

## Article

# Maturity Prediction in Soybean Breeding Using Aerial Images and the Random Forest Machine Learning Algorithm

Osvaldo Pérez <sup>1,2</sup>, Brian Diers <sup>1</sup> and Nicolas Martin <sup>1,\*</sup>

<sup>1</sup> Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA; omperez2@illinois.edu or operez@inia.org.uy (O.P.); bdiers@illinois.edu (B.D.)

<sup>2</sup> Instituto Nacional de Investigación Agropecuaria (INIA), Estación Experimental INIA La Estanzuela, Ruta 50 km 11, Colonia 70000, Uruguay

\* Correspondence: nfmartin@illinois.edu

**Abstract:** Several studies have used aerial images to predict physiological maturity (R8 stage) in soybeans (*Glycine max* (L.) Merr.). However, information for making predictions in the current growing season using models fitted in previous years is still necessary. Using the Random Forest machine learning algorithm and time series of RGB (red, green, blue) and multispectral images taken from a drone, this work aimed to study, in three breeding experiments of plant rows, how maturity predictions are impacted by a number of factors. These include the type of camera used, the number and time between flights, and whether models fitted with data obtained in one or more environments can be used to make accurate predictions in an independent environment. Applying principal component analysis (PCA), it was found that compared to the full set of 8–10 flights ( $R^2 = 0.91$ – $0.94$ ; RMSE = 1.8–1.3 days), using data from three to five flights before harvest had almost no effect on the prediction error (RMSE increase ~0.1 days). Similar prediction accuracy was achieved using either a multispectral or an affordable RGB camera, and the excess green index (*ExG*) was found to be the important feature in making predictions. Using a model trained with data from two previous years and using fielding notes from check cultivars planted in the test season, the R8 stage was predicted, in 2020, with an error of 2.1 days. Periodically adjusted models could help soybean breeding programs save time when characterizing the cycle length of thousands of plant rows each season.



**Citation:** Pérez, O.; Diers, B.; Martin, N. Maturity Prediction in Soybean Breeding Using Aerial Images and the Random Forest Machine Learning Algorithm. *Remote Sens.* **2024**, *16*, 4343. <https://doi.org/10.3390/rs16234343>

Academic Editors: Wei Su, Quanjun Jiao, Bo Liu, Xing Li and Qiaoyun Xie

Received: 27 September 2024  
Revised: 24 October 2024  
Accepted: 5 November 2024  
Published: 21 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** agriculture; plant breeding; high-throughput phenotyping; UAV; physiological maturity; vegetation indices; machine learning

## 1. Introduction

Soybean (*Glycine max* (L.) Merr.) breeding programs need to annually record phenotypic traits for thousands of experimental lines grown as plant rows to select those that should be evaluated in preliminary yield tests. However, much of the data collected are not used because most experimental lines will be discarded due to low grain yield or other undesirable traits, such as when the germplasm has a short or very long cycle length for a given environment. The cycle length is characterized by the date the plants reach their physiological maturity (R8 stage), which is a trait of superlative importance in plant breeding because it is associated with grain yield potential [1]. The R8 growth stage in soybean is defined as when 95% of pods reach their mature color and maximum biomass accumulation in the seed has occurred [2]. Taking maturity notes is time-consuming because it usually requires going to the field every three or four days for about four to five weeks.

As imagery obtained by drones has become readily available and affordable, plant breeders are more interested in applying high-throughput phenotyping (HTP) for classifying and predicting agronomic traits such as date of physiological maturity, pubescence color, lodging, plant biomass, and grain yield [3]. Applying HTP can help identify genomic regions associated with these and other traits of interest; this is by combining phenomics with genomics [4–6].

Compared to standard methods of recording field notes, with an HTP platform, a larger volume of information can be quickly collected, which could help breeders make selections with greater accuracy and efficiency. However, even when research on applying HTP to predict agronomic traits has been published, there has been difficulty in implementing these methods in breeding programs [7]. This difficulty is mainly due to the uncertainty about the validation and repeatability of the predictions and the challenges in processing the data in a short time when data science is still an emerging discipline.

Multi-rotors carrying cheap digital RGB (red, green, blue) cameras are the most common UAVs, while professionals often integrate more expensive multi or hyper-spectral cameras. Although multispectral cameras have a lower resolution than RGB cameras, a significant advantage is that their lenses can record images at frequencies beyond the spectrum visible to the human eye, such as the red edge and near-infrared (NIR) spectral bands. This means that more indices can be calculated with multispectral than with RGB cameras. Two examples are NDVI, which is calculated with the red and NIR spectral bands [8], and the normalized difference red edge (NDRE), calculated with the NIR and red edge bands [9].

Random forest (RF) is one of the most commonly used machine learning algorithms for conducting predictive analyses using an HTP platform [10]. Based on classification and regression trees (CART), which is another machine learning algorithm, RF is an ensemble of individual trees (i.e., the predictors) trained using the bootstrap aggregation method, also known as the bagging method [11,12]. Even with the risk of high bias, the aim of training different random subsets within the training set (i.e., bagging) is to decrease the correlations between trees and the variance, which can cause overfitting of the trained model [13]. Using RF and considering a binary prediction model for analyzing multispectral aerial images taken over plant rows at the University of Illinois, Yu et al. [14] reported an overall accuracy of ~93% for classifying R8 (mature or immature plots). Considering R8 as a classification variable but using RGB images instead, high overall accuracies (>90%) were also reported by other studies applying RF and other methods, including deep learning [15,16]. On the other hand, considering R8 instead as a regression variable, Trevisan et al. [17] and Moeinizade et al. [18] reported a root mean square error (RMSE) of approximately  $\pm 2$  days using RGB images and deep convolutional neural networks (CNN) to train the models.

Other studies to predict physiological maturity in soybean have been conducted using different methods and instruments. Using a ground-based field spectroradiometer to measure the canopy reflectance, Christenson et al. [19] applied partial least squares regression (PLSR) to associate the R8 stage with 91 spectral bands; 27 versions of four vegetation indices (VIs), RENDVI, and the blue, green, and red NDVIs, and 3 versions for the water index. After the PLSR analysis, one model was adjusted with the most significant indices and the other with the most significant spectral bands (RMSE = 5.51 and 5.19 days, respectively). Applying PLSR but using multispectral images taken from a drone, Zhou et al. [20] explained up to 70% of the R8 stage variation (RMSE = 1.7 days). Handheld crop sensors recording NDVI are another example [21], but as well as spectroradiometers, both have the disadvantage of collecting fewer records per unit of time compared to taking images from a drone or a manually operated cart [22].

Narayanan et al. [23] calculated a normalized green excess index and adjusted a piecewise linear regression model in function of time to forecast physiological maturity across 22 sites, achieving a Pearson correlation coefficient ( $r$ ) from 0.79 to 0.92 for a subset of data. Volpato et al. [24] later confirmed these findings but instead used a nonparametric local polynomial regression (LOESS) model ( $r = 0.84\text{--}0.97$ ). Meanwhile, the excess green index (ExG) has become a prominent vegetation index due to its effectiveness in enhancing the contrast between vegetation and soil pixels using only RGB bands for its calculation [25,26]. Yuan et al. [22] used RGB images and five regression and classification models to predict various traits, including physiological maturity, though they reported a lower accuracy ( $R^2 = 0.76$ , RMSE = 3.7 days) than the abovementioned studies.

On the other hand, Trevisan et al. [17] and Moeinizade et al. [18] highlighted the advantages of deep learning methods like CNN for predicting complex traits, such as plant development and plant stress [27]. However, Zhang et al. [28] found no significant accuracy differences among various learning methods for predicting lodging in wheat: support vector machine (SVM), RF, and three neural networks (GoogLeNet, CNN, and VGG-16). When no significant accuracy differences occur, machine learning algorithms like RF would have the advantage of having lower computational complexity than CNN [16].

Regarding complexity and its relationship to the number of features, Teodoro et al. [29] noted improved predictions using RF and deep learning compared to SVM and linear regression when predicting soybean maturity with RGB data. However, when additional variables (VIs) were included in the model (alone or with the RGB bands), the authors observed a prediction error increase when using deep learning, indicating that this method shows higher sensitivity to including redundant features compared to machine learning methods such as RF. This result highlights the importance of optimizing model architecture and variable selection to enhance predictive performance while minimizing overfitting.

A few studies have used RF and CNN to predict the R8 stage with an overall accuracy above 90% [14–16]; however, except for the study by Trevisan et al. [17], as far as we know, no other study tested their models in independent environments. Thus, the hypothesis is that soybean breeding programs could adjust scalable and repeatable models across the years. We also hypothesized that optimizing the drone flights and variable selection would enhance the predictive model performance while minimizing overfitting.

Applying the RF algorithm, we associated more than 32,000 field records across three experiments (2018–2020) with a time series of images taken from a UAV to improve the efficiency of characterizing physiological maturity in breeding populations (plant rows). The objectives were (1) to evaluate whether using a time series of multispectral images can predict the date the plant rows reached the R8 stage more accurately than RGB images; (2) to study how the reliability of predictions is impacted by including classification variables (check cultivars and germplasm relationships), the phenological stage of the plant rows when the images are taken, the number and time between flights, and what image features are the most informative in predictions; and (3) to test the consistency of the fitted models across years to determine if a training set obtained previously can make accurate predictions for an independent experiment or environment.

## 2. Materials and Methods

### 2.1. Plant Material and Experimental Design

Soybean experiments were conducted during three consecutive growing seasons in Savoy, IL, USA. Each experiment was planted nearby in fields under a corn–soybean rotation. The planting dates were 22 May 2018 and 3 June in both 2019 and 2020. The experiments comprised plots of F<sub>4:5</sub> experimental lines developed by the soybean breeding program at the University of Illinois at Urbana–Champaign. The lines were divided into five or six blocks of plant rows each year according to breeding program objectives unrelated to the cycle length (Table 1). Each experimental line was grown in a one-row nonreplicated plot (plant row). The experimental lines were grouped by cross combination with lines from the same cross, approximately 50 to 70 lines, arranged randomly and planted sequentially in a serpentine arrangement. One of five check cultivars, representing the cycle length range of the breeding germplasm (maturity groups II, III, and IV), was planted every 20 rows. The plant row plots measured 1.2 m in length, had a 0.76 m row spacing, and alleys of 1.1 m between ranges of rows.

**Table 1.** Three soybean breeding experiments containing trials of plant rows of F<sub>4.5</sub> experimental lines grown at the University of Illinois Research and Education Center near Savoy, IL. Time series of aerial images over the plant rows were taken from drones carrying a digital RGB camera (2018) or a multispectral camera (2019 and 2020). The number of plant rows associated with images is the difference between the rows in the field (plant rows) and the number of plant rows not included in the analyses (NA and wrong field notes). The number of breeding blocks per experiment, the individuals in the team that took ground-truth notes for physiological maturity (R8 stage), the number of ground control points (GCPs), the sum of aerial images aligned per year (in brackets, the range per flight date), and the orthomosaic resolution range between flights are also indicated.

	2018	2019	2020
Drone	Phantom 4 Pro (DJI, Shenzhen, China)	Inspire 2 (DJI, Shenzhen, China)	
Camera model	Built-in FC6310 (DJI, Shenzhen, China)	Altum DJI SkyPort kit (Micasense, Seattle, WA, USA)	
Spectral bands	RGB (red, green, blue)	R, G, B, red edge, and near-infrared	
Image resolution	5472 × 3648 pixels	2064 × 1544 pixels	
Flight dates <sup>1</sup>	22 and 28 August, 4, 10, 14, 18, 23, and 26 September	1, 10, 25, and 30 July, 6, 15, and 28 August, 10, 17, and 24 September	30 June, 15 July, 4, 18, and 31 August, 10, 14, 22, and 30 September
Flight heights	~80 m	~60 m, and ~40 m (10 and 25 July)	~40 m
Plant rows	9360	11,800	11,400
Plant rows associated with images	9252	11,742	11,197
Breeding blocks	5	6	7
Team taking notes in the field	3	5	5
Ground control points (GCPs)	18	18–17	14
Aligned images (range per date)	1141 (114–221)	3589 (243–570)	3496 (302–558)
Orthomosaic resolution range	2.2–2.5 cm pixel <sup>-1</sup>	1.8–2.7 cm pixel <sup>-1</sup>	1.6–1.9 cm pixel <sup>-1</sup>

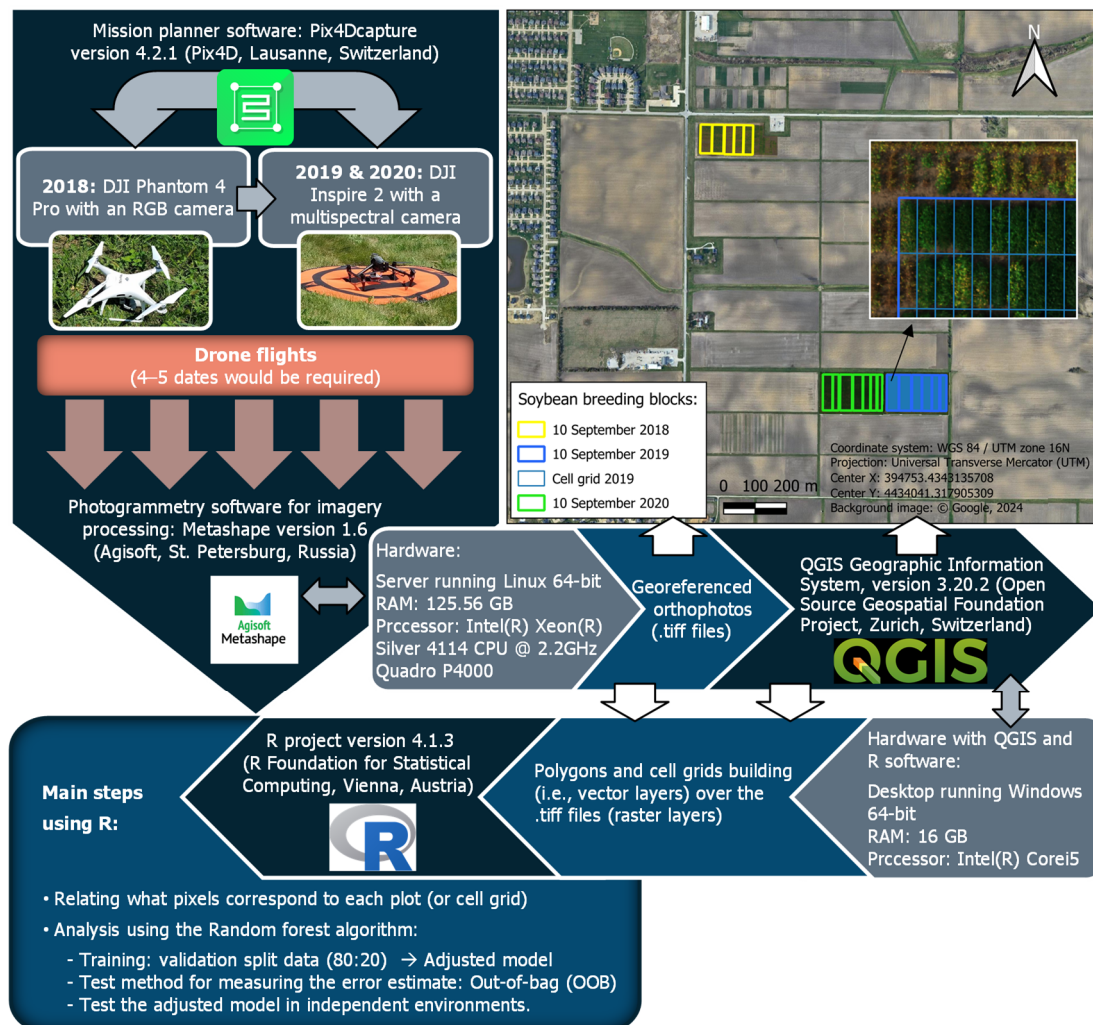
<sup>1</sup> The flight carried out on 25 July 2019 was completed using a Matrice 600 Pro (DJI, Shenzhen, China) drone with a RedEdge-M (Micasense, Seattle, WA, USA) multispectral camera mounted on a T1 gimbal kit (Gremsy, Ho Chi Minh, Vietnam). The model RedEdge-M also records five spectral bands with the same center wavelength as the Altum model, though this is with a lower image resolution (1280 × 960 pixels). Another difference is that the model RedEdge-M does not have a thermal sensor.

## 2.2. Data Collection

On 25 August 2018, 9 September 2019, and 17 September 2020, the plant rows started to be rated every 3–7 days for the date they matured (R8 stage). The field notes were recorded by a team that varied from three to five individuals annually, with each person assigned at least one breeding block (Table 1). In 2019 and 2020, a portion of replications in Block 1 decreased toward maturity (45 check plots each year) because they were destructively sampled for other studies.

RGB images were collected in 2018 with a Phantom 4 Pro (DJI, Shenzhen, China) drone with a built-in FC6310 (DJI, Shenzhen, China) digital camera. Except for one flight date, the multispectral images were collected in 2019 and 2020 with an Altum multispectral sensor DJI SkyPort kit (MicaSense, Seattle, USA) mounted on an Inspire 2 (DJI, Shenzhen, China) drone (Table 1; Figure 1). The Altum camera senses five discrete spectral bands: blue (475 nm), green (560 nm), red (668 nm), red edge (717 nm), and near-infrared (840 nm). The camera also has a thermal sensor, but these data were not used in this work. The flight date exception was on 25 July 2019, when a Matrice 600 Pro (DJI, Shenzhen, China) drone with a RedEdge-M (Micasense, Seattle, WA, USA) multispectral camera mounted on a T1 gimbal kit (Gremsy, Ho Chi Minh, Vietnam) was used. The RedEdge-M camera has a lower image resolution than the Altum model, but the discrete spectral bands are sensed at the same wavelengths. The built-in FC6310 digital camera used in 2018 has different peak responses for the shared spectral bands: red = 594 nm, green = 524 nm, and blue = 468 nm [30]. Finally, both Micasense models, Altum and RedEdge-M, included,

from the manufacturer, a downwelling light sensor and a calibrated reflectance panel for image calibration. Before and after each flight date, with both mentioned cameras, pictures were taken of each panel to calibrate the reflectance between flights. The UAV and its built-in camera used in 2018 do not include a light sensor and a calibration panel; therefore, the flight dates occurred only on cloudless, sunny days.



**Figure 1.** Pipeline workflow diagram of a high-throughput phenotyping platform for predicting soybean physiological maturity (R8 stage) of three breeding experiments (2018–2020) containing trials divided into plant rows of  $F_{4:5}$  experimental lines grown at the University of Illinois Research and Education Center near Savoy, IL. On the top right, overlapped on the satellite image, © Google, 2024 [31], three selected orthophotos corresponding to these experiments were taken from a drone on the same flight date (10 September). The colored polygons indicate the effective area of the soybean breeding blocks (trials) for which physiological maturity was predicted. The magnified orthophoto (10 September 2019) shows the cell grid that was used to associate the pixels within each cell to the day of the year in which the plant row reached the R8 stage.

The software Pix4Dcapture version 4.2.1 (Pix4D, Lausanne, Switzerland) was used to plan the missions of the three experiments. Of the possible options, the autopilot mission for flying over grid-type designs was chosen and an image overlap of 80% forward and 60% of side was selected. There were 16, 17, and 14 georeferenced ground control points (GCPs) placed around each group of trials during 2018, 2019, and 2020, respectively. In 2018, the flights started when plant rows showed initial signs of maturity and continued until almost all plots reached the R8 stage. In 2019 and 2020, the flights began during the

vegetative phase and continued through maturity. The number of flights and the difference in days between flights varied over the years (Table 1).

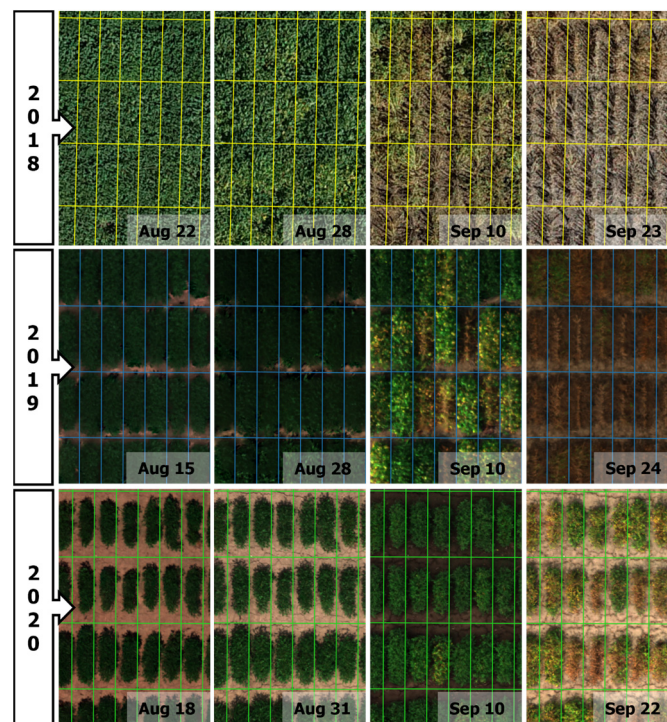
### 2.3. Image Processing

The time series of RGB (2018) and multispectral images (2019–2020) obtained from the three experiments were processed with the software Metashape version 1.6 (Agisoft, St. Petersburg, Russia). After uploading the images from each flight date, redundant images in the field corners and sides were discarded if the image overlapping of trials was not affected. The option ‘use sun sensor’ was not enabled in the software to calibrate the reflectance because the lighting of images was little or not affected by cloudiness during the fly missions ultimately used in this study (~10:00 AM–3:00 PM). For the images recorded in 2019 and 2020, the camera calibration function in Metashape was used to align the reflectance parameters, considering that both the Altum and the RedEdge-M cameras come with calibrated reflectance panels.

After calibration, the images were aligned for each flight date using the highest accuracy software setting. The procedure was repeated for some photos when they could not be aligned the first time. The camera calibration parameters of the first flight in each year were then saved in three .xml files. These files were uploaded for flights that followed during the same year (2018–2020) to adjust reflectance fluctuations due to different lighting conditions across the season. The next step was building dense point clouds setting quality and depth filtering as high and mild, respectively. To save time and space resources, the portion of the dense point clouds not belonging to the trials was selected and deleted. Based on these dense point clouds, a digital elevation model (DEM) for each flight date was generated to represent the surface, and from these models, the orthomosaics were generated.

The georeferenced GCPs were marked on the first orthomosaic for each year and saved as a .csv file (i.e., reference markers). The same GCPs were marked again for the following flight dates (2018–2020), but their geographical localization was realigned with the reference markers previously saved. To the extent the global positioning system (GPS) built-in the drones can yield slight localization differences for the respective GCPs, using the same reference markers for each year is a warranty that the imagery matches across the growing season without requiring the use of a real-time kinematic (RTK) GPS. Finally, with the function ‘enable hole filling’ disabled, orthomosaics were built and saved as .tiff files containing RGB (2018) or multispectral information (2019–2020). When this information in each orthophoto (.tiff file) was RGB values, it was saved as integer values from 0 to 255, and when it was multispectral values, it was saved as normalized data.

The orthophotos were uploaded to the software QGIS Geographic Information System version 3.20.2 (Open Source Geospatial Foundation Project, Zurich, Switzerland, <https://www.qgis.org>, accessed on 20 August 2021), where the correct overlapping between flight dates was visually checked. An orthophoto was then selected for each year (2018–2019) to draw a vector layer (.gpkg file) with a series of polygons representing the area of the breeding blocks (Figure 1). The areas in these three vector layers were used to extract the raster area of interest (i.e., .tiff files with the breeding block pixels). Then, another three vector layers were created using the software R version 4.1.3 (R Foundation for Statistical Computing, Vienna, Austria, <https://www.r-project.org>, accessed on 15 March 2022) and the package ‘sf’ [32] was used to divide the polygons into a grid of cells; this was conducted according to the number of rows and columns per breeding block previously indicated in QGIS with the tool ‘field calculator’. Hence, the pixels in the raster layer assigned by each grid cell represent the values per plant row, which is the plot area of a unique experimental line or a check cultivar replication (Figure 2). The resulting grid from each year was then used to overlap the raster information of the other flight dates within that year.



**Figure 2.** Partial visualization of composed orthophotos obtained from time series of images taken from a drone flying over three soybean breeding experiments (2018–2020). The experiments, containing plant rows of  $F_{4.5}$  experimental lines, were grown at the University of Illinois Research and Education Center near Savoy, IL. The imagery was collected in a total of eight flight dates in 2018, ten in 2019, and nine in 2020, although only four flight dates per year are shown according to the best matching day of the year. The raster information within each cell grid was used to predict the day of the year the plant row reached physiological maturity. All the orthophotos show the three visual spectral bands (red, green, and blue); however, while the images were taken with a digital RGB camera in 2018, in 2019 and 2020, they were with a multispectral camera of five bands: red, green, blue, red edge, and near-infrared.

#### 2.4. Data Analysis Methods

The vector and raster information was read, labeled, and joined using the software R version 4.1.3 with the packages ‘rgdal’ [33], ‘raster’ [34], ‘sf’ [32], and ‘fasterize’ [35]. After the raster was read, a new image feature was created in R through calculating, for each pixel, the vegetation index  $ExG^1$  proposed by Woebbecke et al. [25]:

$$\text{Excess green index (ExG)} = 2g - r - b \quad (1)$$

where  $r$ ,  $g$ , and  $b$  are normalized values of the bands red, green, and blue, respectively; this is,

$$r = \frac{R_{nzed}}{R_{nzed} + G_{nzed} + B_{nzed}} \quad (2)$$

$$g = \frac{G_{nzed}}{R_{nzed} + G_{nzed} + B_{nzed}} \quad (3)$$

$$b = \frac{B_{nzed}}{R_{nzed} + G_{nzed} + B_{nzed}} \quad (4)$$

In turn,  $R_{nzed}$ ,  $G_{nzed}$ , and  $B_{nzed}$  are calculated as  $R/R_{\text{sat}}$ ,  $G/G_{\text{sat}}$ , and  $B/B_{\text{sat}}$ , where the numerators are the band values of the orthophotos (from 0 to 255) and the denominators their saturated values (255). This index combines the subtraction of the green from the red chromatic coordinate ( $r-g$ ) and the blue from the green chromatic coordinate ( $g-b$ ). Hence, pixels with higher index values indicate an excess of green in the image scenario.

In this manner, four image features were obtained from the digital 8-bit RGB imagery collected in 2018 (RGB and *ExG*). From the 16-bit multispectral imagery collected in 2019–2020 with both Micasense camera models, six image features were obtained (red, green, blue, red edge, near-infrared, and *ExG*). This resulted in the spectral bands that are in common between years (red, green, and blue) having different scale values according to the kind of data and how they were saved: 8-bit RGB (2018) and 16-bit multispectral saved as normalized data (2019–2020). In all cases, the resulting values of the *ExG* index varied between 0 and 1.

The median of all pixels per cell grid for each image feature was calculated when the pixel values were equal to zero or greater. After this, the boundaries of the cell grids were read (package *sf*) and each identification label was associated (package *fasterize*) with the respective pixel medians per plant row in each one of the three first flight dates (2018–2020). The joined information was saved in a data table (.csv file) and a new vector layer (.gpkg file). These two files were also generated for the following flight dates but always using the first same cell grid boundaries of the respective experiment.

#### 2.4.1. Data Preparation and Prediction Model

The .csv files were joined and longitudinally arranged to conduct an image time series analysis to predict soybean physiological maturity for each of the three experiments (2018–2020). In the next step, records wrongly annotated by the team in the field were removed, including values entirely out of range (vegetative stage or after harvest) or with an incorrect number of digits. Next, the package ‘*caret*’ [36] was used to randomly divide the data into 80:20 training and test data subsets, and the package ‘*randomForest*’ [37] to run the RF machine learning algorithm [12] and training the model by associating the image features to the response variable (date of R8). Finally, the trained model was tested in the data subset to predict the date the plant rows reached the R8 stage.

#### 2.4.2. Resampling Methods and Hyperparameter Tuning

Three resampling methods of estimating the error of predictions were studied in the experiment 2018 to tune the RF algorithm: 10-fold cross-validation (‘*cv*’), repeated 10-fold cross-validation (‘*repeatedcv*’) with three repeats, and out-of-bag (‘*oob*’) bootstrap samples. The algorithm ran faster with ‘*oob*’, especially compared to ‘*repeatedcv*’, and without affecting the overall prediction accuracy, so ‘*oob*’ was the final method used to train the models and predict physiological maturity in the plant rows for the three experiments (2018–2020). The split length of the number of randomly selected *p*-predictors (‘*mtry*’) was set by default,  $mtry = p^{1/2}$  for categorical and  $mtry = p/3$  for numerical data. The number of trees (‘*ntree*’) and the node size (‘*nodesize*’) were also set by default, *ntree* = 500 and *nodesize* = 1 when the R8 stage was considered a categorical variable, and *ntree* = 500 and *nodesize* = 5 when it was considered a numerical variable.

#### 2.4.3. Predictions Using Categorical or Numerical Data for the Response Variable

A second analysis (also using the RF algorithm) compared predictions by categorical or numerical setting of the response variable. When the R8 stage was predicted as a categorical variable, the field records indicating the day of the year (DOY) the plant rows reached maturity were converted to factor levels. Thus, a classification analysis was made, and the overall prediction accuracy and Cohen’s Kappa value were calculated. After this, the predicted classes were converted to the date format (i.e., numerical values) so that the indicators  $R^2$  and RMSE were also calculated for the categorical predictions, making the comparison with the numerical predictions possible. When the R8 stage was predicted as a numerical variable, DOY was converted to the date format and then to numerical data to be handled by RF as a regression analysis. A regression analysis applies better than a classification analysis because physiological maturity is a continuous process intimately related to plant senescence at the end of ontogeny.



After tuning the RF algorithm with one of the three resampling methods of estimating the error (cv, repeatedcv, or oob) and setting the response variable (categorical or numerical), the three objectives were answered by conducting the following analyses:

#### 2.4.4. Predictions Using Multispectral Images

The analyses were divided into two model scenarios: (1) assume a breeding program has a drone carrying only a digital RGB camera, so the image features for the three experiments included the *ExG* index and only the spectral bands red, green, and blue (2018–2020); and (2) assume a breeding program has a multispectral camera, so the images features also included the red edge and near-infrared spectral bands (2019–2020).

#### 2.4.5. Impact on Predictions by Including Classification Variables in the Model, Plant Rows That Mature After the Last Drone Flight, or Redundant Information from Flights or Images

The scenarios described in Section 2.4.4. were divided according to the inclusion of five classification variables in the models. These variables are (1) the breeding block; (2) the person who annotated the R8 stage to identify if there are biases when different team members take notes in the field; (3) the check cultivar; (4) the  $F_{4.5}$  population the plant rows belong to; and (5) the parental lines from which each  $F_{4.5}$  population was developed. Histograms of the observed physiological maturity distribution and boxplots showing prediction biases, depending on which individual took the field notes, were produced.

Another analysis was conducted to compare two subsets of plant rows according to whether they reached maturity before or after the last drone flight (26, 24 and 30 September, from 2018 to 2020, respectively). These models were trained using the red, green, and blue spectral bands and three classification variables (breeding block, the individual who took the field notes, and the check cultivar).

Two methods were used to study what drone flights and image features were best associated with predicting the R8 stage in soybean: (1) multivariate adaptive regression spline (MARS) to identify the 15 most relevant predictive variables; this is the variable importance measure that typically is conducted after running RF [36,38]; and (2) principal component analysis (PCA) to identify redundant variables and study how the loadings of the image features are associated with the variation of the response variable (the R8 stage). A principal component regression (PCR) analysis was also conducted to compare the results with the RF predictions. According to the PCA results, some image features were discarded to refit the model with those with higher loadings because of their higher association with the R8 stage (i.e., dimensionality reduction).

#### 2.4.6. Are Models Reliable When Tested in an Independent Environment?

A final analysis, using RF, was conducted to evaluate the reliability of the fitted models for predicting physiological maturity when they are tested in an independent environment. The independent test environment included related germplasm (from the University of Illinois soybean breeding program) but grown in a different year. Four different data sets were used to train models that were then tested on the data set from the experiment carried out in 2020: (1) a training data set including field notes collected only in 2019 (Training 2019); (2) a training data set including field notes collected in 2018 and 2019 (Training 2018–2019); (3) the same as Training 2019 but including the records collected from the check cultivars planted in 2020 (Training 2019<sub>plus 2020 checks</sub>); and (4) the same as Training 2018–2019 but including the same mentioned checks (Training 2018–2019<sub>plus 2020 checks</sub>). From the above, the resulting ratio of training:test data for the four models varied as follows: (1) 51:49 (Training 2019:Test 2020); (2) 65:35 (Training 2018–2019:Test 2020); (3) 53:47 (Training 2019<sub>plus 2020 checks</sub>:Test 2020<sub>without checks</sub>); and (4) 67:33 (Training 2018–2019<sub>plus 2020 checks</sub>:Test 2020<sub>without checks</sub>). The time series of images between years was arranged according to the best possible matching dates that the drone flights were carried out.

### 3. Results

#### 3.1. Predictions Using Three Resampling Methods with Categorical or Numerical Data

The comparison between predicting the date of physiological maturity (R8 stage) as a categorical or numerical variable indicates that when this variable was considered numerical, the models explained a greater proportion of the variation ( $R^2$ ) with a lower prediction error (RMSE). This lower error was consistent using three different resampling methods ('cv', 'repeatedcv', and 'oob') to train the models (Table 2). The fact that Cohen's Kappa values ((Agreements – Agreements by chance)/(Test data – Agreement by chance)) decrease relatively little with respect to the overall accuracies ((Total Agreements/Test data)  $\times$  100) indicates that most of the agreements are not due by chance, so the overall accuracy values are reliable. The sum of both agreements is each paired matching between observed and predicted dates (i.e., the diagonal of the confusion matrix).

**Table 2.** Prediction results of soybean physiological maturity (R8 stage) on test data (20%) from a total of 9252 plant rows of F<sub>4:5</sub> experimental lines grown at the University of Illinois Research and Education Center near Savoy, IL (2018). The predictive analysis was conducted using the Random Forest algorithm, where two variable classes and three resampling methods were tested to train the model (80%). The two variable classes were: (1) the date when the plant rows reached the R8 stage (categorical variable); and (2) numerical integer values representing the day of the year in which the same event happened (numerical variable). The field notes were associated with features classified by the breeding block obtained from time series of aerial RGB (red, green, blue) images taken from a drone (eight flight dates). The excess green index (*ExG*) was calculated with the RGB bands and included in the models as another image feature.

Resampling Method	Variable Class					
	Categorical <sup>1</sup>			Numerical		
	Overall Accuracy (%)	Kappa	$R^2$	RMSE ( $\pm$ Days)	$R^2$	RMSE ( $\pm$ Days)
10-fold cross-validation ('cv')	41.46	0.3633	0.8837	2.069	0.9137	1.773
Repeated 10-fold cross-validation with three repeats ('repeatedcv')	41.03	0.3585	0.8830	2.077	0.9138	1.773
Out-of-bag bootstrap samples ('oob')	40.21	0.3494	0.8832	2.072	0.9132	1.778

<sup>1</sup> In this case, after running the Random Forest algorithm, the categorical values were converted to numerical values to calculate a pseudo coefficient of regression ( $R^2$ ) and a pseudo root mean squared error (RMSE). Thus, both indicators are comparable with the results obtained with the variable analyzed as a numerical class.

The overall accuracy for the predicted dates matching the observed R8 date is 40.2% (using the 'oob' method). However, if one or two days are instead assumed as an acceptable error range, the overall prediction accuracy increases to 52.9% and 88.3%, respectively. Hence, if the latter error range ( $\pm 2$  days) was *a priori* assumed, it would be equivalent to the extended error assumed when researchers calculate the RMSE, 2.072 days for the mentioned example (Table 2). The three resampling methods show similar prediction accuracies, but applying the method 'oob' uses less time for data processing, especially compared to the 'repeatedcv' method and when the R8 stage is assumed to be a numerical variable (i.e., ~8 vs. 150 min each time the code was run). When the trade-off efficiency vs. best predictive indicator values ( $R^2$  and RMSE) is considered, the resampling method 'oob' with the R8 stage as a numerical variable was the most effective and efficient in predicting the R8 stage, instead of a classification model as previously used by Yu et al. [14].

#### 3.2. Predictions Including Classification Variables Using RGB or Multispectral Images

The models adjusted with the RF algorithm explained a higher proportion of the variation in 2019 compared to 2018 and 2020; however, the error was lower in 2020, particularly compared to 2018 (Table 3). When a classification variable for individuals who took notes in the field was included in the model, higher  $R^2$  and lower RMSE values were observed in all

cases, indicating a bias among individuals assigning the date that the plant rows reached the R8 stage. There was no advantage in using multispectral images compared to RGB images, indicating that the red edge and the near-infrared bands provide redundant information in predicting soybean physiological maturity (the R8 stage) when the vegetation index *ExG* is included among the images' features. This shows that affordable cameras can be as successful in predicting the R8 stage in soybeans as expensive multispectral cameras.

**Table 3.** Prediction results of soybean physiological maturity (R8 stage) on test data (20%) from a total of 9252, 11,742, and 11,197 plant rows of  $F_{4:5}$  experimental lines grown at the University of Illinois Research and Education Center near Savoy, IL (2018–2020). The predictive analysis was conducted using the Random Forest algorithm where the field notes were associated with the breeding block, the individual who took the field notes, and features obtained from time series of aerial images (8, 10, and 9 drone flight dates, respectively, for 2018, 2019, and 2020).

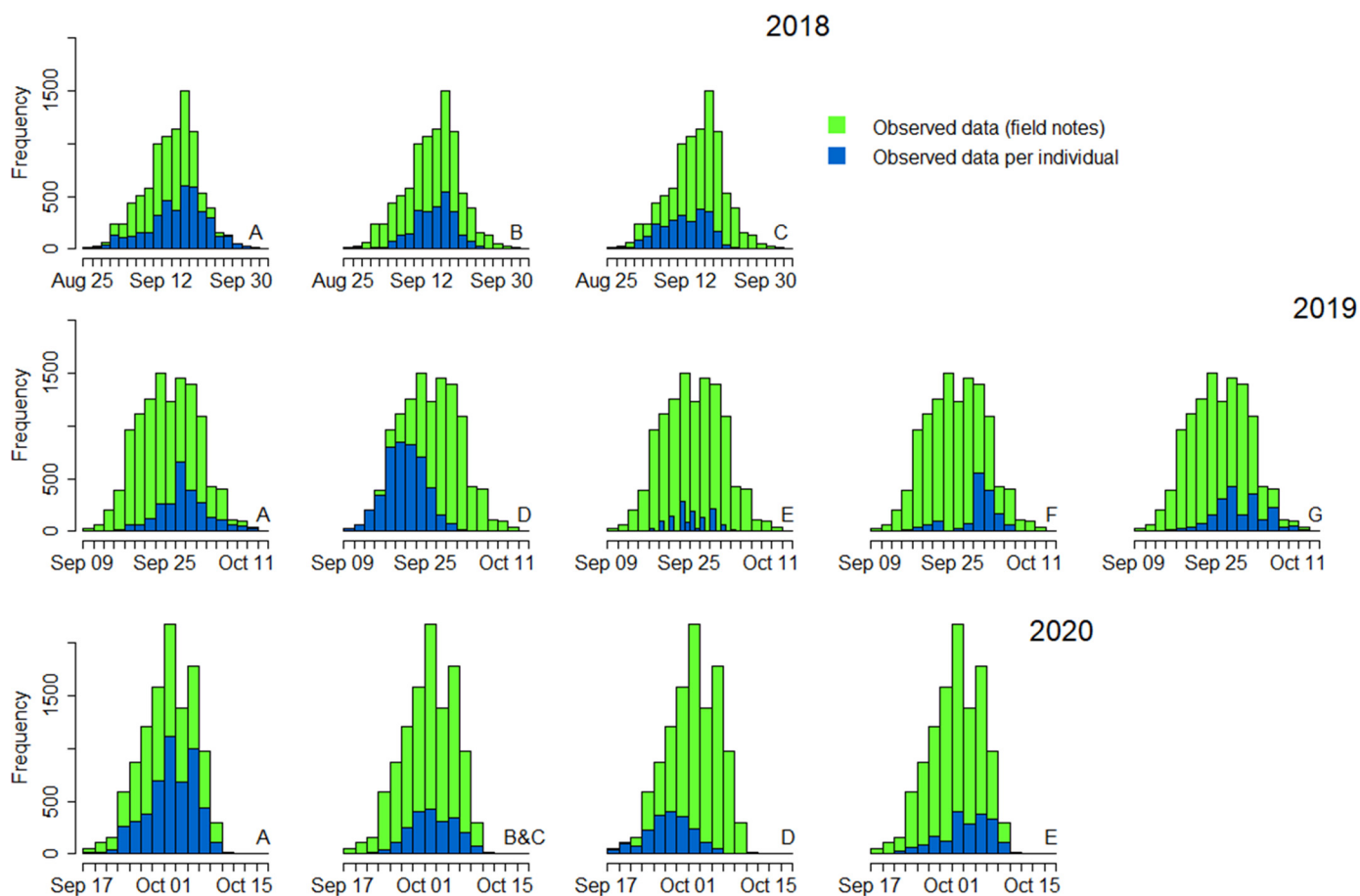
Year <sup>1</sup>	Camera	Image Features <sup>2</sup>	Explanatory and Classification Variables			
			Image Features and Breeding Block		Image Features, Breeding Block, and the Individual Who Took the Field Notes	
			$R^2$	RMSE ( $\pm$ Days)	$R^2$	RMSE ( $\pm$ Days)
2018	Digital RGB	Red, green, blue	0.9132	1.778	0.9141	1.767
2019	Multispectral	(Only the bands) red, green, blue	0.9353	1.448	0.9380	1.417
		Five bands	0.9354	1.448	0.9366	1.435
2020	Multispectral	(Only the bands) red, green, blue	0.9049	1.370	0.9080	1.347
		Five bands	0.9078	1.350	0.9111	1.325

<sup>1</sup> For the latter two years, the predictive analyses were divided into (1) using the five bands: red, green, blue, red edge, and near-infrared (normalized data); and (2) using only the red, green, and blue bands (values from 0 to 255) to compare the results with those obtained in 2018 (only RGB images). <sup>2</sup> For all scenarios, the excess green index (*ExG*) was calculated with the RGB bands and included in the models as another image feature.

The predictions of the R8 date were affected by the individual who took the field notes, the number of plant rows assigned per individual, and the fact that these plant rows were assigned by breeding block and not randomly when the breeding blocks had plant rows that reached the R8 stage with a different distribution (Figures 3 and 4). To the extent that these predictions were made using data collected by each individual and then tested on data collected by the other individuals of the team, biases between observed and predicted dates in which the plant rows reached their physiological maturity, expressed in days, were observed (Figure 4). Because experiments with  $F_{4:5}$  populations typically include thousands of plant rows, breeders need to split the task of taking notes among several individuals. In 2018, this task was given to three individuals (A–C) who had the same 'eye' to assign the R8 stage in soybeans, as no systematic bias was revealed by the individual who took the notes. In 2019, B and C did not take field notes, while in 2020, the field notes taken by both individuals (B and C) could not be distinguished because they shared the same two breeding trials. There was a slight overall bias in days among the five team individuals in 2020. However, in 2019, biases between individuals were noticeably higher due to F and G, who dated plant row maturity more than three days after the values predicted using the data collected by the other three individuals (A, D, and E). As such, these results suggest that breeding programs developing predictive models should consider including a classification variable to account for the individual taking the field notes (Table 3).

Compared to including only the breeding block and the individual who took the field notes in the predictive models (Table 3), a slight improvement in indicators  $R^2$  and RMSE for 2018 and 2020 is observed after including the check cultivar in a stacked manner (Table 4). When the  $F_{4:5}$  population and the female and male parents of lines were also included as classification variables, a consistent but negligible improvement was observed for the three years (2018–2020). Improved prediction using information from repeated

check cultivars was also observed by Trevisan et al. [17], who trained the models with a deep learning method (CNN) instead of the RF algorithm.



**Figure 3.** The histograms (in green) show the distribution of soybean physiological maturity (R8 stage) dates for three experiments of plant rows of  $F_{4.5}$  experimental lines grown at the University of Illinois Research and Education Center near Savoy, IL (2018–2020). The histograms (in blue) also show the distribution of the R8 stage dates, but according to what plant rows were assigned per individual (A–F) to take the field notes.

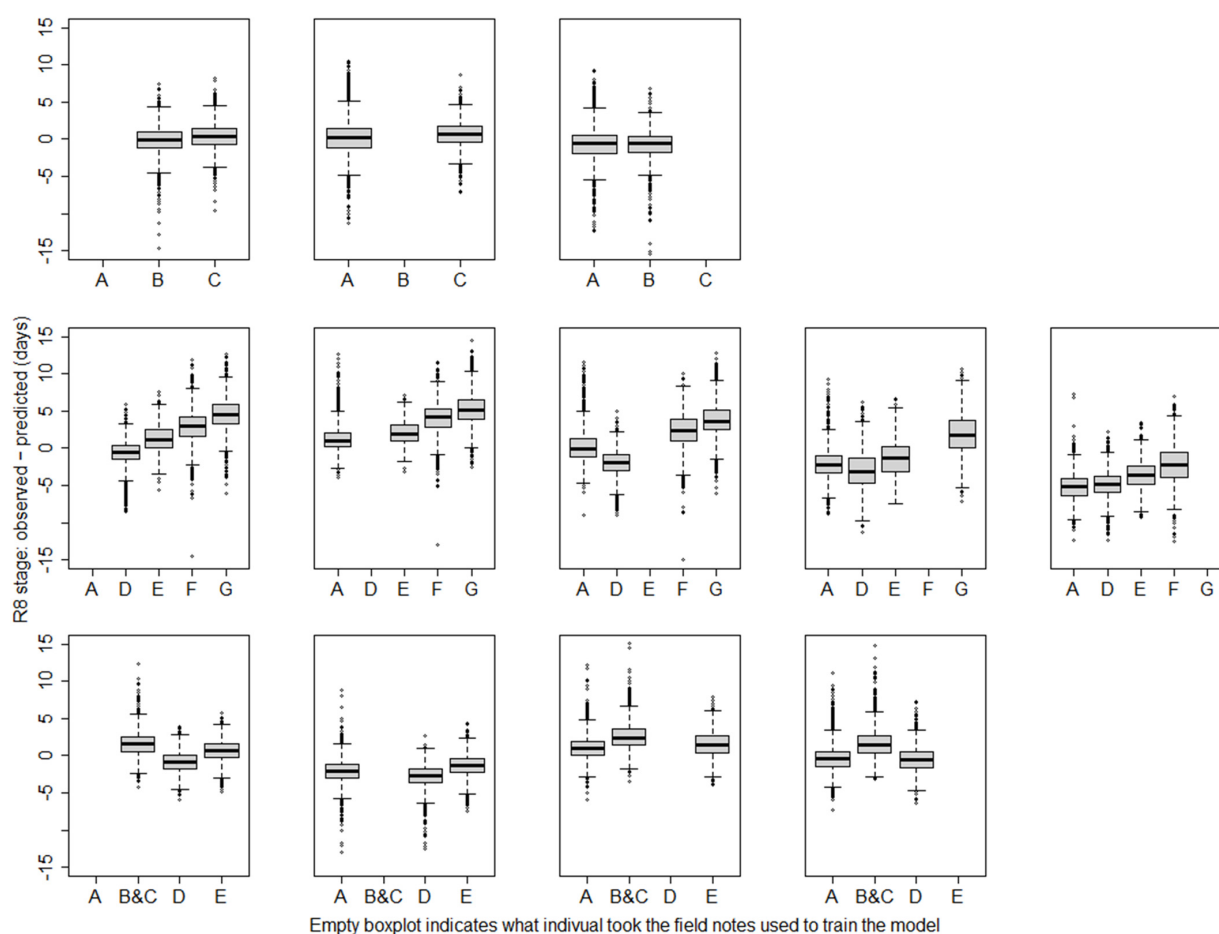
**Table 4.** Prediction results of soybean physiological maturity (R8 stage) on test data (20%) from a total of 9252, 11,742, and 11,197 plant rows of  $F_{4.5}$  experimental lines grown at the University of Illinois Research and Education Center near Savoy, IL (2018–2020). The predictive analysis was conducted using the Random Forest algorithm where the field notes were associated with germplasm information keeping other variables in common: breeding block, the individual who took the field notes, and features obtained from time series of aerial images (8, 10, and 9 drone flight dates, respectively, for 2018, 2019, and 2020).

Year <sup>1</sup>	Camera	Image Features <sup>2</sup>	Germplasm Information Including Variables in Common					
			Check Cultivar		Check Cultivar and $F_{4.5}$ Population		Check Cultivar, $F_{4.5}$ Population, and Parental Lines	
			$R^2$	RMSE ( $\pm$ Days)	$R^2$	RMSE ( $\pm$ Days)	$R^2$	RMSE ( $\pm$ Days)
2018	Digital RGB	Red, green, blue	0.9175	1.732	0.9191	1.714	0.9199	1.706
2019	Multispectral	(Only the bands) red, green, blue	0.9380	1.417	0.9390	1.406	0.9396	1.399
		Normalized data	0.9365	1.435	0.9375	1.425	0.9398	1.397

Table 4. Cont.

Year <sup>1</sup>	Camera	Image Features <sup>2</sup>	Germplasm Information Including Variables in Common					
			Check Cultivar		Check Cultivar and F <sub>4:5</sub> Population		Check Cultivar, F <sub>4:5</sub> Population, and Parental Lines	
			R <sup>2</sup>	RMSE (±Days)	R <sup>2</sup>	RMSE (±Days)	R <sup>2</sup>	RMSE (±Days)
2020	Multispectral	(Only the bands) red, green, blue	0.9087	1.342	0.9097	1.335	0.9107	1.327
		Normalized data	0.9114	1.323	0.9121	1.317	0.9126	1.314

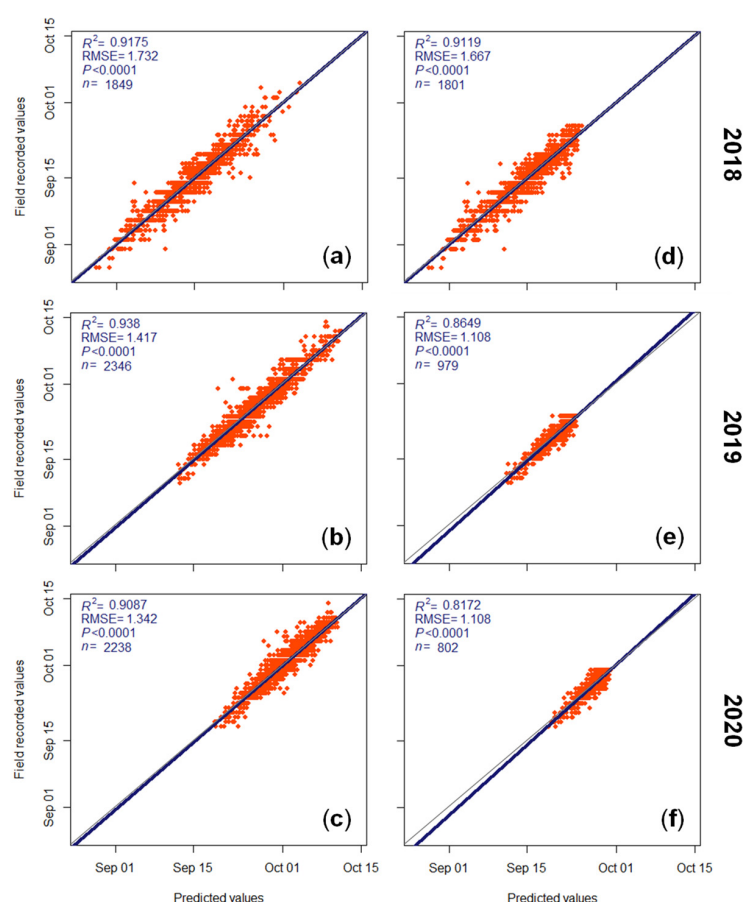
<sup>1</sup> For the latter two years, the predictive analyses were divided into (1) using the five bands: red, green, blue, red edge, and near-infrared (normalized data); and (2) using only the red, green, and blue bands (values from 0 to 255) to compare the results with those obtained in 2018 (only RGB images). <sup>2</sup> For all scenarios, the excess green index (ExG) was calculated with the RGB bands and included in the models as another image feature.



**Figure 4.** The boxplots show the bias of predictions (days) for soybean physiological maturity (R8 stage) according to the individuals (A–F) who together took 9252, 11,742, and 11,197 field notes from three experiments: 2018 (top), 2019 (middle), and 2020 (bottom), respectively. The experiments contained plant rows of F<sub>4:5</sub> experimental lines grown at the University of Illinois Research and Education Center near Savoy, IL. The Random Forest algorithm was used to adjust the predictive models using different training data sizes according to what plant rows were assigned per individual (A–F). The empty boxplot spaces mean that 44.2%, 28.5%, and 27.2% of field notes, taken respectively by A, B, and C, were used to train the models in 2018. In 2019, the proportions were 21.2%, 37.9%, 11.1%, 12.8%, and 17.0% (A, D–G); and in 2020, they were 45.3%, 19.6%, 17.5%, and 17.7% (A, B and C, D, and E).

### 3.3. Predictions Using Plant Rows That Mature Before or After the Last Drone Flight

In 2018, 239 out of 9252 plant rows reached the R8 stage after the last drone flight date (26 September). The comparison between predicting the R8 stage using all data (Table 4) and using only the plant rows that reached their maturity before 26 September resulted in only a small change in the prediction error (RMSE), from 1.73 days using all data (9252 plant rows) to 1.67 days (Figure 5). Conversely, disregarding these records from the distribution's right tail caused the model to explain a slightly lower proportion of the variation ( $R^2 = 0.92$  vs.  $R^2 = 0.91$ ). The above differences were greater in 2019 and 2020 because drone flights could not be carried out late in the season, resulting in many more records being omitted from the analyses. After removal, 6838 plant rows were used out of 11,742 in 2019 and 7197 out of 11,197 in 2020. Consequently, after disregarding the late plant rows, the errors decreased from 1.42 to 1.11 days in 2019 and from 1.34 to 1.11 days in 2020. Despite this, and as happened in 2018, the goodness of fit of the predictions also decreased in these two years (2019 and 2020).



**Figure 5.** Soybean physiological maturity (R8 stage) predictions corresponding to three breeding experiments containing plant rows of  $F_{4:5}$  experimental lines grown at the University of Illinois Research and Education Center near Savoy, IL (2018–2020). The Random Forest algorithm was applied to associate the field recorded values with three classification variables (breeding block, the individual who took the field notes, and the check cultivar) and 32 image features (red, green, blue, and a calculated excess green index—*ExG*—) obtained from eight drone flights. (a–c) The relationship between predicted vs. field recorded values using all the field records, and (d–f) the same, but after filtering records of plant rows that reached the R8 stage after the last drone flight date (26, 24, and 30 September, respectively, for 2018, 2019, and 2020). An equal relationship training:test data ratio (80:20) was maintained for the three experiments ( $n$  = test data). The deviation of the regression line (blue) from the 1:1 line (gray) indicates the model's prediction bias.

Additional predictive analyses were conducted by selecting only the plant rows that matured after the last drone flight to determine if these predictions had a high error. The analysis for 2018 was omitted because only 239 plant rows matured after the last drone flight, as previously mentioned. Surprisingly, in 2019 and 2020, the predictions selecting the later plant rows ( $R^2 = 0.88$  and  $0.82$ ;  $RMSE = 1.28$  and  $1.10$  days) almost did not differ from the predictions of the earlier plant rows (Figure 5e,f). This shows that changes in the image features before the plant rows reached their maturity can be used to predict R8 (Figure 5a–c).

### 3.4. Predictions After Discarding Redundant Information Using MAS and PCA

The variable importance measure of the 15 most relevant predictive variables indicates that for predicting the R8 stage in soybean, ExG was the most important (Figure 6). The variable importance measure, expressed as the cumulative percentage decrease of the generalized cross-validation (GCV) estimate of error when all the predictors are included in the model, is calculated using the backward pass method, subtracting one by one from the least to the most important predictive variable [36,38]. Using RGB data, the second most important predictive variable was the red spectral band, while using multispectral data, the most important were the red and the near-infrared bands. Still, in 2019 and 2020, the multispectral images with the two extra bands, the near-infrared and the red edge, did not improve the maturity predictions compared to using RGB images (Table 3).

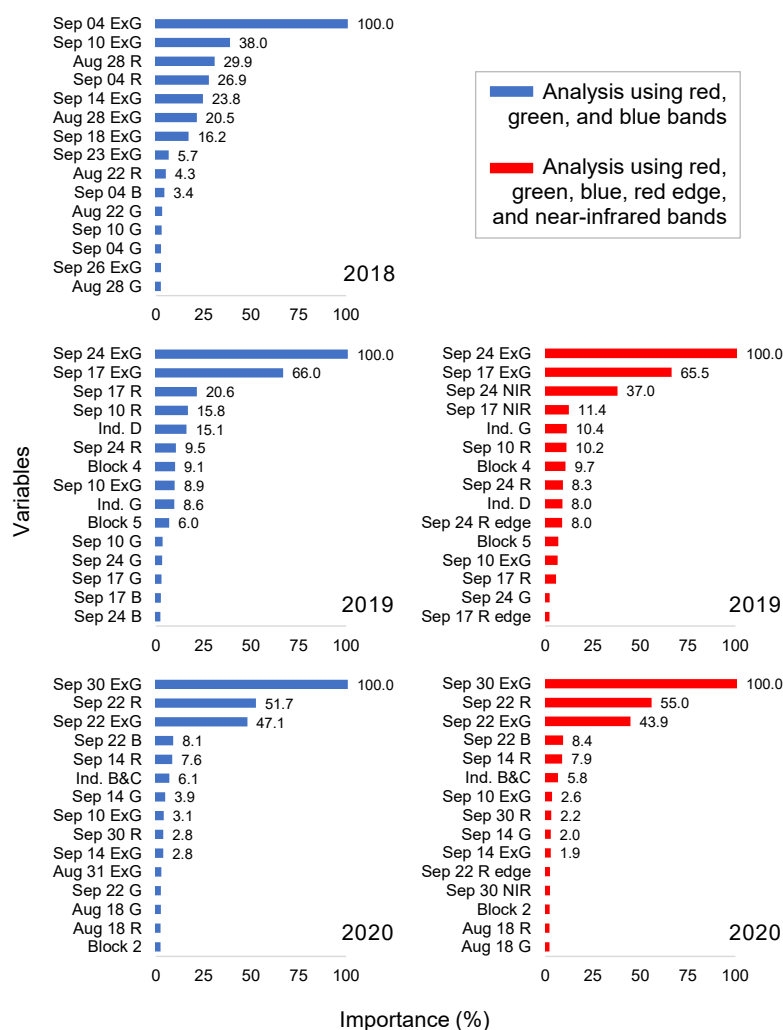


Figure 6. Variable importance measure of 15 most relevant variables for predicting soybean physiological maturity (R8 stage) of three experiments containing plant rows of F<sub>4.5</sub> experimental lines grown

at the University of Illinois Research and Education Center near Savoy, IL. Spectral bands extracted from time series of images taken from a drone and the excess green index (*ExG*) were included in the models as explanatory variables with three other classification variables: the breeding block (Block), the individual who took the field notes (Ind.), and the check cultivar (that does not show relevant importance). In 2018, the images were taken from a drone with a digital RGB (red, green, blue) camera, whereas in 2019 and 2020, they were taken with a multispectral camera. For the latter two years, the analyses were divided into using only the red (R), green (G), and blue (B) bands (simulating a digital RGB camera) and using the five spectral bands: R, G, B, R edge, and near-infrared (NIR).

The latter two drone flight dates in both years (17 and 24 September 2019 and 22 and 30 September 2020) provided better predictive information than the earlier flight dates (Figure 6). This is related to the fact that the mean date that the plant rows reached the R8 stage for both experiments was 26 September ( $n = 11,742$ ;  $sd = 5.7$  days) in 2019 and 2 October ( $n = 11,197$ ;  $sd = 4.5$  days) in 2020. Meanwhile, the mean date of maturity for plant rows in 2018 was 15 September ( $n = 9252$ ;  $sd = 6.1$  days), and the most informative flight dates were 4 and 10 September. These last dates differ from the others mentioned because the 2018 experiment was planted 12 days earlier (22 May) than the latter two experiments (3 June).

The check cultivar was not identified among the 15 most important variables, but the breeding block and the individual who took the field notes in 2019 and 2020 were (Figure 6). This result is consistent with the fact that the  $R^2$  and RMSE were almost unchanged when including the check cultivar in the analyses (Tables 3 and 4). The breeding block and the individual who took the field notes were closely related variables because the task of taking field notes was often assigned by blocks. Still, both variables were much less important than the image features.

The PCA of the 2018 results shows eigenvalues  $> 1$  for the first five principal components (PC), which together explain 87.3% of the total variance of 32 image features collected in 8 drone flights (Table 5). The largest absolute loading scores indicate that PC1 (56.8%) was more associated with the canopy color variation shown on 4, 14, 18, and 23 September (Table 6). On these dates, the loadings were mostly associated with the blue spectral band followed by the red band, which agrees with the literature concerning the relative quantum yield of green leaves [39]. For each PC explaining a portion of the variance (%), a loading score indicates the magnitude (positive or negative) of the effect of a given feature (or variable) relative to the effects of the other features. For PC2, the greatest loading scores were from the first two and the last flight dates: 22 and 28 August and 26 September, respectively. PC1 loadings represented relatively more variation of the three RGB bands, while PC2, PC3, and PC4 loadings represented relatively more *ExG* variation.

**Table 5.** Principal component analysis (PCA) of 32 predictive variables corresponding to a time series of RGB (red, green, blue) images taken across 8 drone flights during 2018 (22 August to 26 September). These predictive variables include the excess green index (*ExG*). Eigenvalues, their proportion explaining the variance, and the cumulative proportion are shown.

Principal Component	Eigen-Values	Proportion of Variance	Cumulative Proportion	Principal Component	Eigen-Values	Proportion of Variance	Cumulative Proportion
1	18.169	0.568	0.568	17	0.072	0.002	0.994
2	4.085	0.128	0.696	18	0.041	0.001	0.995
3	2.873	0.090	0.785	19	0.033	0.001	0.996
4	1.513	0.047	0.833	20	0.027	0.001	0.997
5	1.303	0.041	0.873	21	0.017	0.001	0.998
6	0.875	0.027	0.901	22	0.012	0.000	0.998
7	0.809	0.025	0.926	23	0.011	0.000	0.999
8	0.537	0.017	0.943	24	0.010	0.000	0.999
9	0.433	0.014	0.956	25	0.007	0.000	0.999



Table 5. Cont.

Principal Component	Eigen-Values	Proportion of Variance	Cumulative Proportion	Principal Component	Eigen-Values	Proportion of Variance	Cumulative Proportion
10	0.279	0.009	0.965	26	0.007	0.000	0.999
11	0.218	0.007	0.972	27	0.006	0.000	0.999
12	0.191	0.006	0.978	28	0.006	0.000	1.000
13	0.150	0.005	0.982	29	0.004	0.000	1.000
14	0.116	0.004	0.986	30	0.003	0.000	1.000
15	0.103	0.003	0.989	31	0.002	0.000	1.000
16	0.087	0.003	0.992	32	0.002	0.000	1.000

**Table 6.** Loading scores of a principal component analysis (PCA) that was carried out on 32 predictive variables corresponding to a time series of RGB (red, green, blue) images taken across 8 drone flights (22 August to 26 September 2018). These predictive variables include the excess green index (ExG). The first four principal components (PC) and, in parentheses, their proportion explaining the variance is shown. The score values above  $|0.15|$  are arbitrarily in bold letters to highlight the most important variables.

Variable	Loading Scores				Variable	Loading Scores			
	PC1 (0.568)	PC2 (0.128)	PC3 (0.090)	PC4 (0.047)		PC1 (0.568)	PC2 (0.128)	PC3 (0.090)	PC4 (0.047)
22 August R	<b>-0.16</b>	<b>-0.28</b>	-0.08	<b>0.23</b>	14 September R	<b>-0.20</b>	<b>0.21</b>	-0.03	0.07
22 August G	<b>-0.17</b>	<b>-0.27</b>	-0.07	0.09	14 September G	<b>-0.16</b>	0.15	<b>-0.31</b>	0.02
22 August B	-0.12	<b>-0.24</b>	<b>-0.16</b>	<b>0.28</b>	14 September B	<b>-0.23</b>	0.01	0.00	-0.07
22 August ExG	0.02	<b>0.16</b>	0.14	<b>-0.60</b>	14 September ExG	<b>0.19</b>	-0.08	<b>-0.31</b>	-0.04
28 August R	<b>-0.19</b>	<b>-0.23</b>	0.01	-0.07	18 September R	<b>-0.20</b>	0.13	-0.13	0.06
28 August G	<b>-0.19</b>	<b>-0.17</b>	0.03	<b>-0.19</b>	18 September G	<b>-0.19</b>	0.01	<b>-0.25</b>	-0.07
28 August B	<b>-0.16</b>	<b>-0.19</b>	-0.11	0.02	18 September B	<b>-0.23</b>	0.02	0.01	-0.02
28 August ExG	0.14	<b>0.28</b>	0.08	<b>-0.19</b>	18 September ExG	<b>0.16</b>	<b>-0.18</b>	<b>-0.31</b>	<b>-0.21</b>
4 September R	<b>-0.21</b>	-0.08	<b>0.19</b>	<b>-0.18</b>	23 September R	<b>-0.19</b>	0.14	-0.14	0.05
4 September G	<b>-0.19</b>	0.02	<b>0.17</b>	<b>-0.17</b>	23 September G	<b>-0.21</b>	0.03	<b>-0.18</b>	-0.07
4 September B	<b>-0.21</b>	<b>-0.17</b>	0.02	-0.09	23 September B	<b>-0.23</b>	0.03	-0.01	0.00
4 September ExG	<b>0.20</b>	<b>0.17</b>	-0.15	0.10	23 September ExG	0.12	<b>-0.21</b>	<b>-0.31</b>	<b>-0.32</b>
10 September R	<b>-0.20</b>	0.12	0.08	-0.12	26 September R	-0.05	<b>0.32</b>	<b>-0.24</b>	0.07
10 September G	-0.10	<b>0.27</b>	<b>-0.22</b>	-0.02	26 September G	-0.11	<b>0.23</b>	<b>-0.26</b>	-0.04
10 September B	<b>-0.21</b>	0.00	-0.08	-0.15	26 September B	<b>-0.19</b>	<b>0.18</b>	-0.03	0.04
10 September ExG	<b>0.20</b>	0.07	<b>-0.23</b>	0.12	26 September ExG	0.11	<b>-0.20</b>	<b>-0.29</b>	-0.33

After conducting the PCA, a high correlation coefficient ( $r = 0.82$ ) between PC1 and the R8 stage is observed (Table 7). A principal component regression (PCR) analysis also shows this relationship between PC1 and the R8 stage (Figure 7a). By combining the PCA results with linear regression, this method helps the interpretability and comparison of results with other methods and removes the ‘multicollinearity problem’ [40] that occurs when the explanatory variables are intercorrelated (e.g., plot correlation should occur in time series, and ExG is an index calculated with the RGB bands). In Figure 7a, a PC score with a value of zero indicates the centroid value of the regression, which for 2018 occurred 116 days after the planting date (22 May), when the mean distribution of the plant rows reached their physiological maturity (15 September).

Even though using a single linear combination of the original features (PC1), the PCR of 2018 results indicate that RF is an algorithm that can explain a higher proportion of the variation of the soybean physiological maturity (Figure 7a; Tables 2–4). Nonetheless, applying PCA helps visually identify *a posteriori* what image features were more associated with the response variable (R8 stage), as is shown on the biplot for 2018 (Figure 7b). It is important to mention that the PCA must be carried out without including the response variable to avoid autocorrelation [41], which is why the cumulative proportions of variance

of PC1 and PC2 shown in the figure slightly differ from the values shown in Table 5. Compared to the other five flight dates, the RGB image features collected on 10, 18, and 26 September show less association with the R8 stage (Tables 6 and 7, Figure 7b). Therefore, for the 2018 experiment, the red, green, and blue spectral bands and the calculated *ExG* index were disregarded for these three flight dates before running RF a second time (Table 8). Compared to the results using all data, disregarding three out of eight flight dates did not cause a relevant change in the prediction error (RMSE) and the percentage variation ( $R^2$ ) that the model explained. Not needing to collect data on some dates would save time by not having to acquire and process images.

**Table 7.** Correlation coefficients ( $r$ ) after conducting a principal component analysis (PCA) with 32 predictive variables associated with soybean physiological maturity (R8 stage). The predictive variables corresponded to a time series of RGB (red, green, blue) images taken across eight drone flights (22 August to 26 September 2018) and the excess green index (*ExG*). The correlation between the response variable (R8 stage) and the first four principal components (PC), with their proportion explaining the variance in parentheses, is shown in the last row.

Variable	Correlation Coefficients ( $r$ )				Variable	Correlation Coefficients ( $r$ )			
	PC1 (0.568)	PC2 (0.128)	PC3 (0.090)	PC4 (0.047)		PC1 (0.568)	PC2 (0.128)	PC3 (0.090)	PC4 (0.047)
22 August R	−0.69	−0.56	−0.13	0.28	14 September R	−0.83	0.43	−0.05	0.08
22 August G	−0.73	−0.54	−0.12	0.11	14 September G	−0.68	0.30	−0.53	0.02
22 August B	−0.52	−0.48	−0.27	0.34	14 September B	−0.97	0.03	0.00	−0.08
22 August <i>ExG</i>	0.07	0.32	0.23	−0.74	14 September <i>ExG</i>	0.79	−0.16	−0.53	−0.05
28 August R	−0.81	−0.46	0.01	−0.09	18 September R	−0.85	0.27	−0.22	0.07
28 August G	−0.82	−0.34	0.06	−0.23	18 September G	−0.82	0.01	−0.42	−0.09
28 August B	−0.66	−0.38	−0.19	0.02	18 September B	−0.97	0.04	0.01	−0.02
28 August <i>ExG</i>	0.61	0.56	0.13	−0.24	18 September <i>ExG</i>	0.68	−0.36	−0.53	−0.26
4 September R	−0.88	−0.16	0.32	−0.22	23 September R	−0.83	0.28	−0.24	0.06
4 September G	−0.83	0.04	0.28	−0.21	23 September G	−0.88	0.05	−0.30	−0.08
4 September B	−0.88	−0.34	0.03	−0.11	23 September B	−0.96	0.07	−0.01	0.00
4 September <i>ExG</i>	0.84	0.35	−0.25	0.13	23 September <i>ExG</i>	0.53	−0.43	−0.52	−0.39
10 September R	−0.86	0.25	0.13	−0.15	26 September R	−0.20	0.65	−0.40	0.08
10 September G	−0.42	0.55	−0.38	−0.03	26 September G	−0.48	0.47	−0.44	−0.05
10 September B	−0.91	0.00	−0.13	−0.18	26 September B	−0.82	0.36	−0.05	0.05
10 September <i>ExG</i>	0.86	0.14	−0.39	0.15	26 September <i>ExG</i>	0.49	−0.41	−0.49	−0.41
					R8 stage	0.82	0.23	−0.33	−0.06

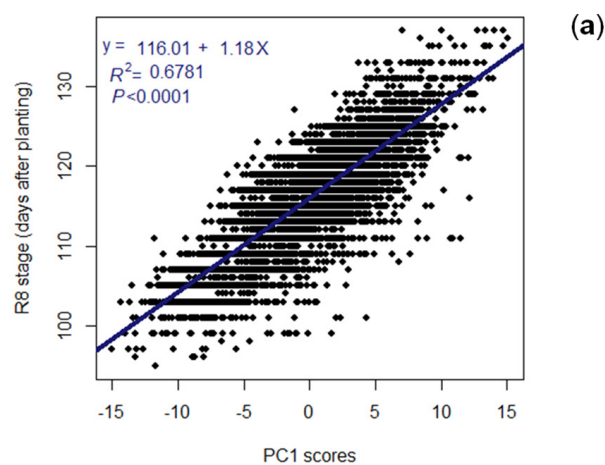


Figure 7. Cont.



**Table 8.** Soybean physiological maturity (R8 stage) predictions based on time series of RGB (red, green, blue) and multispectral images taken from a drone (2018–2020). Predictions were made by applying the Random Forest algorithm and principal component analysis (PCA) to identify what flight dates are best related to the response variable. The percentage of variance explained by the first two principal components (PC1 and PC2), coefficients of regression ( $R^2$ ), and root mean square errors (RMSE) are shown when all drone flight dates or selected drone flight dates were used to conduct both analyses.

Spectral Bands <sup>1</sup>	Flight Dates	Analysis	Indicator	Proportion of Variance (%) Explained by PC1 and PC2, and Random Forest Prediction Accuracy			
				All Drone Flight Dates		Selected Drone Flight Dates	
Red, green, and blue	2018	PCA	PC1	56.8%	22 and 28 August, 4, 10, 14, 18, 23, and 26 September	60.4%	22 and 28
			PC2	12.8%		13.0%	August, 4, 14, and 23
		$R^2$	0.9141	0.9097			
	Random forest	RMSE (days)	1.767	1.812	1.812	September	
		PCA	PC1	36.9%	1, 10, 25, and 30 July, 6, 15, and 28 August, 10, 17, and 24 September	64.1%	10, 17, and 24 September
			PC2	21.3%		13.1%	
$R^2$	0.9380	0.9310					
Random forest	RMSE (days)	1.417	1.492	1.492			
	PCA	PC1	43.8%	30 June, 15 July, 4, 18, and 31 August, 10, 14, 22, and 30 September	58.3%	10, 14, 22, and 30 September	
		PC2	15.8%		12.8%		
$R^2$	0.9080	0.9067					
Random forest	RMSE (days)	1.347	1.356	1.356			
	PCA	PC1	32.4%	1, 10, 25, and 30 July, 6, 15, and 28 August, 10, 17, and 24 September	55.5%	10, 17, and 24 September	
		PC2	19.3%		16.9%		
$R^2$		0.9366	0.9313				
Random forest	RMSE (days)	1.435	1.491	1.491			
	PCA	PC1	39.0%	30 June, 15 July, 4, 18, and 31 August, 10, 14, 22, and 30 September	48.4%	10, 14, 22, and 30 September	
		PC2	17.8%		16.2%		
$R^2$	0.9111	0.9091					
Random forest	RMSE (days)	1.325	1.339	1.339			

<sup>1</sup> In 2018, the images were taken only with a digital RGB (red, green, blue) camera, whereas in 2019 and 2020, they were taken with two multispectral camera models of five bands: red, green, blue, red edge, and near-infrared. For the latter two years, the analyses were divided into (1) using the five bands (normalized data) and (2) using only the red, green, and blue bands (values from 0 to 255) to compare the results with those obtained in 2018 (only RGB images). Also, with the red, green, and blue bands, the excess green index ( $ExG$ ) was calculated.

### 3.5. Reliability of Models When Tested in an Independent Environment

The reliability of the fitted models for predicting soybean physiological maturity in new environments depended mainly on the similarity between the environment used to train the model and the environment in which the model was tested (i.e., the training test data relationship). Two main differences between the years were that the 2018 experiment was planted earlier (22 May) than the other two (3 June 2019–2020) and that the plant rows were taller and had a greater canopy area due to higher water availability during the growing cycle. Using only the calculated  $ExG$  index, this is shown comparing the scenarios Training 2019:Test 2020 vs. Training 2018–2019:Test 2020 (Figure 8a,c), and also when the check cultivars of the test year (2020) were added into the training data (Figure 8b,d). Similarly, the same happened when the RGB bands were included in the models in addition to  $ExG$ . Compared to these last models, the four fitted models using only  $ExG$  (Figure 8) explained, on average, 7.2% more of the R8 stage variation in 2020, while the prediction error decreased by 13.2%. Noticeably, adding the check cultivars into the training data (Figure 8b,d) caused a remarkable change in both indicators, with an increase in the  $R^2$  (12.4% and 12.7%, respectively) and a decrease in the RMSE (−14.1% and −15.6%, respectively). A prediction improvement that was slightly better when, instead of 2019 (Figure 8b), the last two years (2018–2019) were used to train the model (Figure 8d).

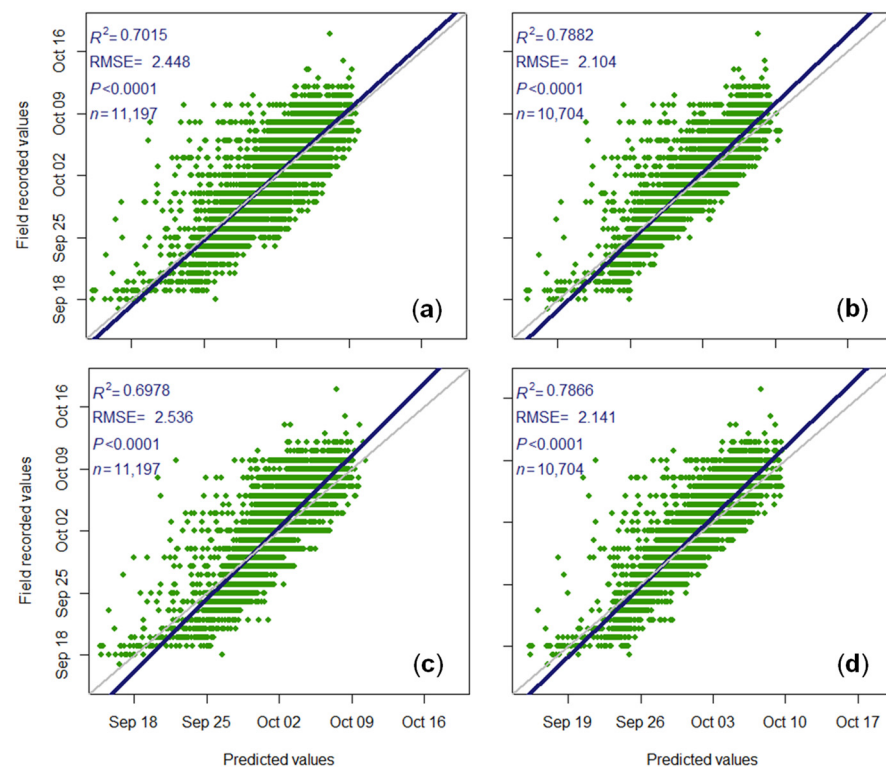
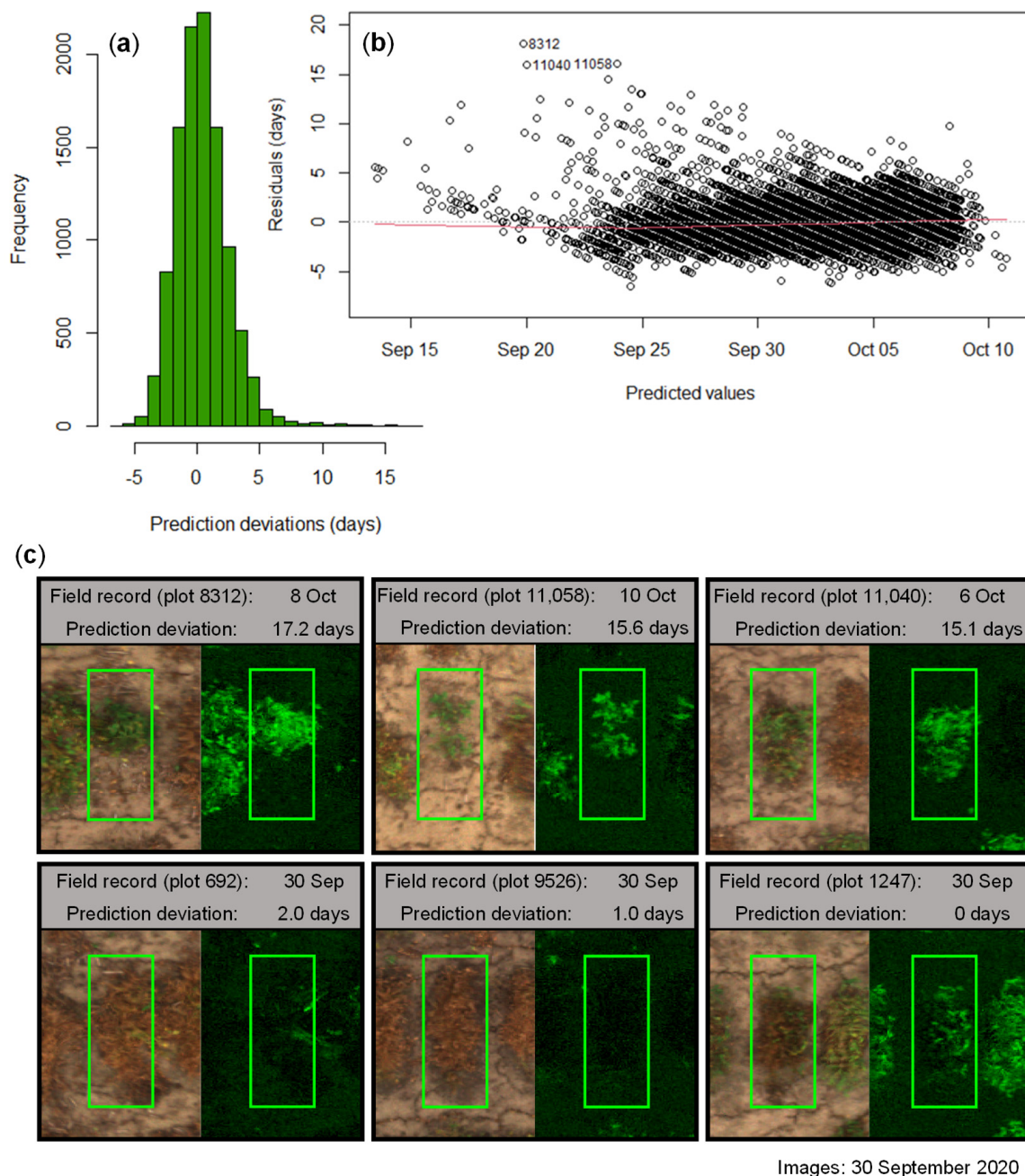


Figure	Year	Training data	Drone flights			
(c) and (d)	2018	All plant rows	Aug 22	Aug 28	Sep 4	Sep 10
(a–d)	2019	All plant rows	Aug 28	Sep 10	Sep 17	Sep 24
(b) and (d)	2020	Only check cultivars	Aug 31	Sep 10	Sep 14	Sep 22

**Figure 8.** Soybean physiological maturity (R8 stage) predictions for 2020 using four models trained with data from field recorded values collected from two previous experiments (2018–2019). The three experiments corresponded to breeding experiments containing plant rows of F<sub>4:5</sub> experimental lines grown at the University of Illinois Research and Education Center near Savoy, IL. The four models were adjusted by applying the Random Forest algorithm to associate the field recorded values with a time series of the excess green index (*ExG*) and three classification variables (breeding block, the individual who took the field notes, and the check cultivar). Calculated from the red, green, and blue spectral bands, *ExG* was obtained from digital images taken with a drone. The four models were adjusted using the following training: test data relationship: (a) Training 2019:Test 2020 ( $n = 51:49$ ); (b) Training 2019<sub>plus 2020 checks</sub>:Test 2020<sub>without checks</sub> ( $n = 53:47$ ); (c) Training 2018–2019: Test 2020 ( $n = 65:35$ ); and (d) Training 2018–2019<sub>plus 2020 checks</sub>:Test 2020<sub>without checks</sub> ( $n = 67:33$ ). The deviation of the regression line (blue) from the 1:1 line (gray) indicates the model's prediction bias. The table below the figures gives the data used to train the models in each figure (a–d).

Prediction deviations of more than five days were infrequent in the current study; still, they were more frequent and extreme on the distribution's right tail (Figure 9a). This histogram and the residuals' analysis (Figure 9b) were produced using the best-fitted model from 2019 and were tested in 2020 (Figure 8b). The residuals' analysis (field observed dates – predicted dates) shows three large outliers that were predicted to mature several days before they actually matured; however, these outliers were plant rows that only had a few plants that reached maturity late in the season (Figure 9b,c).



**Figure 9.** (a) Frequencies, (b) residuals, and (c) images showing prediction deviations for soybean physiological maturity (R8 stage) collected in a breeding experiment with plant rows of  $F_{4.5}$  experimental lines in 2020. The mean residual (red line) indicates in (b) the prediction bias across time compared to predictions with zero bias from the observed R8 dates (gray dashed line). The images on the right show the excess green index ( $ExG$ ), which is calculated with the red, green, and blue bands (images on the left). On the top of (c), the images show the three worst maturity predictions identified on (b); the bottom shows three examples considering predictions with an error of 2, 1, and 0 days from 30 September. The maturity predictions were carried out using a model (Figure 8b) trained with data collected in a breeding experiment planted in 2019 ( $n = 11,197$ ) and in the eight check cultivars replicated in the 2020 experiment. The 2020 experiment minus the checks ( $n = 11,197 - 493$ ) was used to test the model, which was adjusted with the Random Forest algorithm using time series of  $ExG$  and three classification variables (breeding block, the individual who took the field notes, and the check cultivar).

On the other hand, the three images from 2020 at the bottom of Figure 9c show examples where the deviations were 2.0, 1.0, and 0.0 days from 30 September; this is the predicted maturity date for each plant row and the same date these images were taken. The images for plots 692 and 9526 show mature plant rows, suggesting that the deviations between observed and predicted dates (2.0 and 1.0 days, respectively) were because the person rating the plot scored them incorrectly. Inaccurate scoring of maturity can occur because the field raters often score plots several days ahead, given that they cannot return to each plot daily, and may only return to the field every five to seven days, considering that they must rate thousands of plant rows.

#### 4. Discussion

The prediction error ranged between 1.31 and 1.78 days ( $R^2 = 0.91$ ) for any combination of explanatory and classification variables used in the three analyzed years (Tables 2–4). A few other studies predicted soybean physiological maturity with a lower error, including Lindsey et al. [21] ( $R^2 = 0.92$ ; RMSE = 1.29 days), but this was with fewer plots (96), and instead of drone flights and RF, a handheld crop sensor (660- and 780-nm wavelengths) and an inverse logistic model were used. Other studies also reported higher prediction accuracies than the current study [14–16]. However, instead of making predictions based on a regression model carrying out a longitudinal analysis of a time series of images, as in the current study, they used a binary (immature and mature) or a multi-class classification model that included one or two additional labels (harvested and near-mature plots). These labeling methods can lead to some true positive classifications being equally valid if the plant rows reach the R8 stage on the same day or some days before the images are taken. Meanwhile, associating maturity dates recorded in the field with time series of images, as was performed in our study (Tables 2–4), may result in lower prediction accuracy than those obtained with classification models with a few labels but can achieve greater certainty about what date the plant rows matured.

Some authors have mentioned that one advantage of RF is that it runs faster than CNN and other deep learning methods [16,18]. The above may not be trivial when researchers must test several models before choosing the one that best represents the variation of the response variable. An example is that running RF using the germplasm information (i.e., check cultivar, F<sub>4:5</sub> population, and female and male lines) required around one hour per annual analysis (e.g., Intel® Xeon® Platinum 8370C CPU, 2.8 GHz, RAM 32 GB –Intel Corporation, Santa Clara, CA, USA). This processing time was reduced by two-thirds when the germplasm information is disregarded, which requires computing time when interacting with the image features.

The prediction error (RMSE) difference between using all data compared to discarding the late plant rows (Figure 5a–c vs. Figure 5d–f) was similar to the variation observed among members of a team of individuals when they took notes in the field (Figure 4). Therefore, when predicting the R8 stage, conserving a high  $R^2$  value by using all plant rows is preferable so that the predictive power of the models can be maximized, especially when the models are used in different environments (year or location). Concerning this, an advantage of RF over the LOESS model is that the latter cannot make predictions for plant rows that have yet to mature [18,24]. Concerning this, Volpato et al. [24] adjusted a LOESS model for predicting maturity but required several drone flights carried out weekly before and after the R8 stage. Meanwhile, the pipeline we implemented can associate color changes that occur in the plant rows prior to maturity. Thus, predicting physiological maturity several days before harvest would provide breeders with relevant information for plant row selection.

In agreement with Woebbecke et al. [25] and Larrinaga et al. [26], who observed that the ExG index enhances the contrast between vegetation and soil pixels, these results suggest that the contrast helps the RF algorithm to identify which plant rows have not reached maturity. On the other hand, for germplasm showing green-stem syndrome, using ExG could magnify the error of predicting maturity. Based on the variable importance

measure and PCA, for the three years of experiments, *ExG* was the most important image feature associated with the R8 stage (Figures 6 and 7b). However, these results differ from the observed by Teodoro et al. [29], who did not identify significant differences between three groups of features: multispectral data (five bands), the same plus several VIs (all different from *ExG*), and only VIs. Still, by using deep learning instead of RF, they obtained significantly larger prediction errors (RMSE) with the second and third groups of features mentioned above.

After applying PCA to the three experiments (2018–2020) to discard flight date information that was redundant or less associated with the R8 stage, the results after rerunning RF show that the prediction errors (RMSE) increased but only within the range of 0.7–5.3% (Table 8). While applying a deep learning method (CNN-LSTM), Moeinziade et al. [18] analyzed RGB images taken from five drone flights carried out every 7.4 days on average. As a result, the authors could explain the variation of the R8 stage in soybean as well or better than the present work in two out of six environments ( $R^2 = 0.51$ – $0.95$ ; RMSE = 2.5–0.9 days). Our study, applying the RF algorithm after discarding several flight dates based on PCA, shows that a few weekly drone flights (3–5) can explain 91–93% of the total variation with an error of 1.8–1.3 days. This would be performed by starting around the date the earliest germplasm reached maturity and finishing before all the plant rows had reached it (Table 8).

Except for one study, as far as we know [17], others [14–16,18,24,29] tested their models in data subsets of the same trial or environment where they were trained, but not reciprocally or independently across the environments (i.e., different years or locations). In the mentioned exception [17], the authors applied CNN to train models that were reciprocally tested in five environments. One experiment was planted in 2018 (Savoy, IL, USA) and the other four in 2019; one was planted in Rolândia, PR (Brazil), and the other in three locations in Illinois (USA). It is noteworthy that the field notes and the time series of RGB images used for Savoy, IL, are the same we collected in 2018 for our study.

The model trained by Trevisan et al. [17] with data collected in Brazil had a higher prediction error (RMSE ~3 days) than the models trained in Illinois (RMSE ~2 days). This prediction error was almost as small as our study (Table 8). Nonetheless, when the authors reciprocally tested the five trained models independently in the remaining four environments, the prediction errors in some combinations increased noticeably (RMSE ~3 to 4.5 days). This increase was lower when check cultivars (some in common along the environments) were used to correct the raw predictions, which agrees with the findings of the current work (Figure 8). Unlike Trevisan et al. [17], who adjusted the R8 predictions by applying a simple regression analysis after running CNN, we included the checks as a classification variable *a priori* of running the RF algorithm in the current work.

On the other hand, using redundant information as multispectral data in the case of predicting physiological maturity, which did not improve the predictions compared to RGB data when the models were tested in the same environment (Tables 3 and 4), or using too many drone flight dates (Table 8) may cause overfitting and a decrease in the predictive power of the model when it is tested in a new and different environment. In this sense, the models tested in independent environments (Figure 8) that were trained with a single image feature (*ExG*) had a lower prediction error than when they were also trained with the spectral bands (i.e., a RMSE decrease of 13.2%, on average).

When tested in independent environments, the reliability of models is sensitive to overfitting because even when planting similar germplasm in the same location, the growing conditions between experiments differ annually (i.e., planting date associated with day length, sunshine hours, accumulation of growing degree-days, and water availability). As happened with our study (Table 8 vs. Figure 8) and the study by Trevisan et al. [17], predictions made using a model developed in an independent environment should be less accurate than predictions made in a data subset of the same environment. Thus, to avoid training the models with redundant information, including shared check cultivars when planting the experiments would decrease a portion of the prediction error. Despite this, and even though we tested the models in an independent environment, a limitation



could be whether these models were tested in other breeding programs or locations different from where they were trained (Savoy, IL, USA), which could lead to overfitting. Thus, a future challenge will be studying the inclusion of other variables modeling the growing conditions.

## 5. Conclusions

The RF machine learning algorithm was used in this study to inquire how the date of soybean physiological maturity can be predicted using a time series of aerial images. Principal component analysis helped identify redundant or informative information to train the models. In this sense, including or not including redundant information may not make a difference when the models are tested in a data subset of the same environment. However, by discarding redundant information, more general models were fitted in this study, making the predictions more accurate when tested in an independent environment. In this sense, a finding was that for predicting physiological maturity, synthesizing the color information into a single value—the excess green index (*ExG*) calculated only with the RGB bands—resulted in less model overfitting when the models were tested in an independent environment. Another finding was that carrying out more than five weekly drone flights had little impact on decreasing the prediction error by ~0.1 days compared to flying the drone 8–10 times. Images from drone flights only need to be collected starting when the earliest germplasm reaches maturity. Continuing weekly drone flights until all the germplasm is mature would not be necessary. Finally, we also found that when using prediction models from previous experiments, adding field records of the check cultivars from the new environment that maturity will be predicted, the prediction error decreased by ~15% from 2.5 to 2.1 days. Including check cultivars in common as a classification variable would alleviate the bias due to the environmental differences between training and test experiments (i.e., planting date, weather conditions during the growing season, and location).

Studies applying high-throughput phenotyping to predict soybean physiological maturity have been conducted before; nonetheless, few studies have evaluated thousands of plant rows in breeding trials and included studies in more than a single growing season. By applying a pipeline like the one used in this study, soybean breeding programs could fit their models adapted by location to predict soybean physiological maturity for thousands of plant rows before harvest. In addition to saving time in the process, the breeder could easily classify the germplasm by maturity groups, discard the plant rows of long-cycle length before harvest, and, in turn, avoid taking unnecessary field notes.

**Author Contributions:** Conceptualization, O.P., B.D. and N.M.; methodology, O.P. and N.M.; software, O.P.; validation, O.P.; formal analysis, O.P.; investigation, O.P., N.M. and B.D.; resources, B.D. and N.M.; data curation, O.P.; writing—original draft preparation, O.P.; writing—review and editing, O.P., B.D. and N.M.; visualization, O.P.; supervision, B.D. and N.M.; project administration, B.D. and N.M.; funding acquisition, B.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received funding from North Central Soybean Research Program, (NCSRC), “SOYGEN 3: Building capacity to increase soybean genetic gain for yield and composition through combining genomics-assisted breeding with characterization of future environments”.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author, the data are not publicly available due to privacy.

**Acknowledgments:** We want to thank Nathan Schmitz (GDM Seeds Inc.) for providing the drone kit to fly the missions in 2018 and for helping us with the software Metashape; Rodrigo Gonçalves-Trevisan who gave us good tips using Metashape, QGIS, and R; and Sebastián Varela, who made a flight mission in 2019 when the hired digital agricultural specialist could not fly the drone.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Koester, R.P.; Skoneczka, J.A.; Cary, T.R.; Diers, B.W.; Ainsworth, E.A. Historical gains in soybean (*Glycine max* Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. *J. Exp. Bot.* **2014**, *65*, 3311–3321. [[CrossRef](#)] [[PubMed](#)]
2. Fehr, W.R.; Caviness, C.E. *Stages of Soybean Development*; Iowa Agricultural and Home Economics Experiment Station Publications; Special Report; Iowa State University Digital Repository: Ames, IA, USA, 1977; Volume 87, pp. 1–12. Available online: <https://dr.lib.iastate.edu/handle/20.500.12876/90239> (accessed on 3 November 2024).
3. Singh, A.K.; Singh, A.; Sarkar, S.; Ganapathysubramanian, B.; Schapaugh, W.; Miguez, F.E.; Carley, C.N.; Carroll, M.E.; Chiozza, M.V.; Chiteri, K.O.; et al. High-throughput phenotyping in soybean. In *High-Throughput Crop Phenotyping*; Zhou, J., Nguyen, H.T., Eds.; Springer Nature: Cham, Switzerland, 2021; pp. 129–163. [[CrossRef](#)]
4. Edwards, D.; Batley, J. Plant bioinformatics: From genome to phenome. *Trends Biotechnol.* **2004**, *22*, 232–237. [[CrossRef](#)] [[PubMed](#)]
5. Deshmukh, R.; Sonah, H.; Patil, G.; Chen, W.; Prince, S.; Mutava, R.; Vuong, T.; Valliyodan, B.; Nguyen, H.T. Integrating omic approaches for abiotic stress tolerance in soybean. *Front. Plant Sci.* **2014**, *5*, 00244. [[CrossRef](#)] [[PubMed](#)]
6. Krause, M.R.; Mondal, S.; Crossa, J.; Singh, R.P.; Pinto, F.; Haghighattalab, A.; Shrestha, S.; Rutkoski, J.; Gore, M.A.; Sorrells, M.E.; et al. Aerial high-throughput phenotyping enables indirect selection for grain yield at the early generation, seed-limited stages in breeding programs. *Crop Sci.* **2020**, *60*, 3096–3114. [[CrossRef](#)]
7. Araus, J.L.; Kefauver, S.C.; Zaman-Allah, M.; Olsen, M.S.; Cairns, J.E. Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* **2018**, *23*, 451–466. [[CrossRef](#)]
8. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.* **2021**, *32*, 1–6. [[CrossRef](#)]
9. Gitelson, A.; Merzlyak, M.N. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *J. Plant Physiol.* **1994**, *143*, 286–292. [[CrossRef](#)]
10. Eftekhari, M.; Ma, C.; Orlov, Y.L. Editorial: Applications of artificial intelligence, machine learning, and deep learning in plant breeding. *Front. Plant Sci.* **2024**, *15*, 1420938. [[CrossRef](#)]
11. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
12. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
13. Lee, T.-H.; Ullah, A.; Wang, R. Bootstrap aggregating and Random forest. In *Macroeconomic Forecasting in the Era of Big Data*; Fuleky, P., Ed.; Serie: Advanced Studies in Theoretical and Applied Econometrics; Springer Nature: Cham, Switzerland, 2020; Volume 52, pp. 389–429. [[CrossRef](#)]
14. Yu, N.; Li, L.; Schmitz, N.; Tian, L.F.; Greenberg, J.A.; Diers, B.W. Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. *Remote Sens. Environ.* **2016**, *187*, 91–101. [[CrossRef](#)]
15. Hu, J.; Yue, J.; Xu, X.; Han, S.; Sun, T.; Liu, Y.; Feng, H.; Qiao, H. UAV-based remote sensing for soybean FVC, LCC, and maturity monitoring. *Agriculture* **2023**, *13*, 692. [[CrossRef](#)]
16. Zhang, S.; Feng, H.; Han, S.; Shi, Z.; Xu, H.; Liu, Y.; Feng, H.; Zhou, C.; Yue, J. Monitoring of soybean maturity using UAV remote sensing and deep learning. *Agriculture* **2023**, *13*, 110. [[CrossRef](#)]
17. Trevisan, R.; Pérez, O.; Schmitz, N.; Diers, B.; Martin, N. High-throughput phenotyping of soybean maturity using time series UAV imagery and convolutional neural networks. *Remote Sens.* **2020**, *12*, 3617. [[CrossRef](#)]
18. Moeinzade, S.; Pham, H.; Han, Y.; Dobbels, A.; Hu, G. An applied deep learning approach for estimating soybean relative maturity from UAV imagery to aid plant breeding decisions. *Mach. Learn. Appl.* **2022**, *7*, 100233. [[CrossRef](#)]
19. Christenson, B.S.; Schapaugh, W.T.; An, N.; Price, K.P.; Prasad, V.; Fritz, A.K. Predicting soybean relative maturity and seed yield using canopy reflectance. *Crop Sci.* **2016**, *56*, 625–643. [[CrossRef](#)]
20. Zhou, J.; Yungbluth, D.; Vong, C.N.; Scaboo, A.; Zhou, J. Estimation of the maturity date of soybean breeding lines using UAV-based multispectral imagery. *Remote Sens.* **2019**, *11*, 2075. [[CrossRef](#)]
21. Lindsey, A.J.; Craft, J.C.; Barker, D.J. Modeling canopy senescence to calculate soybean maturity date using NDVI. *Crop Sci.* **2020**, *60*, 172–180. [[CrossRef](#)]
22. Yuan, W.; Wijewardane, N.K.; Jenkins, S.; Bai, G.; Ge, Y.; Graef, G.L. Early prediction of soybean traits through color and texture features of canopy RGB imagery. *Sci. Rep.* **2019**, *9*, 14089. [[CrossRef](#)]
23. Narayanan, B.; Floyd, B.; Tu, K.; Ries, L.; Hausmann, N. Improving soybean breeding using UAS measurements of physiological maturity. In Proceedings of the Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV, Baltimore, MD, USA, 15–16 April 2019; Thomasson, J.A., McKee, M., Moorhead, R.J., Eds.; SPIE: Bellingham, WA, USA, 2019; Volume 11008. [[CrossRef](#)]
24. 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. [[CrossRef](#)]
25. Woebbecke, D.M.; Meyer, G.E.; Von Bargaen, K.; Mortensen, D.A. Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. ASAE* **1995**, *38*, 259–269. [[CrossRef](#)]
26. Larrinaga, A.R.; Brotons, L. Greenness indices from a low-cost UAV imagery as tools for monitoring post-fire forest recovery. *Drones* **2019**, *3*, 6. [[CrossRef](#)]
27. Jiang, Y.; Li, C. Convolutional neural networks for image-based high-throughput plant phenotyping: A Review. *Plant Phenomics* **2020**, *2020*, 4152816. [[CrossRef](#)] [[PubMed](#)]

28. Zhang, Z.; Flores, P.; Igathinathane, C.; Naik, D.L.; Kiran, R.; Ransom, J.K. Wheat lodging detection from UAS imagery using machine learning algorithms. *Remote Sens.* **2020**, *12*, 1838. [[CrossRef](#)]
29. Teodoro, P.E.; Teodoro, L.P.R.; Baio, F.H.R.; da Silva Junior, C.A.; dos Santos, R.G.; Ramos, A.P.M.; Pinheiro, M.M.F.; Osco, L.P.; Gonçalves, W.N.; Carneiro, A.M.; et al. Predicting days to maturity, plant height, and grain yield in soybean: A machine and deep learning approach using multispectral data. *Remote Sens.* **2021**, *13*, 4632. [[CrossRef](#)]
30. Burggraaff, O.; Schmidt, N.; Zamorano, J.; Pauly, K.; Pascual, S.; Tapia, C.; Spyrakos, E.; Snik, F. Standardized spectral and radiometric calibration of consumer cameras. *Opt. Express* **2019**, *27*, 19075. [[CrossRef](#)]
31. Google. Google Satellite [Map Tiles]; Google Maps. 2024. Available online: <https://www.google.cn/maps/vt?lyrs=s@189&gl=cn&x={x}&y={y}&z={z}> (accessed on 3 November 2024).
32. Pebesma, E. Simple features for R: Standardized support for spatial vector data. *R J.* **2018**, *10*, 439–446. [[CrossRef](#)]
33. Bivand, R.; Keitt, T.; Rowlingson, B.; Pebesma, E.; Sumner, M.; Hijmans, R.; Baston, D.; Rouault, E.; Warmerdam, F.; Ooms, J.; et al. Package ‘Rgdal’. Bindings for the Geospatial Data Abstraction Library. 2023. Available online: <https://cran.r-project.org/web/packages/rgdal/index.html> (accessed on 6 February 2023).
34. Hijmans, R.J.; Etten, J.V.; Sumner, M.; Cheng, J.; Baston, D.; Bevan, A.; Bivand, R.; Busetto, L.; Canty, M.; Fasoli, B.; et al. Package ‘Raster’. Geographic Data Analysis and Modeling. 2023. Available online: <https://cran.r-project.org/web/packages/raster/index.html> (accessed on 7 February 2023).
35. Ross, N.; Sumner, M.; Ooms, J.; Stevens, A.; EcoHealth Alliance; USAID Predict. Package ‘Fasterize’. Fast Polygon to Raster Conversion. 2022. Available online: <https://cran.r-project.org/web/packages/fasterize/index.html> (accessed on 7 February 2023).
36. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; R Core Team; et al. Package ‘Caret’. Classification and Regression Training. 2022. Available online: <https://cran.r-project.org/web/packages/caret/index.html> (accessed on 13 February 2023).
37. Breiman, L.; Cutler, A. Package ‘randomForest’. Breiman and Cutler’s Random Forests for Classification and Regression. 2022. Available online: <https://cran.r-project.org/web/packages/randomForest/index.html> (accessed on 10 February 2023).
38. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [[CrossRef](#)]
39. McCree, K.J. The action spectrum, absorptance and quantum yield of photosynthesis in crop plants. *Agric. Meteorol.* **1971**, *9*, 191–216. [[CrossRef](#)]
40. Suryanarayana, T.M.V.; Mistry, P.B. *Principal Component Regression for Crop Yield Estimation*; Serie: SpringerBriefs in Applied Sciences and Technology; Springer Nature: Singapore, 2016. [[CrossRef](#)]
41. Zuur, A.F.; Ieno, E.N.; Elphick, C.S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **2010**, *1*, 3–14. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.