



# Article Improving Strawberry Yield Prediction by Integrating Ground-Based Canopy Images in Modeling Approaches

Amr Abd-Elrahman <sup>1,2,\*</sup>, Feng Wu<sup>2</sup>, Shinsuke Agehara <sup>2,3</sup>, and Katie Britt <sup>1,2</sup>

- School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, FL 32611, USA; k.britt@ufl.edu
- <sup>2</sup> Gulf Coast Research and Education Center, University of Florida, Wimauma, FL 33598, USA; fengwu@ufl.edu (F.W.); sagehara@ufl.edu (S.A.)
- <sup>3</sup> Department of Horticulture, University of Florida, Gainesville, FL 32611, USA
- \* Correspondence: aamr@ufl.edu

**Abstract:** Strawberries (*Fragaria* × *ananassa* Duch.) are highly perishable fruit. Timely prediction of yield is crucial for labor management and marketing decision-making. This study demonstrates the use of high-resolution ground-based imagery, in addition to previous yield and weather information, for yield prediction throughout the season at different intervals (3-4 days, 1 week, and 3 weeks preharvest). Flower and fruit counts, yield, and high-resolution imagery data were collected 31 times for two cultivars ('Florida Radiance' and 'Florida Beauty') throughout the growing season. Orthorectified mosaics and digital surface models were created to extract canopy size variables (canopy area, average canopy height, canopy height standard deviation, and canopy volume) and visually count flower and fruit number. Data collected at the plot level (6 plots per cultivar, 24 plants per plot) were used to develop prediction models. Using image-based counts and canopy variables, flower and fruit counts were predicted with percentage prediction errors of 26.3% and 25.7%, respectively. Furthermore, by adding image-derived variables to the models, the accuracy of predicting out-of-sample yields at different time intervals was increased by 10-29% compared to those models without image-derived variables. These results suggest that close-range high-resolution images can contribute to yield prediction and could assist the industry with decision making by changing growers' prediction practices.

Keywords: canopy size metrics; Fragaria × ananassa; high-resolution; image analysis; regression model

# 1. Introduction

Due to their highly perishable nature, strawberry crops are prone to many factors that impact quantity and quality throughout the season [1]. Strawberry production cycles are influenced by many variables such as weather, pollination, planting date, cultivar, pests, and disease impact [2–5]. Crop management practices and logistical choices can also impact crop performance and market variability. Furthermore, the competition for market share and harvest labor affects profitability of this already fluctuating, complicated commodity [6]. High natural variability in strawberry production makes it difficult for growers to anticipate accurate yield distribution and make marketing decisions. A yield model that can predict yield a few days to a few weeks in advance would be a powerful tool for growers, allowing profit maximization by planning more efficiently for marketing costs, labor distribution, and minimization of intra-market competition among regional growers.

Previous studies found that strawberry yield can be predicted using field observations coupled with weather data to some extent. MacKenzie and Chandler [2] used weather information and flower count data collected over two consecutive seasons to predict strawberry yield. Their results found that the number of fruits was more accurately predicted than was fruit weight. A different study successfully predicted weekly strawberry



Citation: Abd-Elrahman, A.; Wu, F.; Agehara, S.; Britt, K. Improving Strawberry Yield Prediction by Integrating Ground-Based Canopy Images in Modeling Approaches. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 239. https://doi.org/10.3390/ijgi10040239

Academic Editors: Wolfgang Kainz and Flavio Lupia

Received: 11 February 2021 Accepted: 4 April 2021 Published: 7 April 2021

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



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). yield using artificial neural networks and soil inputs [7]. The number of crowns, fresh weight, and dry weight were used by Bartczak, Lisiecka, and Knaflewski [8], who found fresh weight to be the most important input in the model. Yield predictions at three to four days before harvest were accomplished using previous yield and fruit counts at specific physiological maturity stages [9]. Previous yield is likely a significant variable given the phenological ripening pattern of berries over time through the growing season [10]. The implications of physical parameters extracted from plant-canopy geometry correlating with yield highlight the value of extracting multiple canopy geometry metrics for use in yield modeling.

While direct field measurements are reliable and have been used to obtain accurate yield model results, there are significant drawbacks to models built entirely on laborintensive, time-consuming, expensive, and often destructive measurements [11]. These methods are often neither practical nor sustainable when targeting large-scale farm operations or research applications. Remote sensing technology allows for large amounts of data to be collected quickly, and the information facilitated by this technology provides a huge potential for extracting many variables quickly and non-destructively [12,13]. The remote sensing imagery is important to build strawberry yield prediction models that can be practically implemented for farm operations. The rapid development of high-throughput imaging technologies and image analysis algorithms has made the derivation of canopy specific information feasible.

Biophysical parameters, such as leaf area and dry weight biomass of strawberry canopies, have been modeled using ground-based remote sensing imagery [13,14]. Lidar is often used for canopy modeling, but this technology is still very costly for extensive data acquisition throughout the growing season. Due to the need not only to model the canopy but also to identify fruit and flower locations, a moving platform with necessary navigation sensors and control points was used in this study to capture true-color and infrared imagery to provide canopy variables. These images have enough overlap and can be used to create orthorectified mosaics (orthomoisaics) and Digital Surface Models (DSM) to extract canopy characteristics. By using a low-altitude sensor as opposed to a piloted aircraft or a satellite sensor, the geometrical characteristics of individual canopies, as well as flower and fruit counts, can be extracted due to the high spatial resolution of the images. Other issues, such as cloud existence and the need for frequent data acquisition sensions, can also be overcome using this type of imaging system.

In this study, our main objective was to demonstrate the feasibility of developing strawberry yield prediction statistical models using image-derived variables such as canopy size variables and flower and fruit counts, weather information, and previous yield data. The study emphasizes the roles of high spatial- and temporal-resolution images in deriving model variables. To achieve our objective, strawberry yield, manual fruit and flower counts, and close-range imagery were collected and analyzed. Field fruit and flower counts were compared to visually interpreted counts from the images. Statistical modeling of field counts as dependent variables and image-based counts as independent variables were developed and validated. Other statistical models were developed to predict strawberry yield at different time intervals ahead of harvest. We compared the results of modeling strawberry yield at 3–4 days ahead, one week ahead, and 3 weeks ahead of harvest using the models' goodness-of-fit and out-of-sample validation.

# 2. Materials and Methods

# 2.1. Study Site

The field experiments of this study were conducted at the University of Florida's Gulf Coast Research and Education Center (GCREC) in Balm, FL (latitude: 27.76030° N; longitude: 82.22798° W), during the 2017–2018 winter strawberry growing season. Two strawberry cultivars ('Florida Radiance' and 'Florida Beauty') were used in the trials. Each cultivar had 6 plots with 24 plants per plot. The plots were arranged in a completely



randomized block design. Figure 1 shows the general location of the study site and experiment layout. Commercial standard management practices were followed.

**Figure 1.** Strawberry yield study site at the Gulf Coast Research and Education Center in Balm, Florida (latitude: 27.760296° N; longitude: 82.227977° W).

#### 2.2. Control Point Establishment and Image Acquisition

The season-long image acquisition process was preceded by Ground Control Point (GCP) establishment. Fixed markings that were visible in the images were set in the field early in the season and used throughout the season to georeference the acquired images. Three Global Navigation Satellite System (GNSS) receivers collected static data for at least four hours on eight control points located throughout the study site. Two points were used as base and backsight points to survey additional control points at the end of each bed and along the plastic covers of the strawberry beds using a surveying total station instrument. The points at the end of each bed were established in the ground using iron rods in the middle of painted circular plastic targets, and points established on the plastic beds were painted on areas of the beds not covered by the canopies, similar to the method adopted by Guan et al. [14] Figure 2 shows the control points established at the end of the strawberry beds and on the plastic beds. The GCP coordinates were determined using GNSS postprocessing in the North American Datum (NAD83) and Universal Transverse Mercator

(UTM) Zone 17N map projection. This fixed set of control points was used throughout the season to create image mosaics and DSM, allowing not only accurate extraction of the canopies' geometric properties but also the comparison of sequential datasets captured during the season.



Figure 2. Targets at ends of beds and on beds for 3D modeling.

Images were acquired using a custom-built platform towed by a tractor driven through the strawberry beds (Figure 3). Two consumer-grade Nikon D-300 digital cameras were used. The first camera captured Red-Green-Blue (RGB) imagery, and the other was a camera modified by removing the near-infrared (NIR) filter to capture NIR imagery. The cameras were mounted 3.5 m above ground and driven at 0.5 m per second to allow for about 70% forward overlap between images. Factoring in the distance between adjacent beds created 60% sidelap between the acquired images, which was needed to construct 3D models. The cameras were automatically triggered by a software program and hardware interface developed in-house that synchronized the RGB and NIR camera triggers and acquired a GNSS time stamp for each camera trigger instance [15,16]. Imaging trajectory was collected during the image acquisition mission using a geodetic-grade GNSS receiver mounted on the cameras' carrying platform. The GNSS trajectory data were analyzed using kinematic post-processing analysis techniques. Camera trigger time stamps were matched with the GNSS trajectory to produce camera trigger locations. The images captured by the RGB and NIR cameras had about 0.5 mm spatial resolution in their raw form. Based on the time of day for image collection, which was the same across the season, the images were captured only while the tractor was moving in the south to north direction to avoid shadows from the platform and to keep shadows in the canopy consistent.



**Figure 3.** Imaging of strawberry plots using an imaging platform built in this study. The platform is equipped with (**A**) timing GPS, (**B**) Survey-quality GNSS receiver, (**C**) Nikon D300 RGB, (**D**) Nikon D300 IR, (**E**) synchronization hardware and acquisition software (laptop), and (**F**) camera trigger box hardware.

Two beds to the east and west of the experiment bed were also imaged to strengthen the image acquisition geometry, provide multi-view images, and facilitate three-dimensional data extraction. The beds were imaged twice per week throughout the strawberry season (early November to late February). Approximately 1300 (RGB and NIR) images were collected in each of the 31 acquisition sessions conducted throughout the season. These images, along with location information for each image, were processed to produce different products such as orthomosaics, and dense 3D point clouds.

# 2.3. Image Pre-Processing

Both the RGB and NIR images acquired in each collection session were processed using the Agisoft Photoscan software (version 1.4) [17] to produce orthomosaic images as well as DSM. Combining the RGB and NIR images in the structure from motion (SfM) analysis increased the density of the point cloud and enabled the creation of spatially co-registered RGB-NIR orthomosaics, as shown in the example view of a 3D, dense point cloud in Figure 4. The Agisoft software, which utilizes SfM analysis, was used to recreate the image acquisition geometry using features matched in the overlapped images [18]. The surveyed ground control points were identified in the captured images, and their coordinates were input to the Agisoft software. This information was processed mathematically by the Agisoft software to produce accurate image location and orientation at the time each image was triggered. Further processing of this information combined with extensive matching of the corresponding features in the overlapped images produced dense 3D point-clouds of image content, which is used to create orthorectified mosaics and DSM.



Figure 4. Dense point cloud of the surface created from overlapping RGB and IR images.

A 2 mm DSM representing the relief of the objects (canopies, plants, and soils) [19] was created for each image acquisition date. Similarly, a 1 mm orthorectified mosaic (orthomosaics) free from the geometric distortion caused by topographic relief and camera tilt [20] was created for each of the RGB and IR bands. The DSM and orthomosaics were then used to derive strawberry vegetation masks and to extract canopy structural properties, such as canopy area and volume using the ESRI ArcMap (v 10.3) [21] software.

Visual interpretation was used to identify and count strawberry fruits and flowers visible in the images. Although seven different fruit development stages were used as described in Figure 5, they were grouped into just two categories (flowers and fruits) for this study (Figure 5). The fruits and flowers were identified in the individual images and their locations on the orthomosaic were marked using ESRI ArcMap software. These point locations were aggregated to compute the flower and fruit counts at the plot level.



Figure 5. Strawberry fruit and flower categorization. Classes 0–2 are combined as flowers, while 3–6 are combined as fruits.

# 2.4. Strawberry Yield, and Flower and Fruit Count Data Collection

Harvesting was performed on the entire plot (24 plants per plot). Only fully ripened fruits were harvested and graded following the United States Department of Agriculture

(USDA) grading standards [22] which produced yield in weight. All data were collected following the same schedule as image acquisition (Monday and Thursday of each week). Flowers and fruits were manually counted in the field and categorized as flowers (categories 0–2) and fruits (categories 3–6) following the classification categories shown in Figure 5. The data were collected from six plants per plot, which were randomly selected and tagged prior to the first counting incidence, with many fruits being counted on multiple data collection days due to the phenological development through categories.

The georeferenced orthorectified images (orthomosaics), produced by mosaicking individual overlapped images taken by the platform in the field, have a resolution that is half of the native resolution of the individual images. The orthomosaic images also often have artifacts resulting from the mosaicking process that can be seen when zooming closer in to the canopy. These factors made flower and fruit identification on the orthorectified images difficult and necessitated the use of the original (before mosaicking) individual images that show the canopy from different directions (multi-view). In this context, the image analysis operator looked at each canopy in the orthorectified image and identified the fruit or flower in at least 4 individual images. The operator then identified the fruits and flowers on the individual images and marked their locations on the orthomosaic image.

### 2.5. Weather Variables

Prior studies have reported that a great number of weather variables influence the rate of physiological progress in all parts of the strawberry plant [4,23,24]. For example, MacKenzie and Chandler found that strawberry flower bud initiation is greatly affected by day length and temperature [2], while Chandler et al. [25] observed that the time from flowering to fruiting would lengthen by 2.2 days in response to a 1 °C decrease in temperature. Kadir et al. [26] also observed the effect of crown temperature on fruit yield. In addition, Crespo et al. [27] reported a close correlation between yield and photosynthetically active radiation. These correlations are prominent at the early-season and late-season growth stages [4]. By contrast, Li et al. [28] indicated negative impacts of solar radiation and air temperature associated with water loss on the responses of cool-weather strawberry plants and, consequently, on fruit formation. Both Li et al. [28] and Pires et al. [29]. observed a yield change in response to soil moisture, implying the effect of rainfall. Relationships between weather conditions and strawberry growth, development, and yield are complex. To maximize the predictive strength of the model, all available weather variables collected from the Florida Automated Weather Network (FAWN) [30] were tested for prediction purposes. Among them, three air temperature variables were calculated from daily measurements from probes 60 cm, 2 m and 10 m above ground (Balm FAWN station, FL). Other variables include soil temperature, relative humidity, rainfall, barometric pressure, solar radiation, wind speed, wind direction, dew point temperature, and evapotranspiration (Table 1).

#### 2.6. Canopy Size Variables Extraction

In this section, the methods and assumptions used to extract the geometrical properties of the canopies are introduced. These methods have also been assessed and discussed in Abd-Elrahman et al. [31]. The ESRI's ArcMap software v10.3 [21] was used to analyze the RGB-IR orthomosaics and DSM resulting from the SfM analysis. The RGB-IR orthomosaic bands were further used to create normalized difference vegetation index (NDVI) [32], shadow ratio [33] and Hue-Intensity saturation bands [34], which were used to create a binary vegetation mask. This mask was created by marking pixels with NDVI, saturation, and shadow ratio values greater than zero, 0.4, and zero, respectively, as vegetation. All other pixels were considered non-vegetation (i.e., soils, plastic bed, paint, etc.). These thresholds were experimented with by manually changing the thresholds and visually inspecting the results to determine the most accepted output. Once the thresholds were identified, they were held constant for all the image acquisition sessions acquired throughout the season. Only pixels marked as vegetation were used to extract canopy size variables. Figure 6 shows the orthomosaic and overlaid vegetation mask created for the data captured on 8 January 2018.

Table 1.	Weather	Variables.

Weather Variables	Variable Description	
Temperature—Air (60 cm)	Average daily temperature from a probe 60 cm above ground (°C)	
Temperature—Air (2 m)	Average daily temperature from a probe 2 m above ground ( $^{\circ}C$ )	
Temperature—Air (10 m)	Average daily temperature from a probe 10 m above ground (°C)	
Soil temperature	Average daily temperature from a probe 4 inches below ground (°C)	
Relative humidity	Average percentage of saturation of a specific volume of air (%)	
Rainfall	Total rainfall (inches)	
Barometric pressure	Pressure within the atmosphere of Earth (millibars)	
Solar radiation	Radiant energy emitted by the sun (watts per square meter)	
Wind speed	Average wind speed (miles per hour)	
Wind direction	Direction from which the wind blows (degrees)	
Dew point temperature	Temperature at which dew starts to form on solid surfaces (°C)	
Evapotranspiration	Sum of evaporation and transpiration (inches per day)	



Figure 6. Vegetation mask shown overlaid on orthomosaic.

Since the images were georeferenced based on centimeter-level GNSS GCPs, manual delineation of the plant and plot boundaries was used to exclude soil area, which was a relatively simple task that did not require intensive labor between one acquisition date and the next. The DSM produced by the SfM analysis provided the height of the surface of the objects shown in the scene above a specific height datum (NAD83 in our analysis). In order to produce canopy heights, a Digital Terrain Model (DTM) layer, representing the ground (plastic bed surface) under the canopies, was created using spatial interpolation. The elevation of the pixels on the soil and beds (excluding canopy pixels) was used to interpolate soil and bed elevation under the canopies. In other words, interpolation was used to fill in the gaps in the DTM under the canopies in order to produce a continuous DTM for the entire bed surface. This process was performed using the ArcMap v10.3 spatial analyst extension. Canopy height model was then computed as the difference between the DSM and DTM. Figure 7 shows 3D point cloud visualization and a canopy height calculation schematic as the difference between DSM and DTM for one of the plots captured on 11 January 2018.



**Figure 7.** (a) Nadir visualization of a point cloud, (b) oblique 3D visualization of a point cloud, and (c) profile showing canopy height as the difference between DSM (top solid line) and DTM (bottom dashed line).

Canopy height above the bed for each pixel of the canopies was then used to extract canopy size variables used in this analysis. Several canopy size variables were computed from the canopy height raster layer for each of the 12 plots established in this experiment. These variables were aggregated per plot and computed for each of the 31 image acquisition dates. Table 2 lists the canopy size variables used in this study and their definitions. Several models developed using the ArcMap v10.3 software model builder were used to automatically compute and export the canopy size variables for all 31 data acquisition sessions.

Canopy Size Variables	Variable Description		
Canopy Average Height	Average canopy height within each plot		
Canopy Height Standard Deviation (std)	Standard deviation of canopy height within each plot		
Canopy Area	Canopy planimetric area within each plot		
Canopy Volume	Canopy volume computed from canopy heights and summarized at the plot level		

Table 2. Canopy Size Variables.

# 2.7. Statistical Analysis Methods

Although many possible predictor variables (e.g., weather and imagery metrics) were available to predict dependent variables of interest in the study, only a few of them are likely to contribute to improved prediction accuracy. Two processes were implemented to select optimal predictor variables. First, a statistical test was conducted to test whether variables had significant effects on counts or yield. Only those that had significant effects were selected as basic variables in prediction models. Second, a variety of prediction models built on these basic variables were compared in terms of prediction accuracy to determine the optimal model with corresponding predictor variables. Linear regression models were used to test variables and develop prediction models. The models were estimated by the least square method.

Linear models were first developed to predict field-observed flower and fruit counts. The comparison between image-derived and field-observed flower and fruit counts shows that fruits and flowers were identified from the imagery very accurately at the early-season growth stage. Due to canopy growth and the increased density throughout the season, however, fewer fruits and flowers became visible from the imagery compared to the established field counts as the season progressed. Therefore, identification of the actual flower and fruit counts is affected by time (e.g., days after planting) and canopy size. As expected, the results of the statistical tests showed that the effects of these variables were significant. In particular, each canopy size variable had a significant effect on one-by-one testing because they were highly correlated. Therefore, time, image-derived counts and four canopy size variables were used as basic variables to predict the observed counts. The general prediction model is expressed as

$$o\_count_t = f(time_t, i_{count_t}, height_t, std_t, area_t, vol_t),$$
(1)

where *o\_count* and *i\_count* are the observed and image-derived counts, respectively, *time* is the days after planting, and *height*, *std*, *area*, and *vol* are canopy average height, canopy height standard deviation, canopy area, and canopy volume, respectively, as explained earlier.

The rolling prediction was adopted to continuously predict next-interval counts (per plot) over the season. That is, data available at the *ith* interval were used to estimate the model for the observed counts at the *ith* interval. Using the estimated equation and available data at the (i + 1)th interval, we calculated the (i + 1)th-interval predictive counts and compared them to the observed counts. The prediction performance is measured by the root mean squared error (RMSE). For predicting the (i + 2)th interval counts, the prediction model was re-estimated with data available at the (i + 1)th interval. The rolling process continued through the last counts of the season. The prediction accuracy of the model over the season was represented by the total value of RMSEs spanning from the first to the last prediction. The general model (1) contained a rich set of models using different combinations of basic predictor variables, e.g., using fewer canopy size variables or using cross or squared terms of basic variables. Each model was examined in terms of the total RMSE over the season. The model with the lowest RMSE was identified as the optimal prediction model.

Next, linear models were developed to predict out-of-sample yields (flat per acre, one flat equals 8 lbs.) at different time intervals (3–4 days ahead of harvest, 1 week ahead of harvest, and 3 weeks ahead of harvest). We considered three general yield prediction models. The first one used time and previous yields as basic predictor variables since strawberry fruit development is characterized by growth waves [35] and has an overall upward trend over time. The test results showed that only the previous two interval yields had significant effects on yields. Therefore, the following first general model was considered

$$yield_t = f(yield_{t-1}, yield_{t-2}, time_t).$$
<sup>(2)</sup>

The second general model assumed that weather information is available. Air temperature, relative humidity, rainfall, barometric pressure, solar radiation, wind speed, wind direction, and evapotranspiration during intervals were found to have effects on yield at the 5% significance level. Incorporating them into the model produced the second general model

$$yield_t = f(yield_{t-1}, yield_{t-2}, time_t, weather_t),$$
(3)

where *weather* is a vector of variables including all weather variables that have significant effects on yields. The first two general models were taken as a benchmark to compare with the third one, which included variables extracted from the images (e.g., canopy size variables and flower and fruit counts)

$$yield_t = f(yield_{t-1}, yield_{t-2}, time_t, weather_t, image),$$
(4)

where *image* is a vector of variables including not only flower counts, fruit counts, and canopy sizes at time *t*, but also previous values of these variables. For each yield prediction model, we repeated the procedure used for predicting the flower or fruit counts. That is, the rolling prediction was adopted to predict next-step yield, and the average RMSE from 12 plots was generated at each step until the final yield prediction of the season. The optimal prediction model was identified by comparing the total RMSEs over the season. The contribution of imagery metrics (image-derived counts and canopy size) in the prediction accuracy was determined by the reduction of RMSEs from the optimal model of (4) to that of (2) or (3).

# 3. Results

# 3.1. Image-Derived and Field-Observed Flower and Fruit Counts

We started by predicting the 6th interval count because enough data must be left out to estimate the model. When imagery metrics were available, the optimal model of predicting flower counts was an equation built on time and its squared term, the image-derived flower count, and canopy volume (model 1.a in Table 3). The total value of RMSEs of prediction over the season was 105, suggesting a percentage prediction error of 26.3% because the actual total number of flower counts over the same period was averaged at 400. Also, the goodness of fit of the model (1) when being estimated with over-the-season data reached 88.2%, which means that 88.2% of the variation of actual flower counts during the season can be explained by the predictor variables.

a.	Flower Count Prediction			
_		<b>Optimal Prediction Model</b>	RMSE of Prediction in Flower Count per Plot	Goodness of Fit
	Model 1.a	$o\_fl_t = \\ \alpha_0 + \alpha_1 \cdot time_t + \alpha_2 \cdot time_t^2 + \alpha_3 \cdot i\_fl_t + \alpha_4 \cdot vol_t + \varepsilon$	105	88.2%
b	b. Fruit Count Prediction			
		<b>Optimal Prediction Model</b>	RMSE of Prediction in Fruit Count per Plot	Goodness of Fit
	Model 1.b	$o_{fr_{t}} = \alpha_{0} + \alpha_{1} \cdot time_{t} + \alpha_{2} \cdot time_{t}^{2} + \alpha_{2} \cdot i  fr_{t} + \alpha_{4} \cdot vol_{t} + \varepsilon$	268	92.6%

Table 3. Count Prediction Models and their Performance.

Note: o\_fl and o\_fr are the observed flower and fruit count while i\_fl and i\_fr are the image-derived flower and fruit count.

The optimal model identified for predicting fruit counts (Table 3, model 1.b) was similar to that for flower count prediction. The actual total fruit counts during the forecast period was 1042, whereas the model generated a RMSE of 268, suggesting a percentage prediction error of 25.7%. The goodness of fit for fruit counts was as high as 92.6%. In summary, these results show that prediction relationships could be built between the image-derived fruit and flower counts and those observed in the field, which provides justifications to use image-derived counts and canopy size variables for yield prediction.

## 3.2. Yield Prediction Based on Imagery, Weather, and Canopy Characteristics

Plot-level yield prediction models were developed using linear regression models relating yield to all available predictor variables, including time, within-season previous yields, weather, image-derived flower and fruit counts, and canopy size variables. Similarly, the RMSE was used to assess prediction accuracy and select predictor variables. A prediction model was first started with easily accessible, basic predictor variables, such as time and previous yields, and then added weather variables, until all imagery metrics were included in the model (Models 2, 3, and 4 in Table 4). Linear regression analysis was implemented to predict yield at different time intervals, including 3–4 days ahead of harvest, 1 week ahead of harvest, and 3 weeks ahead of harvest, and the results are shown in Table 4.

We started by predicting the 13th interval yield at 3–4 days ahead of harvest. The RMSE generated from the optimal prediction model based on previous yields and time was 1222 flats and the goodness of fit reached only 75.1% (Model 2.a in Table 4). Growers often take into account future weather conditions, given that the weather plays an important role in fruit growth. Although many weather variables have statistically significant effects on yield, only air temperature and rainfall contributed to the improvement of prediction accuracy. However, the optimal prediction model, even with actual weather data, generated a RMSE of 1172 flats per acre (Model 3.a in Table 4), only 4% up in prediction accuracy. The prediction accuracy could be even worse since weather forecasts are not always accurate for actual conditions. Nevertheless, the prediction error was large given the total yield over the season was averaged at 2528 flats per acre. The poor yield prediction performance indicates the need for industry or researchers to develop new tools or methods to improve prediction accuracy. Our results show that imagery data, like image-derived fruit counts and canopy volume, were particularly instrumental in improving the prediction performance. The RMSE of prediction was reduced to 866 flats per acre when both image-derived fruit counts and canopy volume were incorporated into the model (Model 4.a in Table 4), which is 29% and 26% lower than that from Model 2.a and from Model 3.a, respectively. The goodness of fit also increased to 83.4%.

a.	Yield Prediction at 3–4 Days Ahead of Harvest				
		Optimal Prediction Model	RMSE of Predictioin in Flat (8lb) per Acre	Goodness of Fit	
	Model 2.a	$yield_{t} = \\ \beta_{0} + \beta_{1} \cdot time_{t} + \beta_{2} \cdot yield_{t-1} + \beta_{3} \cdot yield_{t-2} + \varepsilon$	1222	75.1%	
	Model 3.a	$\begin{aligned} yield_t &= \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot yield_{t-1} + \\ \beta_3 \cdot yield_{t-2} + \beta_4 \cdot tem_t + \beta_5 \cdot rain_t + \varepsilon \end{aligned}$	1172	79.7%	
_	Model 4.a	$\begin{aligned} yield_t = \\ \beta_0 + \beta_1 \cdot yield_{t-1} + \beta_2 \cdot yield_{t-2} + \beta_3 \cdot tem_t + \\ \beta_4 \cdot rain_t + \beta_5 \cdot i\_fr_t + \beta_6 \cdot time_t \cdot i\_fr_t + \beta_7 \cdot vol_t + \varepsilon \end{aligned}$	866	83.4%	
b	Yield Prediction at 1 Week Ahead of Harvest				
		Optimal Prediction Model	RMSE of Prediction in Flat (8lb) per Acre	Goodness of Fit	
_	Model 2.b	$yield_t = \beta_0 + \beta_1 \cdot time_t + \beta_2 \cdot yield_{t-1} + \varepsilon$	1362	88.3%	
	Model 3.b	$yield_{t} = \beta_{0} + \beta_{1} \cdot time_{t} + \beta_{2} \cdot yield_{t-1} + \beta_{3} \cdot tem_{t} + \beta_{4} \cdot rain_{t} + \varepsilon$	1307	89.9%	
	Model 4.b	$yield_{t} = \beta_{0} + \beta_{1} \cdot yield_{t-1} + \beta_{2} \cdot tem_{t} + \beta_{3} \cdot rain_{t} + \beta_{4} \cdot i_{f}r_{t} + \beta_{5} \cdot time_{t} \cdot i_{f}r_{t} + \beta_{6} \cdot vol_{t} + \varepsilon$	1122	92.1%	
c	. Yield Prediction at 3 Weeks Ahead of Harvest				
		Optimal Prediction Model	RMSE of Prediction in Flat (8lb) per Acre	Goodness of Fit	
_	Model 2.c	$yield_t = \beta_0 + \beta_1 \cdot yield_{t-1} + \varepsilon$	1178	95.4%	
	Model 3.c	$yield_t = \beta_0 + \beta_1 \cdot yield_{t-1} + \beta_2 \cdot tem_t + \varepsilon$	1193	95.6%	
	Model 4.c	$yield_t = \beta_0 + \beta_1 \cdot yield_{t-1} + \beta_2 \cdot i_{-}fl_t + \beta_3 \cdot i_{-}fr_{t-1} + \beta_4 \cdot i_{-}fr_{t-2} + \beta_5 \cdot i_{-}fr_{t-3} + \varepsilon$	1055	96.7%	

Table 4.	Yield	Prediction	Models	and	their	Performance.
----------	-------	------------	--------	-----	-------	--------------

Note: *tem* is the average daily temperature from a probe 60 cm above ground (°C) during intervals, *rain* is the rainfall, *i\_fr* is the image-derived fruit count, and *i\_fl* is the image-derived flower count.

The performance of yield prediction at 1 week ahead of harvest was also encouraging. The RMSE of prediction reduced from 1362 flats in the prediction model using only previous fruit yields and time (Model 2.b in Table 4) to 1307 flats using weather variables (Model 3.b in Table 4) to 1122 flats using imagery variables (Model 4.b in Table 4). Imagery metrics contributed to increasing prediction accuracy by 14–18%. Meanwhile, the goodness of fit increased to 92.1%. In MacKenzie and Chandler's [2] prediction equations with hand-collected flower counts as inputs, the goodness of fit was only 89%. Finally, three-week-ahead yield prediction models presented an even stronger fit when exploiting imagery data. The goodness of fit was as high as 97%, and the RMSE was reduced by 10% in Model 4.c compared to Model 2.c (Table 4). Note that image-derived flower and fruit counts at previous harvest times were used in Model 4 because the ripe strawberries were normally ready for harvest several weeks after plants blossomed.

# 4. Discussion

This study demonstrated the feasibility of using high temporal resolution imagery to extract canopy size information as well as fruit and flower counts to predict strawberry yield along the strawberry season. The results show significant improvement in prediction accuracy when compared to the models using only previous yield as adopted in current practices.

Image capturing was implemented by mounting the cameras on farm equipment, which could be easily integrated into standard farm operations. Although we used a survey-quality GNSS to process the data used in this study, we experimented using a lower

costs system that does not use the GNSS data from the mobile platform and achieved the same results counting only on ground control points established in the beginning of the season. This leaves the image acquisition and triggering system, which costs less than \$1500 as the main investment to acquire the data used in this study. Such cost is considered a small overhead that can be afforded by all growers.

Canopy information extraction was performed automatically using geospatial analysis models. Although these models can analyze thousands of plants in a few hours [31], we believe that the models have the potential to be served as a server service for web clients as a step towards commercial implementation. Flower and fruit counting was conducted manually to allow visual identification of the flowers and fruits on images captured by the platform from different directions (multi-view) and projecting them to the orthorectified mosaic. This process allowed for detecting more of the flowers and fruits hidden under the canopy compared to using the orthorectified mosaic image. A deep learning algorithm utilizing Multiview flower and fruit detection and counting is being developed by the authors. Integrating strawberry prediction models in large scale field operations can be achieved by the individual growers if special geospatial analysis expertise are utilized. However, small farms would probably benefit from strawberry yield prediction provided as a subscription-service by specialized vendors.

The RMSEs of the prediction models were generally high. The smallest RMSE was achieved for the 3–4-day prediction model. This model, however, was achieved after incorporating the canopy image metrics. Incorporating the canopy image metrics variables derived from imagery substantially improved the prediction accuracy for all models by 10–29%, which highlighted the importance of using the imaging technologies in strawberry yield prediction.

Our results show that statistical model prediction with imagery data and other already available weather and previous yield data matched the measured yields within a margin that could be a first step towards strawberry prediction models that help growers plan their harvest and marketing operations. Nevertheless, we believe that incorporating the strawberry physiological parameters is essential in the next stage of strawberry model prediction. Models that incorporate the strawberry yield waves that could be associated with climate conditions and genotypes are essential to achieve improved prediction accuracy. The effect of pest and disease stressors could also be incorporated in the prediction model through spectral image information.

# 5. Conclusions

The data and results analyzed in this study provide strong evidence that close-range high-resolution images captured in the field throughout the strawberry season could be a valuable tool for strawberry yield prediction at different time scales, which could be a valuable asset for strawberry farm management and marketing. Canopy size variables extracted from the acquired images such as canopy area, volume, height standard deviation, and fruit and flower counts visually interpreted from the images were used to predict actual flower and fruit counts with percentage prediction errors of 26.3% and 25.7%, respectively. Similarly, this study demonstrates the feasibility of developing statistical strawberry yield prediction models at different time intervals (3–4 days ahead of harvest, 1 week ahead of harvest, and 3 weeks ahead of harvest) using canopy size variables as well as flower and fruit counts, weather variables, and previous yield data. The rolling out-of-sample prediction method shows that prediction accuracy from models with image-derived variables could be increased by 10–29% compared to those without these variables, implying the importance of imagery information to yield prediction.

Author Contributions: Conceptualization: Amr Abd-Elrahman, Shinsuke Agehara, Feng Wu, and Katie Britt; Canopy Metric Extraction Software Development: Amr Abd-Elrahman; Statistical Analysis: Feng Wu; Results Validation: Shinsuke Agehara; Investigation: Amr Abd-Elrahman, Shinsuke Agehara, Feng Wu, and Katie Britt; Data Curation: Amr Abd-Elrahman and Katie Britt; Writing— Original Draft Preparation: Amr Abd-Elrahman; Writing—Review and Editing: Amr Abd-Elrahman, Shinsuke Agehara, Feng Wu, and Katie Britt; Visualization: Katie Britt; Project Administration: Amr Abd-Elrahman; Funding Acquisition: Amr Abd-Elrahman, Shinsuke Agehara, Feng Wu, and Katie Britt. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by a grant from the Florida Strawberry Research and Education Foundation (FSREF).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The datasets generated during and/or analyzed during the current study are not publicly available due to their continued use in ongoing, unpublished research. The code generated and used during and/or during the current study is available from the corresponding author on reasonable request.

Acknowledgments: We would like to thank Ali Gonzalez Perez, Weining "Bill" Wang, and other members of Horticultural Crop Physiology Lab at Gulf Coast Research and Education Center for their technical assistance. This study was supported by a grant from the Florida Strawberry Research and Education Foundation (FSREF).

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

# References

- 1. Sanz, C.; Pérez, A.G.; Olías, R.; Olías, J.M. Quality of Strawberries Packed with Perforated Polypropylene. *J. Food Sci.* **1999**, *64*, 748–752. [CrossRef]
- 2. MacKenzie, S.J.; Chandler, C.K. A Method to Predict Weekly Strawberry Fruit Yields from Extended Season Production Systems. *Agron. J.* 2009, 101, 278–287. [CrossRef]
- 3. Tuovinen, T.; Parikka, P. Monitoring Strawberry Pests and Diseases: A Method to Estimate Yield Losses. *Acta Hortic.* **1997**, 439, 941–946. [CrossRef]
- 4. Palencia, P.; Martinez, F.; Medina, J.J.; López-Medina, J. Strawberry Yield Efficiency and Its Correlation with Temperature and Solar Radiation. *Hortic. Bras.* 2013, *31*, 93–99. [CrossRef]
- 5. Pathak, T.B.; Dara, S.K.; Biscaro, A. Evaluating Correlations and Development of Meteorology Based Yield Forcasting Model for Strawberry. *Adv. Meteorol.* 2016, 2016, 1–7. [CrossRef]
- 6. Guan, Z.; Wu, F.; Whidden, A. Top Challenges Facing the Florida Strawberry Industry: Insights from a Comprehensive Industry Survey. *EDIS* **2016**, *2*, 3.
- Misaghi, F.; Dayyanidardashti, S.; Mohammadi, K.; Ehsani, M.R. Application of Artificial Neural Network and Geostatistical Methods in Analyzing Strawberry Yield Data. In Proceedings of the 2004 ASAE Annual Meeting, Ottawa, ON, Canada, 15 August 2004.
- 8. Bartczak, M.; Lisiecka, J.; Knaflewski, M. Correlation between Selected Parameters of Planting Material and Strawberry Yield. *Folia Hortic.* **2010**, *22*, 9–12. [CrossRef]
- Lopez, A.; Perez, C.; Arias, A.; Palanco, J.; Gomez, A.; Torres, M.; Rodriguez, M. Strawberry Fruit Yield Forecast Based on Montecarlo Methodology and Artificial Vision. In Proceedings of the 7th International Strawberry Symposium, Beijing, China, 12 January 2019; pp. 551–552.
- 10. Poling, E.B. Strawberry Plant Structure and Growth Habit. In Proceedings of the Empire State Producers Expo, New York, NY, USA, 12–15 January 2012.
- Jones, J.W.; Antle, J.M.; Basso, B.; Boote, K.J.; Conant, R.T.; Foster, I.; Godfray, H.C.J.; Herrero, M.; Howitt, R.E.; Janssen, S.; et al. Toward a New Generation of Agricultural System Data, Models, and Knowledge Products: State of Agricultural Systems Science. *Agric. Syst.* 2017, 155, 269–288. [CrossRef]
- 12. Xu, Y.; Smith, S.E.; Grunwald, S.; Abd-Elrahman, A.; Wani, S. Effects of Image Pansharpening on Soil Total Nitrogen Prediction Models in South India. *Geoderma* **2018**, *320*, 52–66. [CrossRef]
- 13. Zheng, C.; Abd-Elrahman, A.; Whitaker, V. Remote Sensing and Machine Learning in Crop Phenotyping and Management, with and Emphasis on Applications in Strawberry. *Remote Sens.* **2021**, *13*, 531. [CrossRef]
- 14. Guan, Z.; Abd-Elrahman, A.; Fan, Z.; Whitaker, V.M.; Wilkinson, B. Modeling Strawberry Biomass and Leaf Area Using Object-Based Analysis of High-Resolution Images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *163*, 171–186. [CrossRef]
- 15. Abd-Elrahman, A.; Sassi, N.; Wilkinson, B.; Dewitt, B. Georeferencing of Mobile Ground-Based Hyperspectral Digital Single-Lens Reflex Imagery. J. Appl. Remote Sens. 2016, 10, 014002. [CrossRef]
- 16. Abd-Elrahman, A.; Pande-Chhetri, R.; Vallad, G. Design and Development of a Multi-Purpose Low-Cost Hyperspectral Imaging System. *Remote Sens.* **2011**, *3*, 570–586. [CrossRef]
- 17. Agisoft, L.L.C. Agisoft PhotoScan User Manual: Professional Edition. Available online: https://www.agisoft.com/pdf/photoscan-pro\_1\_4\_en.pdf (accessed on 15 May 2020).

- 18. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle Adjustment—A Modern Synthesis. *Int. Work. Vis. Algorithms* **1999**, *1883*, 298–372.
- 19. Caliper Mapping and Transportation Glossary. DEM. 1991. Available online: https://www.caliper.com/glossary/default.htm (accessed on 15 May 2020).
- Smith, G. Digital Orthophotography and GIS. In Proceedings of the 1995 ESRI User Confernce, Palm Springs, CA, USA, 22–26 May 1995.
- 21. ArcMAP, v.10.3; ESRI: Redlands, CA, USA, 2014.
- 22. USDA. United States Standards for Grades of Strawberries. 2006. Available online: https://www.hort.purdue.edu/prod\_quality/ quality/strawber.pdf (accessed on 15 May 2020).
- Fernandez, G.E.; Butler, L.M.; Louws, F.J. Strawberry Growth and Development in an Annual Plasticulture System. *HortScience* 2001, 36, 1219–1223. [CrossRef]
- Lobell, D.B.; Cahill, K.N.; Field, C.B. Historical Effects of Temperature and Precipitation on California Crop Yields. *Clim. Chang.* 2007, *81*, 187–203. [CrossRef]
- 25. Chandler, C.K.; MacKenzie, S.J.; Herrington, M. Fruit Development Period in Strawberry Differs among Cultivars, and Is Negatively Correlated with Average Postbloom Air Temperature; Florida State Horticultural Society: Daytona Beach, FL, USA, 2004; Volume 117.
- 26. Kadir, S.; Carey, E.; Ennahli, S. Influence of High Tunnel and Field Conditions on Strawberry Growth and Development. *HortScience* **2006**, *41*, 329–335. [CrossRef]
- Crespo, P.; Ançay, A.; Carlen, C.; Stamp, P. Strawberry Cultivar Response to Tunnel Cultivation. *Acta Hortic.* 2009, 838, 77–82.
   [CrossRef]
- 28. Li, H.; Li, T.; Gordon, R.J.; Asiedu, S.K.; Hu, K. Strawberry Plant Fruiting Efficiency and Its Correlation with Solar Irradiance, Temperature, and Reflectance Water Index Variation. *Environ. Exp. Bot.* **2010**, *68*, 165–174. [CrossRef]
- 29. Pires, R.C.M.; Folegatti, M.V.; Passos, F.A.; Arruda, F.B.; Sakai, E. Vegetative Growth and Yield of Strawberry under Irrigation and Soil Mulches for Different Cultivation Environments. *Sci. Agric.* 2006, *63*, 417–425. [CrossRef]
- University of Florida. Florida Automated Weather Network. Available online: https://fawn.ifas.ufl.edu (accessed on 15 May 2020).
- Abd-Elrahman, A.; Guan, Z.; Dalid, C.; Whitaker, V.; Britt, K.; Wilkinson, B.; Gonzalez, A. Automated Canopy Delineation and Size Metrics Extraction for Strawberry Dry Weight Modeling Using Raster Analysis of High-Resolution Imagery. *Remote Sens.* 2020, 12, 3632. [CrossRef]
- 32. Crippen, R.E. Calculating the Vegetation Index Faster. Remote Sens. Environ. 1990, 34, 71–73. [CrossRef]
- Sirmacek, B.; Unsalan, C. Damaged Building Detection in Aerial Images Using Shadow Information. In Proceedings of the 2009 4th International Conference of Recent Advances in Space Technologies, Zeytinburnu, Turkey, 1–4 June 2021; pp. 249–252.
- Joblove, G.H.; Greenberg, D. Color Spaces for Computer Graphics. In Proceedings of the 5th annual conference on Computer graphics and interactive techniques, New York, NY, USA, 12–19 August 1978; pp. 20–25.
- 35. Wu, F.; Guan, Z.; Whitaker, V. Optimizing Yield Distribution under Biological and Economic Constraints: Florida Strawberries as a Model for Perishable Commodities. *Agric. Syst.* **2015**, *141*, 113–120. [CrossRef]