Optimization of temporal UAS-based imagery analysis to estimate plant maturity date for soybean breeding

Leonardo Volpato | Austin Dobbels | Aluízio Borem | Aaron Joel Lorenz

1 Department of Agronomy, Federal University of Viçosa, University Campus, Viçosa, Minas Gerais, 36570–900, Brazil
2 Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA

Correspondence
Aaron J. Lorenz, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA.
Email: lore0149@umn.edu

Abstract
Estimating the date of maturity of soybean breeding field plots is necessary for breeding line characterization and for informing yield comparisons among varieties. The main drawback of visually dating soybean maturity is the sheer scale of note recording entailed and the frequency at which these notes need to be taken. The overall aim of this study was to build upon prior work in using low-cost UAS-based RGB cameras to estimate soybean maturity date by examining the effect of vegetation index, summary statistic of the pixel values from each region of interest (plot), statistical model, and flight frequency. Maturity dates collected from five environments with 53 experimental trials (4,415 plots) were both visually dated and imaged using a RGB camera carried by a UAS. Using the mean greenness leaf index on each plot combined with LOESS regression, we achieved high correlations between ground and UAS-based estimates ($r = 0.84–0.97$). Precision, quantified by broad-sense heritability estimates, was greater for UAS-based dates in 29 of 53 field trials, and nearly equivalent in 11 more field trials. We found that 54% of the significant deviations between ground and UAS-based estimates were caused by inaccurate UAS-based estimates, while errors in the ground-based estimates accounted for 46% of the deviations. Reasons for these inaccurate estimates were attributed to lodging, presence of weeds, low germination, and within-line genetic heterogeneity in the plots. A detailed description of the analysis pipeline, a user-friendly R script, and all of the images and ground data have been made publicly available to help other researchers and breeders test and adopt these methods.

Abbreviations:
1 FlyW, one flight per week; 2FlyW, two flights per week; AllFly, all flights missions available; BLUP, best linear unbiased predictions; CI, confidence interval; DPMaerial, days to plant maturity estimated using UAS; DPMground, days to plant maturity by ground-based measurements; GCP, ground control point; GIS, geographic information system; GLI, greenness leaf index; GSD, ground sample distance; $H$, broad-sense heritability; HI, hue index; HTP, high-throughput phenotyping; LA, Lambertton; LOESS, non-parametric local polynomial regression; MG, maturity group; NGE, green excess index; $r$, Pearson’s correlation coefficient; $R^2$, coefficient of determination; RF, random forest classification; RGB, red, green, and blue; RM, relative maturity; RMSE, root-mean-square errors; RO, Rosemount; ROI, region of interest; SEG, segmented linear regression; TGI, triangular greenness index; UAS, unmanned aerial systems; VI, vegetation index; WA, Waseca

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *The Plant Phenome Journal* published by Wiley Periodicals LLC on behalf of American Society of Agronomy and Crop Science Society of America
1 | INTRODUCTION

Measuring the days from sowing to physiological maturity of each plot is a necessary and ubiquitous feature of soybean (Glycine max L.) variety trials, both in breeding programs and commercial variety testing programs. Soybean varieties vary in their time from germination to physiological maturity, which is controlled by genetic factors that respond to both photoperiod and temperature (Schoving et al., 2020; Setiyono et al., 2007). Collecting reliable data on days to maturity for each variety is essential for determining proper adaptation to a particular latitudinal zone. Varieties that mature too late may be damaged by frost, while those that mature too early may not maximize yield in a latitudinal zone. Another key reason for the collection of maturity dates at the field plot level is to inform the comparison of yield between varieties. Varieties that mature later can be expected to have higher yields than those that mature earlier, and therefore yields should only be compared between varieties that mature on similar dates (Mourtzinis & Conley, 2017; Zdziarski et al., 2018). Maturity date can also be used as a covariate in a statistical analysis so that yield values are all expressed on the basis of an equivalent maturity date (Diers et al., 2018).

As soybean plants begin to mature, their color progresses from light green to yellow to fully brown with senesced leaves. The duration of this progression is approximately 3 wk, but it can greatly vary with the environment and variety. Soybean breeders record the date of soybean maturity by visually observing when 95% of the pods in a plot reach their mature color (tan, brown, or black) (Fehr & Caviness, 1977). The main drawback of visually dating soybean maturity is the sheer scale of note recording entailed and the frequency at which these notes need to be taken. A typical soybean breeding program could quickly require visual maturity dating on tens of thousands of plots spread across multiple locations, with the frequency of dating being two times per week until all plots have reached maturity. This level of phenotyping imposes a high cost in human resources. Moreover, human errors and subjectivity can unnecessarily reduce the reliability of maturity date data. The use of hand-held sensors for estimating soybean maturity date has been investigated with the goal of increased precision and throughput (Christenson et al., 2016; Lindsey et al., 2020; Ma et al., 2001), but the manual nature of these devices is still limiting when thousands of plots need to be evaluated.

Unmanned aircraft systems (UASs) have become less expensive and more versatile in recent years, allowing a large variety of cameras to be flown to collect aerial imagery at high spatial and temporal resolution (Araus et al., 2018; Holman et al., 2016; Peña et al., 2015; Reynolds et al., 2020; Torres-Sánchez et al., 2013). Many different applications of UAS technology to improve soybean breeding programs have been studied for traits including soybean abiotic stress tolerance, canopy coverage, maturity date, and yield (Bai et al., 2016; Castelao Tetila et al., 2017; Dobbels & Lorenz, 2019; Gao et al., 2018; Huang et al., 2018; Jarquin et al., 2018; Yu et al., 2016b). The success of any image collection and analysis pipeline depends on various factors, including choice of suitable analysis algorithms and statistical models (van Eeuwijk et al., 2019); temporal and spatial resolution of imagery (Belgiu & Stein, 2019; Holman et al., 2016); and sensor type and treatment of reflectance data (Chawade et al., 2019; Krause et al., 2019). Yu et al. (2016) Maes & Steppe, 2019; Yang et al., 2017 used UAS-collected imagery combined with a random forest statistical model and found that they could correctly classify over 95% of the plots within one flight date of the ground-based maturity date estimate. However, the binary maturity classification (mature versus not mature on each flight date) used by Yu et al. (2016) limited the resolution of dating to the frequency of flights, which ranged from one flight every 3–6 d. Achieving daily resolution using this method would require collection of images each day throughout the maturation of all plots in a trial. Another approach to predicting maturity is to extract many features from an image taken with a multispectral camera, including changes in reflectance values between flight dates, and use techniques in dimension reduction and regression to estimate the maturity date from the extracted image features. Zhou et al. (2019) used this method by extracting features from imagery collected with a five-band multispectral sensor and associated these predictor variables with maturity date in a calibration set using partial least-squares regression, allowing maturity date estimates to take on any value in between flight dates. Using a 10-fold cross-validation, UAS-based maturity estimates explained 81% of the variation in ground-based estimates. Disadvantages to this approach include the need to collect ground data on a calibration set at each field trial site, a requirement for a multispectral sensor, and overall greater complexity of feature extraction and optimization of multivariate models.
TABLE 1 Details on each environment, unmanned aircraft systems (UASs), and flight settings used in this study. A total of five environments were used between 2018 and 2019 to estimate days of plant maturity (DPMaerial). Each field site varied in area (ha), number of trials, the total number of plots, planting date, number of ground control points (GCPs), and type of unmanned aircraft system (UAS) platforms used. Locations in Southern Minnesota: LA, Lamberton; WA, Waseca; RO, Rosemount.

| Item                  | 2018 WA | 2018 LA | 2019 WA | 2019 LA | 2019 RO |
|-----------------------|---------|---------|---------|---------|---------|
| No. trials            | 12      | 11      | 6       | 9       | 15      |
| No. plots             | 873     | 794     | 482     | 684     | 1,382   |
| Planting date         | May 17  | May 16  | May 15  | May 16  | May 26  |
| Area total, ha        | 3.04    | 3.40    | 2.42    | 2.14    | 3.61    |
| No. flights           | 9       | 12      | 18      | 18      | 16      |
| Flight period, start/end date | Aug. 8 /Oct. 12 | Aug. 15 /Oct. 11 | Aug. 30 /Oct. 14 | Aug. 28 /Oct. 16 | Aug. 28 /Oct. 7 |
| GCP no.               | 5       | 4       | 4       | 4       | 9       |
| UAS platform          | DJI Phantom3 Pro | DJI Phantom3 Pro | DJI Phantom3 Pro | DJI Phantom3 Pro | DJI Inspire 1 |

An ideal method for estimating the maturity date of soybean plots using UAS imagery would use simple red-green-blue (RGB) cameras and simple modeling relating the spectra to maturity date, not requiring model calibration. Narayanan et al. (2019) accomplished this with reflectance data obtained from an RGB camera attached to a UAS, calculated a “green excess index” (NGE), and modeled the decay of the NGE values of each plot over time with a segmented (or piece-wise) linear regression model. Using this approach, Narayanan et al. (2019) found they could estimate soybean plot maturity with good reliability ($R^2 = 0.58 – 0.96$) and precision. The broad-sense heritabilities of the UAS-based estimates were comparable to those of the ground-based estimates across 22 trials. However, Narayanan et al. (2019) did not systematically investigate many variables that may affect the accuracy and precision of estimates, such as flight frequency, vegetative index, and summary statistic of pixels in the region of interest. Moreover, the influence of various practical challenges in a field trial setting on the accuracy of UAS-based maturity date estimates has not been fully explored and quantified. These challenges include presence of weeds, genetic heterogeneity within a plot, low germination, and lodging.

The overall aim of this study was to build upon the work of Narayanan et al. (2019) in using low-cost UAS-based RGB cameras to estimate soybean maturity date by examining the effect of (a) three different vegetation indices (VIs) calculated from RGB bands, (b) five different summary statistics on the pixel values extracted from each region of interest (ROI), (c) two different statistical models, and (d) four flight frequencies. In addition to studying the effect of these factors on accuracy of UAS-based maturity date estimates, we also aimed to categorize and quantify underlying reasons for significant deviations between UAS-based (DPMaerial) and ground-based (DPMground) maturity dates. Finally, another goal of ours was to make available a detailed description of our analysis pipeline, an R script to implement the analysis, and all images and data so that other researchers and breeders can test and adopt these methods to their own programs.

2 | MATERIAL AND METHODS

2.1 | Experimental design and ground-based dating of plant maturity

A set of 53 trials of breeding lines at different stages of testing in the University of Minnesota (UMN) soybean breeding program were visually dated for maturity and aerially imaged via UAS. The experiments were carried out over two growing seasons (2018 and 2019) at three locations in southern Minnesota, USA: Waseca (WA) (44°03′56.0″N, 93°33′01.8″W), Lamberton (LA) (44°16′57.0″N, 95°16′09.6″W), and Rosemount (RO) (44°42′33.6″N, 93°04′28.4″W) (Table 1). Individual trials consisted of breeding lines and check cultivars of similar relative maturity so that the total range of maturity within a trial did not exceed one full maturity group. Across trials, maturity groups ranged from late MG 0 to early MG II cultivars and breeding lines. Field plots (experimental units) consisted of two- or four-row rows. Rows were 3.66 m in length and 0.76 m apart. Alleys between the plots were perpendicular to the rows and were 0.91-m wide. Plots were arranged in a randomized complete block design (RCBD) with two or three replications, depending on the stage of the trial.

Locations were visited for ground-based visual dating of maturity (DPMground) as soon as the earliest maturing varieties in a given location began to senesce. A plot was determined to be mature when approximately 95% of pods exhibited their mature color (Fehr & Caviness, 1977). The DPMground notes were taken by observing and revisiting plots as they matured, eventually assigning a date of maturity to each plot. Plots were visited every 3 d, on average, and
maturity dates were interpolated when it was clear a plot reached maturity in between site visits. The date of plot maturity was expressed as the number of days after 31 Aug. for all environments, and thus negative values indicate maturity dates before 31 Aug.

2.2 UAS platform and image data collection

The DPMaerial was collected using either a DJI Phantom 3 Pro (DJI Technology Co., Ltd.) or DJI Inspire 1 model T600 (SZ DJI Technology Co.) equipped with digital red-green-blue (RGB) cameras (Zenmuse X3 equipped to Inspire 1 platform). A single platform was used within each environment. Flights were conducted across the same time range as DPM-ground were recorded (Table 1). All flights were performed within 1 h of solar noon to limit shadow effects. In 2018, flights were conducted once per week, while in 2019, we aimed to fly three times per week. The exact number of days between flights slightly varied if weather conditions (rain, wind) prevented flying. The flight frequency in 2019 allowed the simulation of one (1FlyW) or two (2FlyW) flights per week through the masking of individual flights (Figure 1). The 1FlyW scenario was created in two different ways as the flights did not evenly stagger within 1 wk (Figure 1). The DPMaerial-DPMground correlations from these two different 1FlyW scenarios were averaged because the results were very similar.

All flight missions were conducted at an altitude of 61 m with a UAS speed of 5 m/s. Images were captured with 75%
end, and side lap with area covered higher than the actual trial area to ensure an accurate orthomosaic product. As a result, images had a ground sampling distance (GSD) of approximately 2.4 to 2.6 cm. Waypoints and flying routes were automatically generated using the flight planning software, Pix4Dcapture (v4.8.0; Pix4D, Prilly). A tablet was used to prepare missions prior to going to the field, which were then uploaded to the main control board on the UAS before flight. During each flight, the UAS traveled a predetermined route via the software’s autonomous flying mode.

Painted red blocks (30 x 30 cm) were placed as permanent ground control points (GCPs) to help ensure accurate georeferencing of images. The GCPs were placed randomly throughout the entire field trial area. The GCPs remained in the field for the duration of the season. The GCPs were surveyed using a Garmin eTrex 30x hand-held device (Garmin). Specific platform and flight information for each environment are provided in Table 1.

2.3 Image data processing

Orthomosaics were generated using Pix4D Mapper software (v4.5.4; Pix4D, Prilly), which provides an automated pipeline including photo alignment, matching, and bundle adjustment. The WGS 84 datum was used with a projected coordinate system of UTM zone 15 N. Images were imported into Pix4D and then optimized and matched using the default processing options template ‘Ag RGB’ for generating georeferenced orthomosaics. This option generates orthomosaics from overlapping nadir images and outputs a full resolution GeoTIFF file. The GCPs were input and identified into the Pix4D project using the basic manual editor before initial processing.

Python 2.7.1 (Python Software Foundation) software was used to generate the polygon shapefiles (ROI) according to plot boundary delimitation following Haghighattalab et al. (2016). Shapefiles were defined using images collected just prior to the full flower development stage (R2). ArcGIS (version 10.7, Esri Inc.) buffer tool was used for polygon adjustments of the ROI. The spatial polygon vectors (or shapefiles) were outlined with a buffer zone of 10 cm for each plot to prevent any influence of neighboring plots. Additionally, the ROI of the plots were manually aligned as needed on the center of the plot.

Using the ‘raster’ (Hijmans, 2020) and ‘FIELDImageR’ (Matias et al., 2020) packages in R, 17 VIs (Supplementary Table S1) were calculated using the digital numbers (DN) from RGB-bands via images collected in 2018. Resulting VI geo-TIFF files were imported into R for polygon value extractions using the mean pixel values from each ROI. Plot boundary delineation was performed using the package ‘rgdal’ (Bivand et al., 2020) to import the shapefile, and the package ‘raster’ was used to manipulate spatial data (pixels) in raster format file.

A preliminary analysis was performed using the ‘randomForest’ R package (Liaw & Wiener, 2018) to relate the 17 VIs to DPMground. The DPMground data was treated as a binary variable (mature/not mature), and 70% of the data, at random, was used to train the model, with the remaining 30% of data used for validation. The evaluation of the importance of each variable (17 VIs) in the prediction process suggested three indices based on variable importance rankings (Strobl et al., 2007) (data not shown): greenness leaf index (GLI), triangular greenness index (TGI), and hue index (HI). The GLI was initially created for use with a digital RGB camera to measure crop cover (Louhaichi et al., 2001) and the TGI to estimate the area of a triangle bounding a leaf reflectance spectrum (Hunt et al., 2012). Both indices have been applied in studies to measure differences in leaf chlorophyll content at both individual leaf and whole canopy scales (Hunt et al., 2011; Saberioon et al., 2014). The HI index has also been used in vegetation applications (Escadafal et al., 1994; Zhou et al., 2019). These three indices were subsequently used to model the decay in plot greenness through the senescence phase in this study.

The threshold value of each index corresponding to plant maturity was calculated by taking the average pixel values for each VI of all plots that were visually scored as mature on the day of imaging (Narayanan et al., 2019; Supplementary Figure S1). The effect of the predetermined threshold was evaluated by the correlation between DPMground and DPMaerial. For the GLI index, we found the optimal threshold to be very similar to that of Narayanan et al. (2019). Nevertheless, we recommend that users of these methods first predetermine the optimal threshold for their specific germplasm and geographic region. Once a threshold is determined, it may be used in future trials without prior collection of ground data.

Using only the top three VIs selected (GLI, HI, and TGI), the same procedures were used to manipulate pixel value extractions from the entire dataset with the R packages ‘rgdal’ and ‘raster.’ Five different summary statistics on the distribution of pixel values from the ROI were explored: overall mean pixel value; overall median pixel value; and the mean of central 75, 25, and 5% of the pixel value distribution from each ROI (Figure 1). The purpose of testing different ways in which to summarize the ROI pixel values was to determine if they can be used in an attempt to avoid off-target pixels (e.g., weeds, plants of neighboring plots, bare ground).

A script for data extraction from images and downstream analyses was developed to the procedures described above. The DPMaerial scripts and processes used to perform the image analyses and trait extract are available at https://github.com/volpatoo/HTP-via-drone-imagery. All raw and processed image output files from this study are publicly available at Cyverse (Lorenz et al., 2020).
2.4 | Models to estimate days to plant maturity

Two models to estimate DPMaerial from RGB spectra were investigated: Non-parametric local polynomial regression (LOESS) and segmented linear regression (SEG). The SEG model was used in Narayanan et al. (2019). Non-parametric local polynomial regression is a localized regression approach that fits simple linear models to localized subsets of data determined by a nearest neighbor algorithm. It was our hypothesis that this type of localized regression model could capture decay in greenness well by accounting for chance fluctuation in the rate of decay caused by random weather events. This could make it superior to the SEG model that fits a single linear model over the whole senescence phase. However, this feature of LOESS also makes it susceptible to outliers in the data. In this study, the default settings of the LOESS function were implemented using the ‘stats’ package from the base R software in order to interpolate VIs values between flights following the methodologies described in Cleveland and Devlin (1988).

The DPMaerial based on decaying canopy greenness using SEG was similar to what was done by Narayanan et al. (2019). Segmented linear regression is a form of regression in which multiple linear models are fit to the data for different ranges of the explanatory variable (x). The explanatory variable in this study was days relative to 31 Aug. We used the ‘segmented’ package (Muggeo, 2019) in R to identify the senescence phase curve. The segmented function first tests for differences between slopes using the Davies test (Davies, 1987) and then estimates the break-points between the linear models using maximum likelihood (Muggeo, 2003). Break-points are the values of x where the linear slope of the function changes. When the number of flights was equal to or greater than two per week, two break-points were modeled, resulting in three curves. Otherwise, only one break-point was modeled because of difficulty in converging on solutions when two break-points were attempted to model one flight per week. The curve with the greatest absolute value in slope was selected as that representing the senescence phase of each plot. The scales of the indices were such that senescence phase curve slopes were negative for GLI and TGI, but were positive for HI. Using the senescence phase curve, the estimated DPMaerial was calculated as \(\frac{(T_i-a_{ij})}{b_{ij}}\), where \(T_i\) is the threshold for the VI i (Supplementary Figure S1); \(a_{ij}\) is the intercept of the senescence phase curve model for VI i and plot j; and \(b_{ij}\) is the slope of the senescence phase model of VI i and plot j. The function ‘seg.control’ from the ‘segmented’ package was used as an auxiliary function to fitting the model using 50 bootstrap samples.

The georeferenced orthomosaics were generated by Pix4D across multiple time points, and the shapefiles to ROI delimitation were created in Python and tuned using ArcGIS. The DPMaerial R script made available as part of this study first converts each orthomosaic into a user-defined RGB based index (GLI, HI, TGI) and extracts a user-defined summary statistics (mean of central 75%, 25%, 5%, overall mean, or overall median) of pixel values within the ROI. The user also can execute either a LOESS or SEG regression model, and choose the VI threshold corresponding to plot maturity.

2.5 | Data analysis and model validation

Each of the 53 trials included in this study was arranged in a RCBD with at least two replications in each environment. Within each environment, the best linear unbiased predictions (BLUPs) of the genotype were calculated for the DPMground and DPMaerial dates. BLUPs were obtained for each trial using the following model:

\[ y_{ij} = \mu + g_i + r_j + e_{ij} \]

where \(y_{ij}\) is the value for the plot of genotype i and replicate j, \(\mu\) is the grand mean, \(g_i\) is the random effect for genotype i, \(r_j\) is the fixed effect of replicate j, and \(e_{ij}\) is the random residual. Both the genotype and residual effects were assumed to be independent and identically distributed random variables from normal distributions. Genotypic (\(\sigma^2_g\)) and residual (\(\sigma^2_e\)) variance estimates from each trial were used to calculate broad-sense heritability (\(H\)) estimates on a plot basis (\(H = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e}\)) for both DPMground and DPMaerial. Statistical analyses were carried out using ASReml-R (version 4.1) software (ASReml, VSN International).

The DPMground and DPMaerial were compared using Pearson correlation coefficients (r), coefficients of determination (R²), root-mean-square errors (RMSE), and \(H\) calculated on individual plot values in each environment. Confidence intervals (CIs - 95%) of Pearson correlation coefficients were calculated using the pairwise correlation function in JMP Pro software analysis (JMP, Pro 14, SAS Institute Inc.), which implements Fisher’s z’ transformation. Also, 95% CIs of \(H\) were computed using the standard errors (Schweiger et al., 2016) from each trial with the ‘pin’ function through the R package ‘nadiv’ (Wolak, 2019).

Outliers were defined in two ways: (a) based on DPMground alone and (b) by analyzing deviations between DPMaerial and DPMground. For (a), outliers were identified using studentized residuals of DPMground by JMP Pro software analysis. An analysis of variance was conducted for each of the 53 trials, including the genotype and replicate sources of variation. Studentized residuals were calculated for the DPMground value of each plot, and individual values were considered as outliers if the absolute value of the studentized residual
Correlations between visual maturity date scores (DPMground) and UAS-based imagery (DPMaerial) for local regression (LOESS) and segmented regression (SEG) models using all flights available for each environment. Horizontal lines represent 95% confidence intervals. Shaded boxes show results for the three different VIs (GLI, TGI, and HI) for the overall pixels value mean of each plot, while the unshaded portions represent the results for the comparison of five-pixel filtering criteria of different central quantiles and summary statistics (mean, median, 75q, 25q and 5q) for GLI only. Locations in Southern Minnesota: LA, Lamberton; WA, Waseca; RO, Rosemount.

was greater than three. For (b), a regression model between DPMaerial and DPMground was analyzed using the function `outlierTest` from the `car` package (Fox et al., 2019) in R software with the default settings from the package. The outlier test ($P < .01$) uses a Bonferroni $p$ value to test for outlying deviations between the DPMaerial and DPMground in the linear model.

### 3 RESULTS

Variation in DPMground was first analyzed for the 53 trials and 5 environments. Across all trials DPMground ranged from 5 to 50. The range within trials varied substantially, with standard deviations within trials ranging from 2.05 to 8.14 in 2018 LA, 1.87 to 6.65 in 2018 WA, 3.56 to 6.27 in 2019 LA, 2.96 to 4.2 in 2019 WA, and 1.79 to 5.4 in 2019 RO (Supplementary Figure S2). Breeding lines were a significant source of variation for DPMground in all trials (results not shown).

Using correlations between DPMground and DPMaerial, we first set out to compare the VIs GLI, HO, and TGI. Overall, GLI provided the highest correlations between DPMaerial and DPMground, although differences did not exceed the CI with the exception of a couple of instances comparing HI to GLI in LA 2019 and RO 2019 (Figure 2). These results, combined with those reported by Narayanan et al. (2019), persuaded us to choose GLI as the single index value in all subsequent analyses.

The LOESS model resulted in significantly higher correlations to DPMground in most scenarios, and in no scenarios was SEG significantly better than LOESS (Figure 2). Overall, the LOESS model showed a slight advantage in 2019 LA and RO and a significant difference against the SEG model 2018 LA and WA. The only exception was in 2019 WA, in which the SEG model showed a higher correlation than the LOESS model, however, the difference did not exceed 1 CI (Figure 2).

The DPMaerial – DPMground correlations for different ROI pixel value summary statistics were also very similar across all environments (Figure 2). Because each of the summary statistics provided similar overall results across uniform full vegetation soybeans plot, the mean value of each ROI was considered in all subsequent analyses. However, we did find that for plots affected by genetic heterogeneity, lodging, and weeds (creating outliers as discussed below) that
only using pixels from the central quantiles of the distribution reduced the average deviation size between DPMground and DPMaerial (Supplementary Figure S3). This indicates that pixel filtering may increase the accuracy of DPMaerial estimates for problematic plots because it effectively reduces the influence of non-target pixels. Since the proportion of plots affected by these factors was small, the overall DPMground-DPMaerial correlation was not affected.

Flight frequency was combined with two statistical models – LOESS and SEG – in determining whether a statistical model was particularly best suited to a specific flight frequency. The DPMground-DPMaerial correlation plateaued for the LOESS model after one flight per week in all 2019 environments (Figure 3). In the 2019 LA and WA environments, DPMground-DPMaerial correlations using the SEG model were not maximized until flights were conducted at least two times per week (Figure 3). In RO, the predictive ability of DPMaerial clearly plateaued at one flight per week for both models, with the SEG model providing numerically better correlations for all flight frequencies. All results clearly showed that flights conducted only biweekly were insufficient to maximize the predictive ability of ground data (Figure 3). All further analyses used only one flight per week because of the diminishing returns in predictive ability we observed and the fact that using only one flight per week allowed us to include the 2018 data.

The general accuracy of DPMaerial estimates was assessed by choosing pipeline settings determined to maximize the correlations per the previously described results above. The chosen settings were one flight per week, GLI, LOESS, and mean pixel value of the ROI. As described above, results from different settings were very similar, and thus these settings were also selected on the basis of simplicity. The mean of $R^2$ value across the five study environments was 0.80, ranging from 0.57 in RO-2019 to 0.93 in LA-2018. The 2018 environments realized $R^2$ values between DPMaerial and DPMground of 0.92 and 0.93 (Supplementary Figure S4).

The $H$ for DPMaerial was numerically higher than the $H$ of DPMground in 29 of the 53 trials and nearly equivalent (within 0.05) in nine trials (Figure 4). In no trials was the $H$ of DPMground significantly greater than that of DPMaerial. Averaged across all trials, the $H$ of DMPaerial was only 0.02 higher than DPMground, with differences ranging from −0.36 to 0.23 (Figure 4). The DPMaerial $H$ of many of the RO 2019 trials was markedly better than DPMground, reflecting poor precision of ground notes in this environment, which is the likely reason why the correlation between DPMaerial and DPMground was lower in RO 2019 compared to the other environments.

Five percent of the total plots were defined as outliers either on the basis of the studentized residual from the analysis of variance (criteria $a$ described in the Methods section) or the deviation between DPMground and DPMaerial (criteria $b$). To determine whether the outliers resulted from errors in DPMaerial or DPMground, we visually inspected the UAS-collected imagery of the plots. This re-inspection of the plots made possible by the collected imagery revealed that for 46% of the outliers, the DPMground data point was the
FIGURE 4 Broad sense heritability ($H$) estimates for 53 trials (X-axis) across five environments for DPMground and DPMaerial. Greenness leaf index (GLI) and LOESS were combined along with the overall mean pixel value and one flight per week. Vertical and horizontal bars represent the 95% confidence intervals of the $H$ estimated. Locations in Southern Minnesota: LA, Lamberton; WA, Waseca; RO, Rosemount.

FIGURE 5 A breakdown of the 212 (5%) outliers according to the cause of the outlier determined upon close inspection of the images source of error, while 54% of the outliers could be attributed to error in the DPMaerial estimate. The DPMground errors were attributed to human errors. We inspected the images further to determine the cause of DPMaerial errors. It appeared that most DPMaerial errors were caused by within-plot heterogeneity (48%), which could be due to either genetic segregation at major maturity loci or seed lot contamination. The remaining errors were attributed to lodging (24%), presence of weeds (18%), and low germination (10%) (Figure 5).

The way in which the four identified plot issues affected the DPMaerial estimates varied (Figure 6). In the case of lodging, an adjacent plot is leaning into the ROI, and the greenness value of the ROI remains above the predetermined threshold, effectively delaying the DPMaerial relative to DPMground estimates. Similar to adjacent lodging plots, weeds within a plot also keep the plot greenness above the threshold past the true maturation date (Figure 6). Low germination results in plots not developing a full canopy before senescence. In this case, the plots tended to be classified by the UAS pipeline as mature before the true date of maturity. The removal of these outliers from the dataset improved the DPMground-DPMaerial correlation as expected. Thus, this factor could better reflect the predictive ability of DPMaerial estimates because the correlation is less affected by human errors (Figure 7). Improvements in the DPMaerial-DPMground correlation averaged 0.04 across trials and ranged from 0 to 0.24 (Supplementary Table S2).

4 | DISCUSSION

4.1 | Temporal greenness modeling and flight frequency

Modeling the decay of soybean canopy greenness through remote sensing and statistical models is useful for estimating soybean maturity date. It has also been shown that taking a classification approach using random forest models on individual time points can be accurate (Yu et al., 2016). The challenge with the classification approach, however, is that soybean plots will mature in between flights, and by using a binary classification scheme, it is not possible to interpolate between flights, thus limiting the resolution to the flight frequency. Narayanan et al. (2019) proposed a method to overcome this limitation by modeling greenness decay over time.
FIGURE 6  Visualizations of plots representing each one of the four issues causing significant deviations between DPMaerial and DPMground. The index used was GLI, and the relationship between GLI and date was smoothed using LOESS. The green dashed horizontal line represents the index values threshold corresponding to a fully mature plot in this study. Images at the bottom are time-series images of the same plot from which the data were collected and displayed in the plot.

FIGURE 7  Coefficient of determination ($R^2$), root mean square deviation (RMSE), and number of plots (N) for the linear relationship between DPMground and DPMaerial. Ninety-eight outliers caused by errors in ground notes were removed in total. DPMaerial was estimated using GLI, LOESS, overall mean pixel value, and one flight per week. “DPMaerial deviance” is the absolute difference between DPMaerial and DPMground. Locations in Southern Minnesota: LA, Lamberton; WA, Waseca; RO, Rosemount.

by first characterizing three different phases of each plot and then using regression on the senescence phase to estimate the date of plot maturity. Our study builds upon Narayanan et al. (2019) by exploring different levels of settings in the image collection and maturity analysis in terms of balancing simplicity with accuracy and precision. Another output of this study is the release of a user-friendly R script to calculate DPMaerial, allowing the researchers and breeding programs to implement these methods easily. Additionally, all data and images used in this study have been made publicly available.
A threshold for index values is necessary to determine when a plot has reached a level of greenness corresponding to maturity (Figure 1). Threshold values were derived by extracting VIs from images of all fully mature plots in which images were collected on the same day of manual scoring (Supplementary Figure S1). A higher threshold indicates that maturity occurred early while there was still green residue apparent in the plots. Narayanan et al. (2019) explored the effect of threshold adjustment for NGE on accuracy and concluded that threshold values between 0.01 and 0.05 did not affect estimation accuracy. Our results were similar and suggested that a GLI threshold of 0.02 was optimal. Nevertheless, testing different threshold levels is recommended in each case because it may change based on region, sensor hardware, and germplasm. An optimization study could be performed to determine the best threshold, and once that is determined, maturity dates could be estimated using DPMaerial without the need to collect additional ground-truth data. The R script accompanying this manuscript takes the threshold setting as an input so it can be changed by the user.

We found that a single flight per week (1FlyW) achieved similar accuracy compared to flying as many as two or three times per week (Figure 3). Minimizing flight frequency would ease data collection logistics such as travel to potentially distant sites for image collection, navigating unfavorable weather conditions, and time dedicated to processing imagery. We also found that it is not necessary to have all three phases of the canopy greenness decay captured to accurately estimate maturity date, as was reported to be a requirement by Narayanan et al. (2019). Depending on flight timing, either the pre-senescence or post-senescence phases might not be captured due to logistical concerns within a breeding program. However, as long as the senescence phase is captured, an accurate calculation of the maturity phenotype for both models (LOESS and SEG) can be made. Since LOESS fits linear models to localized subsets of data, flights are required both before and after the date of maturity. Practically, this means that flights will need to continue until every plot reaches maturity. On the other hand, the SEG model can estimate DPMaerial even if plots do not reach maturity within the range of flight dates.

### 4.2 Index values, spectral calibration, and pixel summary statistics

One purpose of VIs is to enhance the vegetation signal while minimizing soil background and solar irradiance effects (Huete et al., 2002; Wang et al., 2017). Some authors have shown the importance of excluding soil pixels when performing analysis of VIs (Schirrmann et al., 2016; Thorp & Tian, 2004; Torres-Sánchez et al., 2013). Spectral calibration is a recommended remote-sensing technique to correct the reflected light detected by the sensors (i.e., RGB or multispectral cameras) (Yu et al., 2016; Holman et al., 2019; Zhou et al., 2019) to avoid the effect of changing natural light conditions over the course of a flight day, as well as sky conditions such as humidity and clouds. Many authors have discussed protocols to calibrate remote sensing imagery collected by UAS (Cao et al., 2019; Iqbal et al., 2018; Xu et al., 2019). However, our results, as well as those of Narayanan et al. (2019), suggest that these time-consuming steps may not be critical for tracking soybean plot senescence through time. The dramatic change in plant color and canopy density through leaf senescence is easily captured, making masking and spectral calibration possibly unnecessary.

Selecting alternative statistics to summarize the pixel value distribution from a ROI could be useful to reduce the influence of non-representative pixels from areas such as neighboring plots or field alleys. This approach has been used by other authors to select target pixels inside each ROI to estimate plant height (Hassan et al., 2019; Madec et al., 2017) and lodging (Singh et al., 2019). Testing different measures applied to pixel extraction is recommended to assist researchers in choosing which is most appropriate for a specific crop, trait, and field design. Haghighattalab et al. (2016) developed an automatic image classification method but pointed out that post-processing and manual adjustments were still necessary on images of wheat plots late in the season. In our study on soybean maturity date evaluated on elite germplasm under uniform field conditions, we found that the mean of the pixel values was just as good as the alternatives we tested. However, these results were obtained after precisely defining the ROI using GIS tools (i.e., ArcGIS software) to adjust the polygons to cover each plot manually. Moreover, we gathered early-season imagery (~R1 growth stage) to facilitate the creation of accurate and centralized plot boundaries made possible by visible rows before complete canopy closure. For fine adjustment of UAS-based imagery applications and to reduce non-target pixel influence (i.e., neighborhood plots) in images, methods for pixel values extractions can be explored, plus an additional buffer zone inside each target plot, as performed in this study.

### 4.3 Underlying causes of outliers

There are practical issues to consider in field trials that can introduce errors when implementing UAS-based imagery. In this study, significant deviations between DPMaerial and DPMground measures were visually inspected in the images to determine the sources of these differences. As mentioned, human visual scoring methods may include human bias and mistakes, which ultimately decreases correlations between DPMground and DPMaerial (Chapman et al., 2014; Jin et al., 2017; Song & Wang, 2019). Accurate phenotyping is
extremely important in experimental validation (Furbank & Tester, 2011; Grońkowski et al., 2015), and it has been a bottleneck when studying HTP technology (Han et al., 2018; Loladze et al., 2019). One aim of this study was to enumerate and characterize these errors in an attempt to provide insight into any issues related to the practical application of these technologies.

Among the outliers caused by field plot issues, within-plot genetic heterogeneity was the primary cause (Figure 6). This could be due to either seed contamination or segregation of major maturity loci because breeding lines were not derived from completely homozygous plants. Lodging was also an important cause of errors in DPMaerial, followed by weeds and low germination. There are some options to avoid these errors in the future, including the obvious ones such as ensuring pure source seed lots, minimizing the presence of weeds, and only planting seeds with high germination rates. The issue of lodging would be less problematic in elite commercial trials as compared to earlier generation materials or special studies involving diverse germplasm. Another helpful step would be to take field notes on plots experiencing such issues so that DPMaerial estimates can be flagged for added scrutiny or just visually dated outside the image analysis pipeline. Finally, it may also help to filter out all pixels save for only the central part of the distribution (central 5%) so that non-representative pixels from soil or weeds, for example, do not unduly influence estimates. We examined the five different pixel summary statistics on only those plots deemed as outliers and found that this technique may reduce the magnitude of these errors (Supplementary Figure S3).

5 | CONCLUSIONS

Findings from this study suggest there is flexibility in the optimization of a UAS-based image pipeline for scoring maturity dates in soybean. A simple approach using either local polynomial regression or segmented linear regression relating VIs capturing canopy greenness to date provided reliable estimations for date of soybean maturity in field trial plots. We found that flights could occur as infrequently as one time per week without loss in accuracy. The mean of all pixel values from a ROI worked equally well as when pixels were filtered for only those that have more central values, except in the case of specific plot issues that led to errors in the UAS-based maturity dates. In accordance with findings from previous studies, UAS-based maturity estimates were at least as precise as human visual dates and often more precise. An analysis of the causes of large deviations between human and UAS-based dates found that 46% of these large deviations were the result of human errors. Among the remaining 54%, within-plot genetic heterogeneity, lodging, presence of weeds, and low germination were important sources of errors in UAS-based dates. We have made our image and statistical analysis script publicly available, along with the data and images used in this study, to help others adopt these methods to their own programs.

ACKNOWLEDGMENTS

We acknowledge the United Soybean Board, Minnesota Soybean Research and Promotion Council, the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 for funding this research, and PepsiCo for partially funding the development of the script. In addition, we thank members of the Lorenz Lab at the University of Minnesota for helping to plant, manage field plots, and take field measurements. Special thanks for due to Thomas Hoverstad, Leonardo Moros, and Steve Quiring for collecting UAS imagery.

AUTHOR CONTRIBUTIONS

Leonardo Volpato: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing-original draft; Writing-review & editing. Austin Dobbs: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Visualization; Writing-original draft; Writing-review & editing. Aluízio Borem: Project administration; Visualization; Writing-review & editing. Aaron Joel Lorenz: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing-original draft; Writing-review & editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Leonardo Volpato https://orcid.org/0000-0003-1119-0615
Aaron Joel Lorenz https://orcid.org/0000-0002-4361-1683

REFERENCES

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. Trends in Plant Science, 23(5), 451–466. https://doi.org/10.1016/j.tplants.2018.02.001

Bai, G., Ge, Y., Hussain, W., Baenziger, P. S., & Graef, G. (2016). A multi-sensor system for high throughput field phenotyping in soybean and wheat breeding. Computers and Electronics in Agriculture, 128, 181–192. https://doi.org/10.1016/j.compag.2016.08.021

Belgu, M., & Stein, A. (2019). Spatiotemporal image fusion in remote sensing. Remote Sensing, 11(7),. https://doi.org/10.3390/rs11070818

Bivand, R., Keitt, T., & Rowlingson, B. (2020). Package ‘rgdal’. R. https://cran.r-project.org/web/packages/rgdal/index.html

Castellao Tetila, E., Brandoli Machado, B., Belete, N. A. D. e S., Guimarães, D. A., & Pistori, H. (2017). Identification of soybean
Haghhighatalab, A., González-Pérez, L., Mondal, S., Singh, D., Schinstock, D., Rutkoski, J., Ortiz-Monasterio, I., Singh, R. P., Goodin, D., & Poland, J. (2016). Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods, 12*(1). https://doi.org/10.1186/s13007-016-0134-6

Han, L., Yang, G., Yang, H., Xu, B.o, Li, Z., & Yang, X. (2018). Clustering field-based maize phenotyping of plant-height growth and canopy spectral dynamics using a UAV remote-sensing approach. *Frontiers in Plant Science,* 9. https://doi.org/10.3389/fpls.2018.01638

Hassan, M. A., Yang, M., Fu, L., Rasheed, A., Zheng, B., Xia, X., Xiao, Y., & He, Z. (2019). Accuracy assessment of plant height using an unmanned aerial vehicle for quantitative genomic analysis in bread wheat. *Plant Methods, 15*(1). https://doi.org/10.1186/s13007-019-0419-7

Hijmans, R. J. (2020). Package ‘raster’ R. https://cran.r-project.org/web/packages/raster/raster.pdf

Holman, F. H., Riche, A. B., Castle, M., Wooster, M. J., & Hawkesford, M. J. (2019). Radiometric calibration of “commercial offthe shelf” cameras for UAV-based high-resolution temporal crop phenotyping of reflectance and NDVI. *Remote Sensing, 11.* https://doi.org/10.3390/rs11141657

Holman, F., Riche, A., Michalski, A., Castle, M., Wooster, M., & Hawkesford, M. (2016). High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sensing, 8*(12). https://doi.org/10.3390/rs8121031

Huang, Y., Reddy, K. N., Fletcher, R. S., & Pennington, D. (2018). UAV low-altitude remote sensing for precision weed management. *Weed Technology, 32*(1), 2–6. https://doi.org/10.1017/wet.2017.89

Löwy, M. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment, 83*(2 Special Issue), 195–213. https://doi.org/10.1016/S0925-8574(02)00062-5

Hunt, E. R., Doraisswamy, P. C., Mcmurtrey, J. E., Daughtry, C. S. T., Perry, E. M., & Akhmedov, B. (2012). A visible band index for remote sensing leaf chlorophyll content at the Canopy scale. *International Journal of Applied Earth Observation and Geoinformation, 21*(1), 103–112. https://doi.org/10.1016/j.jag.2012.07.020

Hunt, E. R., Daughtry, C. S. T., Eitel, J. U. H., & Long, D. S. (2011). Remote sensing leaf chlorophyll content using a visible band index. *Agronomy Journal, 103*(4), 1090–1099. https://doi.org/10.2134/agronj2010.0395

Iqbal, F., Luceeer, A., & Barry, K. (2018). Simplified radiometric calibration for UAS-mounted multispectral sensor. *European Journal of Remote Sensing, 51*(1), 301–313. https://doi.org/10.1080/22797254.2018.1432293

Jarquin, D., Howard, R., Xavier, A., & Das Choudhury, S. (2018). Increasing predictive ability by modeling interactions between environments, genotype and canopy coverage image data for soybeans. *Agronomy, 8*(4). https://doi.org/10.3390/agronomy8040051

Jin, X., Liu, S., Baret, F., Hemerlé, M., & Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sensing of Environment, 198*, 105–114. https://doi.org/10.1016/j.rse.2017.06.007

Krause, M. R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R. P., Dreissigacker, S., Poland, J., Rutkoski, J., Sorrells, M., Gore, M. A., & Mondal, S. (2019). Hyperspectral reflectance-derived relationship matrices for genomic
prediction of grain yield in wheat. *G3: Genes, Genomes, Genetics*, 2008856. https://doi.org/10.1534/g3.118.200856

Liaw, A., & Wiener, M. (2018). Package ‘randomForest.’ https://www.stat.berkeley.edu/~breiman/RandomForests/

Lindsey, A. J., Craft, J. C., & Barker, D. J. (2020). Modeling canopy senescence to calculate soybean maturity date using NDVI. *Crop Science*, 60(1), 172–180. https://doi.org/10.2135/csc2.20079

Loladze, A., Rodrigues, F. A., Toledo, F., San Vicente, F., Gérard, B., & Boddupalli, M. P. (2019). Application of remote sensing for phenotyping tar spot complex resistance in maize. *Frontiers in Plant Science*, 10. https://doi.org/10.3389/fpls.2019.00552

Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, 16(1), 65–70. https://doi.org/10.1080/10106040108542184

Lorenz, A. J. (2020). Optimization of temporal UAS-based imagery analysis to estimate plant maturity date for soybean breeding. CyVerse Data Commons. https://doi.org/10.25739/bc4f-x382

Ma, B. L., Dwyer, L. M., Costa, C., Cober, E. R., & Morrison, M. J. (2001). Early prediction of soybean yield from canopy reflectance measurements. *Agronomy Journal*, 93(6), 1227–1234. https://doi.org/10.2134/agronj2001.1227

Madec, S., Baret, F., De Solan, B., Thomas, S., Dutartre, D., Jezequel, S., Hemmerlé, M., Colomeau, G., & Comar, A. (2017). High-throughput phenotyping of plant height: comparing unmanned aerial vehicles and ground LiDAR estimates. *Frontiers in Plant Science*, 8. https://doi.org/10.3389/fpls.2017.02002

Maes, W. H., & Steppe, K. (2019). Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends in Plant Science*, 24(2), 152–164. https://doi.org/10.1016/j.tplants.2018.11.007

Matias, F. I., Caraza-Harter, M. V., & Endelman, J. B. (2020). FIELDimageR: An R package to analyze orthomosaic images from agricultural field trials. *The Plant Phenome Journal*, 3, e20005; https://doi.org/10.1002/ppj2.20005

Mourtzinis, S., & Conley, S. P. (2017). Delineating soybean maturity groups across the United States. *Agronomy Journal*, 109(4), 1397–1403. https://doi.org/10.2134/agronj2016.10.0581

Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19), 3055–3071. https://doi.org/10.1002/sim.1545

Muggeo, V. M. R. (2019). Package ‘segmented.’ https://cran.r-project.org/web/packages/segmented/segmented.pdf

Narayanan, B., Floyd, B., Tu, K., Ries, L., & Hausmann, N. (2019). Improving soybean breeding using UAS measurements of physiological maturity. (May): 27. https://doi.org/10.11711/2519072

Peña, J., Torres-Sánchez, J., Serrano-Pérez, A., De Castro, A., & López-Granados, F. (2015). Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. *Sensors (Switzerland)*, 15(3), 5609–5626. https://doi.org/10.3390/s150305609

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., Pinto, F., Piner-Chavez, F. J., Poland, J., Rivera-Amedo, C., Saint Pierre, C., & Sukumaran, S. (2020). Breeder friendly phenotyping. *Plant Science*, 110396. https://doi.org/10.1016/j.plantsci.2019.110396

Saberioun, M. M., Amin, M. S. M., Anuar, A. R., Gholizadeh, A., Wayayok, A., & Khairunniza-Bejo, S. (2014). Assessment of rice leaf chlorophyll content using visible bands at different growth stages at both the leaf and canopy scale. *International Journal of Applied Earth Observation and Geoinformation*, 32(1), 35–45. https://doi.org/10.1016/j.jag.2014.03.018

Schirrmann, M., Hamdorf, A., Garz, A., Ustuyuzhanin, A., & Dammer, K. H. (2016). Estimating wheat biomass by combining image clustering with crop height. *Computers and Electronics in Agriculture*, 121, 374–384. https://doi.org/10.1016/j.compag.2016.01.007

Schoving, C., Stöckle, C. O., Colombet, C., Champolivier, L., Debaeke, P., & Maury, P. (2020). Combining simple phenotyping and photothermal algorithm for the prediction of soybean phenology: Application to a range of common cultivars grown in Europe. *Frontiers in Plant Science*, 10(1An.), 1–14. https://doi.org/10.3389/fpls.2019.01755

Schweiger, R., Kaufman, S., Laaksonen, R., Kleber, M. E., März, W., Eskin, E., Rosset, S., & Halperin, E. (2016). Fast and accurate construction of confidence intervals for heritability. *American Journal of Human Genetics*, 98(6), 1181–1192. https://doi.org/10.1016/j.ajhg.2016.04.016

Setiyono, T. D., Weiss, A., Specht, J., Bastidas, A. M., Cassman, K. G., & Dobermann, A. (2007). Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. *Field Crops Research*, 100(2–3), 257–271. https://doi.org/10.1016/j.fcr.2006.07.011

Singh, D., Wang, X.u, Kumar, U., Gao, L., Noor, M., Imtiaz, M., Singh, R. P., & Poland, J. (2019). High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. *Frontiers in Plant Science*, 10. https://doi.org/10.3389/fpls.2019.00394

Song, Y., & Wang, J. (2019). Winter wheat canopy height extraction from UAV-based point cloud data with a moving cuboid filter. *Remote Sensing*, 11, 10–14. https://doi.org/10.3390/rs11012329

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8. https://doi.org/10.1186/1471-2105-8-25

Thorp, K. R., & Tian, L. F. (2004). A review on remote sensing of weeds in agriculture. *Precision Agriculture*, 5, 477–508. https://doi.org/10.1007/s11119-004-5321-1

Torres-Sánchez, J., López-Granados, F., De Castro, A. I., & Peña-Barragán, J. M. (2013). Configuration and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. *PLoS One*, 8(3). https://doi.org/10.1371/journal.pone.0058210

Van Eeuwijk, F. A., Bustos-Korts, D., Millet, E. J., Boer, M. P., Kruijer, W., Thompson, A., Malosetti, M., Iwata, H., Quiroz, R., Kuppe, C., Muller, O., Blazakis, K. N., Yu, K., Tardieu, F., & Chapman, S. C. (2019). Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Science*, 282, 23–39. https://doi.org/10.1016/j.plantsci.2018.06.018

Wang, C., Nie, S., Xi, X., Luo, S., & Sun, X. (2017). Estimating the biomass of maize with hyperspectral and LiDAR data. *Remote Sensing*, 9(1). https://doi.org/10.3390/rs9010011

Wolak, M. (2019). Package nadjiv. https://cran.r-project.org/web/packages/nadjiv/nadjiv.pdf

Xu, K., Gong, Y., Fang, S., Wang, K.e, Lin, Z., & Wang, F. (2019). Radiometric calibration of UAV remote sensing image with spectral angle constraint. *Remote Sensing*, 11(11). https://doi.org/10.3390/rs1111291

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B.o, Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Z., Li, H., & Yang, H. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives.
Yu, N., Li, L., Schmitz, N., Tian, L. F., Greenberg, J. A., & Diers, B. W. (2016). Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. *Remote Sensing of Environment, 187*, 91–101. https://doi.org/10.1016/j.rse.2016.10.005

Zdziarski, A. D., Todeschini, M. H., Milioli, A. S., Woyann, L. G., Madureira, A., Stoco, M. G., & Benin, G. (2018). Key soybean maturity groups to increase grain yield in Brazil. *Crop Science, 58*(3), 1155–1165. https://doi.org/10.2135/cropsci2017.09.0581

Zhou, J., Yungbluth, D., Vong, C. N., Scaboo, A., & Zhou, J. (2019). Estimation of the maturity date of soybean breeding lines using UAV-based multispectral imagery. *Remote Sensing, 11*(18). https://doi.org/10.3390/rs11182075

**Supporting Information**
Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article**: Volpato, L., Dobbels, A., Borem, A., & Lorenz, A. J. Optimization of temporal UAS-based imagery analysis to estimate plant maturity date for soybean breeding. *Plant Phenome J*. 2021;4:e20018. https://doi.org/10.1002/ppj2.20018