biomass. We went about selecting close locations for the three times with the assumption that they might be achieving similar growth. While it is no brainer to sample the same spot repeatedly, the best practice could be ensuring similar growth conditions in the closely located sampling spots.

5. Conclusions

In conclusion, integrating linear mixed modeling framework enabled us to incorporate hierarchical structure of multi-field multi-temporal winter cover crop data into the modeling process and assess multiple sources of variability. Two modeling strategies with and without explanatory covariates were studied to assess the role of RS-derived and biophysical characteristics of winter cover crop in explaining springtime biomass variability. Incorporating RS-derived VIs like BGratio and SRrededge, along with biophysical characteristic such as CCpercent data, into the linear mixed model was found to improve overall model precision by explaining the leftover non-systematic noise in the data. While CCpercent and SRrededge showed positive relationship, BGratio showed no clear pattern relative to cereal rye biomass. There was evidence for varying relationships between the two variables (i.e., CCpercent and SRrededge) and cereal rye biomass across different management groups, determined by combination of planting timing and methods in the fall. Evidence was observed for relatively greater biomass accumulation throughout the spring season for cereal rye planted early in fall using drilled approach compared to late planted with aerial and broadcasting methods. In summary, this study underscores the efficacy of mixed modeling, coupled with biophysical and RS-derived measurements, in improving the understanding of winter cover crop growth and differences among management practices over a larger landscape.

Acknowledgments

This work was supported by funds from OSU L&L Grant# PG107271, SI Grant # PG107338, Ohio Soybean Council, USDA-AFRI Grant # GR130726, and Hatch Project #NC1195. We also want to thank Matthew Romanko, Boden Fisher, Brigitte Moneymaker, Abha Bhattarai, and Gaoshoutong Si for their support during field data collection.

References

- Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705
- Avneri, A., Aharon, S., Brook, A., Atsmon, G., Smirnov, E., Sadeh, R., Abbo, S., Peleg, Z., Herrmann, I., Bonfil, D. J., & Nisim Lati, R. (2023). UAS-based imaging for prediction of chickpea crop biophysical parameters and yield. Computers and Electronics in Agriculture, 205, 107581. https://doi.org/10.1016/j.compag.2022.107581
- Balkcom, K. S., Read, Q. D., & Gamble, A. V. (2023). Rye planting date impacts biomass production more than seeding rate and nitrogen fertilizer. Agronomy Journal, 115(5), 2351–2368. https://doi.org/https://doi.org/10.1002/agj2.21418
- Boyd, N. S., Brennan, E. B., Smith, R. F., & Yokota, R. (2009). Effect of Seeding Rate and Planting Arrangement on Rye Cover Crop and Weed Growth. Agronomy Journal, 101(1), 47–51. https://doi.org/https://doi.org/10.2134/agronj2008.0059
- Brennan, E. B., & Smith, R. F. (2023). Predicting cereal cover crop biomass using shoot length in California vegetable systems. Agricultural \& Environmental Letters, 8(1), e20099. https://doi.org/https://doi.org/10.1002/ael2.20099
- Broge, N. H., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sensing of Environment, 76(2), 156–172. https://doi.org/https://doi.org/10.1016/S0034-4257(00)00197-8
- Brooker, A. P., Renner, K. A., & Basso, B. (2020). Interseeding cover crops in corn: Establishment, biomass, and competitiveness in on-farm trials. Agronomy Journal, 112(5), 3733–3743. https://doi.org/https://doi.org/10.1002/agj2.20355
- Burnett, E., Wilson, R. S., Heeren, A., & Martin, J. (2018). Farmer adoption of cover crops in the western Lake Erie basin. Journal of Soil and Water Conservation, 73(2), 143–155. https://doi.org/10.2489/jswc.73.2.143
- Burnham, K., & Anderson, D. (2002). Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.
- Cao, Q., Miao, Y., Shen, J., Yu, W., Yuan, F., Cheng, S., Huang, S., Wang, H., Yang, W., & Liu, F. (2016). Improving in-season estimation of rice yield potential and responsiveness to topdressing nitrogen application with Crop Circle active crop canopy sensor. Precision Agriculture, 17(2), 136–154. https://doi.org/10.1007/s11119-015-9412-y
- Chen, P., Tremblay, N., Wang, J., Philippe, V., Huang, W., & Li, B. (2010). New index for crop canopy fresh biomass estimation. Spectroscopy and Spectral Analysis. https://doi.org/https://doi.org/10.3964/j.issn.1000-0593(2010)02-0512-06
- Cipoletti, N., Jorgenson, Z. G., Banda, J. A., Kohno, S., Hummel, S. L., & Schoenfuss, H. L. (2020). Biological consequences of agricultural and urban land-use along the Maumee River, a major tributary to the Laurentian Great Lakes watershed. Journal of Great Lakes Research, 46(4), 1001–1014. https://doi.org/https://doi.org/10.1016/j.jglr.2020.04.013
- CTIC. (2016). Annual report 2015-2016 cover crop survey.
- Daryanto, S., Fu, B., Wang, L., Jacinthe, P. A., & Zhao, W. (2018). Quantitative synthesis on the ecosystem services of cover crops. Earth-Science Reviews, 185(June), 357–373. https://doi.org/10.1016/j.earscirev.2018.06.013
- Duiker, S. W. (2014). Establishment and Termination Dates Affect Fall-Established Cover Crops. Agronomy Journal, 106(2), 670–678. https://doi.org/https://doi.org/10.2134/agronj2013.0246
- Duke, J. M., Johnston, R. J., Shober, A. L., & Liu, Z. (2022). Barriers to cover crop adoption: Evidence from parallel surveys in Maryland and Ohio. Journal of Soil and Water Conservation, 77(2), 198–211. https://doi.org/10.2489/jswc.2022.00062
- Farsad, A., Randhir, T. O., Herbert, S. J., & Hashemi, M. (2011). Spatial Modeling of Critical Planting Date for Winter Rye Cover Crop to Enhance Nutrient Recovery. Agronomy Journal, 103(4), 1252–1257. https://doi.org/10.2134/agronj2010.0433
- Feyereisen, G. W., Wilson, B. N., Sands, G. R., Strock, J. S., & Porter, P. M. (2006). Potential for a rye cover crop to reduce nitrate loss in southwestern Minnesota. Agronomy Journal, 98(6), 1416–1426. https://doi.org/10.2134/agronj2005.0134
- Finney, D. M., White, C. M., & Kaye, J. P. (2016). Biomass Production and Carbon/Nitrogen Ratio Influence Ecosystem Services from Cover Crop Mixtures. Agronomy Journal, 108(1), 39–52. https://doi.org/https://doi.org/10.2134/agronj15.0182
- Flombaum, P., & Sala, O. E. (2007). A non-destructive and rapid method to estimate biomass and aboveground net

- primary production in arid environments. Journal of Arid Environments, 69(2), 352–358. https://doi.org/https://doi.org/10.1016/j.jaridenv.2006.09.008
- Gamon, J., & Surfus, J. (1999). Assessing leaf pigment content and activity with a reflectometer. The New Phytologist, 143(1), 105–117. https://doi.org/DOI: 10.1046/j.1469-8137.1999.00424.x
- Gitelson, A., & Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. Journal of Photochemistry and Photobiology B: Biology, 22(3), 247–252. https://doi.org/10.1016/1011-1344(93)06963-4
- Goslee, S. C. (2020). Estimating pasture species biomass from canopy cover. Crop, Forage \& Turfgrass Management, 6(1), e20038. https://doi.org/https://doi.org/10.1002/cft2.20038
- Haramoto, E. R. (2019). Species, Seeding Rate, and Planting Method Influence Cover Crop Services Prior To Soybean. Agronomy Journal, 111(3), 1068–1078. https://doi.org/https://doi.org/10.2134/agronj2018.09.0560
- Hayden, Z. D., Ngouajio, M., & Brainard, D. C. (2015). Planting Date and Staggered Seeding of Rye–Vetch Mixtures: Biomass, Nitrogen, and Legume Winter Survival. Agronomy Journal, 107(1), 33–40. https://doi.org/https://doi.org/10.2134/agronj14.0237
- Hillel, D. (2003). Introduction to environmental soil physics. Elsevier.
- Huete, A. (1988). A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295–309. https://doi.org/https://doi.org/10.1016/0034-4257(88)90106-X
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83(1), 195–213. https://doi.org/https://doi.org/10.1016/S0034-4257(02)00096-2
- Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. Ecology, 50(4), 663–666. https://doi.org/https://doi.org/10.2307/1936256
- Kanke, Y., Tubaña, B., Dalen, M., & Harrell, D. (2016). Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. Precision Agriculture, 17(5), 507–530. https://doi.org/10.1007/s11119-016-9433-1
- M. A. Cho, A. K. S., & Atzberger, C. (2008). Towards red-edge positions less sensitive to canopy biophysical parameters for leaf chlorophyll estimation using properties optique spectrales des feuilles (PROSPECT) and scattering by arbitrarily inclined leaves (SAILH) simulated data. International Journal of Remote Sensing, 29(8), 2241–2255. https://doi.org/10.1080/01431160701395328
- Marcillo, G. S., Mirsky, S., Poncet, A., Reberg-Horton, C., Timlin, D., Schomberg, H., & Ramos, P. (2020). Using statistical learning algorithms to predict cover crop biomass and cover crop nitrogen content. Agronomy Journal, 112(6), 4898–4913. https://doi.org/https://doi.org/10.1002/agj2.20429
- Milliken, G., & Johnson, D. (2009). Analysis of Messy Data. Volume I: Designed Experiments (2nd ed.). Chapman & Hall/CRC Press.
- Moges, S. M., Raun, W. R., Mullen, R. W., Freeman, K. W., Johnson, G. V, & Solie, J. B. (2005). Evaluation of Green, Red, and Near Infrared Bands for Predicting Winter Wheat Biomass, Nitrogen Uptake, and Final Grain Yield. Journal of Plant Nutrition, 27(8), 1431–1441. https://doi.org/10.1081/PLN-200025858
- Montès, N., Ballini, C., Bonin, G., & Faures, J. (2004). A comparative study of aboveground biomass of three Mediterranean species in a post-fire succession. Acta Oecologica, 25(1), 1–6. https://doi.org/https://doi.org/10.1016/j.actao.2003.10.002
- Moore, V. M., & Mirsky, S. B. (2020). Cover crop biomass production across establishment methods in mid-Atlantic corn. Agronomy Journal, 112(6), 4765–4774. https://doi.org/https://doi.org/10.1002/agj2.20414
- Motew, M., Chen, X., Booth, E. G., Carpenter, S. R., Pinkas, P., Zipper, S. C., Loheide, S. P., Donner, S. D., Tsuruta, K., Vadas, P. A., & Kucharik, C. J. (2017). The Influence of Legacy P on Lake Water Quality in a Midwestern Agricultural Watershed. Ecosystems, 20(8), 1468–1482. https://doi.org/10.1007/s10021-017-0125-0
- Muñoz, J. D., Finley, A. O., Gehl, R., & Kravchenko, S. (2010). Nonlinear hierarchical models for predicting cover crop biomass using Normalized Difference Vegetation Index. Remote Sensing of Environment, 114(12), 2833–2840. https://doi.org/https://doi.org/10.1016/j.rse.2010.06.011
- Murrell, E. G., Schipanski, M. E., Finney, D. M., Hunter, M. C., Burgess, M., LaChance, J. C., Baraibar, B., White, C. M., Mortensen, D. A., & Kaye, J. P. (2017). Achieving Diverse Cover Crop Mixtures: Effects of Planting Date and Seeding Rate. Agronomy Journal, 109(1), 259–271. https://doi.org/https://doi.org/10.2134/agronj2016.03.0174
- Mutanga, O., & Skidmore, A. K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. International Journal of Remote Sensing, 25(19), 3999–4014. https://doi.org/10.1080/01431160310001654923

- Myers, R., & Watts, C. (2015). Progress and perspectives with cover crops: Interpreting three years of farmer surveys on cover crops. Journal of Soil and Water Conservation, 70(6), 125A--129A. https://doi.org/10.2489/jswc.70.6.125A
- OSUE. (2021). Soil Health Indicator Measurement Protocols. https://digitalag.osu.edu/sites/digitag/files/imce/publications/Soil_Health/OSUE Soil Health Indicator Measurement Protocol.pdf
- Patrignani, A., & Ochsner, T. E. (2015). Canopeo: A powerful new tool for measuring fractional green canopy cover. Agronomy Journal, 107(6), 2312–2320. https://doi.org/10.2134/agronj15.0150
- Pix4D. (2022). Pix4Dmapper. https://www.pix4d.com/product/pix4dmapper-photogrammetry-software
- Plastina, A., Liu, F., Miguez, F., & Carlson, S. (2020). Cover crops use in Midwestern US agriculture: perceived benefits and net returns. Renewable Agriculture and Food Systems, 35(1), 38–48. https://doi.org/DOI: 10.1017/S1742170518000194
- Prabhakara, K., Dean Hively, W., & McCarty, G. W. (2015). Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, United States. International Journal of Applied Earth Observation and Geoinformation, 39, 88–102. https://doi.org/10.1016/j.jaq.2015.03.002
- Roesch-Mcnally, G. E., Basche, A. D., Arbuckle, J. G., Tyndall, J. C., Miguez, F. E., Bowman, T., & Clay, R. (2018). The trouble with cover crops: Farmers' experiences with overcoming barriers to adoption. Renewable Agriculture and Food Systems, 33(4), 322–333. https://doi.org/10.1017/S1742170517000096
- Rosle, R., Che'Ya, N. N., Roslin, N. A., Halip, R. M., & Ismail, M. R. (2019). Monitoring Early Stage of Rice Crops Growth using Normalized Difference Vegetation Index generated from UAV. IOP Conference Series: Earth and Environmental Science, 355(1), 12066. https://doi.org/10.1088/1755-1315/355/1/012066
- Roth, L., & Streit, B. (2018). Predicting cover crop biomass by lightweight UAS-based RGB and NIR photography: an applied photogrammetric approach. Precision Agriculture, 19(1), 93–114. https://doi.org/10.1007/s11119-017-9501-1
- Rouse, J. W., Haas, R. H., Deering, D. W., Schell, J. A., & Harlan, J. C. (1974). Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation. https://doi.org/No. E75-10354
- Ruis, S. J., Blanco-Canqui, H., Creech, C. F., Koehler-Cole, K., Elmore, R. W., & Francis, C. A. (2019). Cover Crop Biomass Production in Temperate Agroecozones. Agronomy Journal, 111(4), 1535–1551. https://doi.org/10.2134/agronj2018.08.0535
- Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6(2), 461–464. https://doi.org/10.1214/aos/1176344136
- Sekaluvu, L., Zhang, L., & Gitau, M. (2018). Evaluation of constraints to water quality improvements in the Western Lake Erie Basin. Journal of Environmental Management, 205, 85–98. https://doi.org/https://doi.org/10.1016/j.jenvman.2017.09.063
- Sellaro, R., Crepy, M., Trupkin, S. A., Karayekov, E., Buchovsky, A. S., Rossi, C., & Casal, J. J. (2010). Cryptochrome as a Sensor of the Blue/Green Ratio of Natural Radiation in Arabidopsis. Plant Physiology, 154(1), 401–409. https://doi.org/10.1104/pp.110.160820
- Singer, J. W. (2008). Corn Belt Assessment of Cover Crop Management and Preferences. Agronomy Journal, 100(6), 1670–1672. https://doi.org/https://doi.org/10.2134/agronj2008.0151
- Smith, D. R., King, K. W., Johnson, L., Francesconi, W., Richards, P., Baker, D., & Sharpley, A. N. (2015). Surface Runoff and Tile Drainage Transport of Phosphorus in the Midwestern United States. Journal of Environmental Quality, 44(2), 495–502. https://doi.org/https://doi.org/10.2134/jeq2014.04.0176
- St Aime, R., Noh, E., Bridges, W. C., & Narayanan, S. (2022). A Comparison of Drill and Broadcast Planting Methods for Biomass Production of Two Legume Cover Crops. Agronomy, 12(1). https://doi.org/10.3390/agronomy12010079
- Staver, K. W., & Brinsfield, R. B. (1998). Using cereal grain winter cover crops to reduce groundwater nitrate contamination in the mid-Atlantic coastal plain. Journal of Soil and Water Conservation, 53(3), 230–240. https://www.jswconline.org/content/53/3/230
- Stroup, W. (2013). Generalized Linear Mixed Models: Modern Concepts, Methods and Applications (1st ed.). Chapman & Hall/CRC Press.
- Teshome, F. T., Bayabil, H. K., Hoogenboom, G., Schaffer, B., Singh, A., & Ampatzidis, Y. (2023). Unmanned aerial vehicle (UAV) imaging and machine learning applications for plant phenotyping. Computers and Electronics in Agriculture, 212, 108064. https://doi.org/https://doi.org/10.1016/j.compag.2023.108064

- Thaler, E. A., Kwang, J. S., Quirk, B. J., Quarrier, C. L., & Larsen, I. J. (2022). Rates of Historical Anthropogenic Soil Erosion in the Midwestern United States. Earth's Future, 10(3), e2021EF002396. https://doi.org/10.1029/2021EF002396
- Tucker, C. J., & Sellers, P. J. (1986). Satellite remote sensing of primary production. International Journal of Remote Sensing, 7(11), 1395–1416. https://doi.org/10.1080/01431168608948944
- USDA. (2018). Summary Report: 2017 National Resources Inventory.
- USDA. (2024). Agriculture in the Midwest. https://www.climatehubs.usda.gov/hubs/midwest/topic/agriculture-midwest
- USDA NRCS. (2022). Soil Survey Geographic (SSURGO) Database for Ohio. https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx
- Wallander, S., Smith, D., Bowman, M., & Claassen, R. (2021). Cover Crop Trends, Programs, and Practices in the United States. https://doi.org/10.22004/aq.econ.309562
- Wilson, M. L., Baker, J. M., & Allan, D. L. (2013). Factors Affecting Successful Establishment of Aerially Seeded Winter Rye. Agronomy Journal, 105(6), 1868–1877. https://doi.org/https://doi.org/10.2134/agronj2013.0133
- Woebbecke, D., Meyer, G., Von Bargen, K., & Mortensen, D. (1995). Color Indices for Weed Identification Under Various Soil, Residue, and Lighting Conditions. Transactions of the ASAE, 38(1), 259–269. https://doi.org/10.13031/2013.27838
- Yang, R.-C. (2010). Towards understanding and use of mixed-model analysis of agricultural experiments. Canadian Journal of Plant Science, 90(5), 605–627. https://doi.org/10.4141/CJPS10049
- Zhang, J., Qiu, X., Wu, Y., Zhu, Y., Cao, Q., Liu, X., & Cao, W. (2021). Combining texture, color, and vegetation indices from fixed-wing UAS imagery to estimate wheat growth parameters using multivariate regression methods. Computers and Electronics in Agriculture, 185, 106138. https://doi.org/https://doi.org/10.1016/j.compag.2021.106138
- Zhou, Q., Guan, K., Wang, S., Jiang, C., Huang, Y., Peng, B., Chen, Z., Wang, S., Hipple, J., Schaefer, D., Qin, Z., Stroebel, S., Coppess, J., Khanna, M., & Cai, Y. (2022). Recent Rapid Increase of Cover Crop Adoption Across the U.S. Midwest Detected by Fusing Multi-Source Satellite Data. Geophysical Research Letters, 49(22), e2022GL100249. https://doi.org/10.1029/2022GL100249