Statistical Modeling of Strawberry Fruit Yield Using Weather and Image Information (Year 1)

Amr Abd-Elrahman, Shinsuke Agehara, Feng Wu, and Katie Britt

Summary

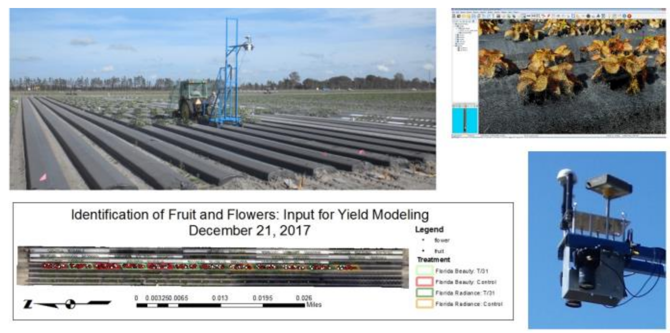
The main goal of this project is to develop and test different strawberry yield prediction models. Strawberry prediction models using flower and fruit counts, canopy area, and weather data produced the best prediction models for the 'Florida Radiance' and 'Florida Beauty' cultivars. Prediction models using historical harvest data from five farms as well as weather and satellite image data were also developed. Using image-derived flower and fruit counts produced results similar to the models utilizing field counts, which encourages image utilization in yield prediction.

Strawberry Yield Prediction

Statistical models were developed to predict strawberry yield to aid in marketing and operational decisions across the season. The project used information derived from field images as well as weather and previous yield data to develop yield prediction models. The project also utilized long-term farm harvest records, weather and satellite image data to develop regional yield prediction models.

Methods

Analysis was performed on two datasets. The first dataset was collected (Fig. 1) from twelve plots of two cultivars ('Florida Radiance' and 'Florida Beauty') at the GCREC. A total of 5 beds were imaged (around 1000 images) approximately two times per week (31 acquisition sessions) through the whole strawberry season (11/02/2017 – 03/01/2018). The images were analyzed to extract canopy area of each plot as well as visual flower and fruit counts. Field observations were also collected from 6 plants per plot to provide actual flower and fruit counts and to document the physiological development cycle of the fruit. Yield and weather data were also used in the analysis.



Clockwise from top left: imaging platform collecting data in the field, 3d point cloud model of strawberry canopy viewed in Agisoft Photoscan, imaging platform showing GPS and camera suite, intermediate step in modeling process showing input of fruit and flowers for specific dates and treatments on a specific date as input for yield model.

Figure 1. Ground based image collection using a tractor-trailed platform.

The second dataset analyzed in this study consists of historical harvest, weather, variety, and satellite image data, collected for 5 farms in West Central Florida. Statistical modeling was conducted to (1) model image-derived vs field-observed fruits and flowers, (2) predict yield using plot-level fruit/flower count, canopy size data, and weather data using the first dataset, and (3) predict yield using historical farm harvest, weather, and satellite image data of the second dataset.

Results

1. Image-derived vs field-observed strawberry flower and fruit counts

The precision of image-derived vs. field-observed counts for number of flowers was 60-133% for 'Florida Radiance' and 67-130% for 'Florida Beauty', and for number of fruits was 49-60% for Radiance and 52-74% for Beauty (Table 1). Image-derived counts were generally more precise for flowers than for fruits. For both cultivars, the precision of image-derived fruit counts declined over time, as canopy size increased. Image-derived fruit counts were slightly more precise for Beauty than for Radiance, probably because of its compact canopy size allowing more fruits to be exposed.

Table 1. Image and field flower and fruit count comparison

Cultivar	Variable	Precision of image-derived vs field-observed counts (%) 11/20	11/27	12/11	12/27	1/11
Radiance	Flower	100	117	133	140	60
	Fruit	60	50	49	48	50
Beauty	Flower	67	130	105	75	80
	Fruit	74	55	56	52	52

2. Plot-level yield prediction models using close range image data

The plot-level yield prediction models are described by the linear equations relating yield to various predicting factors, including within-season previous yield, weather, actual and visual flower and fruit counts, and canopy area. The past multiple periods' (lagged) observations of these factors are used to predict the current yield for the two cultivars. The stepwise regression method is used to fit regression models and choose predictive variables that yield the highest predictive power. The predictive power is assessed by the R-squared, which measures how close the predicted values are to the actual values. Table 2 presents the performance of prediction models using different predictor variables.

Table 2. Performance of various prediction models using close range image data

Prediction Model	**R-squared**		
	Radiance	Beauty	Aggre-gate
Previous yield	10.52%	0.20%	4.41%
Weather	58.50%	74.11%	58.12%
Canopy, nflower & nfruit	52.11%	43.34%	31.50%
Canopy, iflower & ifruit	48.18%	51.02%	36.27%
Weather & iflower & ifruit & Canopy	69.36%	75.43%	61.98%

The results show that the past yield has little predictive power, while weather factors can play an important role in prediction, in particular, for 'Florida Beauty'. These weather factors include air temperature, soil temperature, humidity, rainfall, wind speed, and so on. Additionally, the actual fruit and flower counts and canopy size can predict yield with good accuracy for each cultivar. The close-range image technology can identify fruit and flowers and the prediction models using visual flower and fruit counts, and canopy size provided similar predictive powers. The predictive power of the models using both weather and image data is substantially improved for 'Florida Radiance', compared to those only using weather data, while the improvement is slight for 'Florida Beauty'. These models will be fine-tuned by incorporating more flower and fruit categories along the growth cycle in future research.

3. Yield prediction models using historical farm data

The models using historical farm data are used to predict yield at a one-week interval. Predictor variables include historical yield (yield at the prior 1-week interval and yield at the same time in previous years), weather, and satellite imagery. The MODIS satellite images available in the past 15 years are used to derive the enhanced vegetation index (EVI). Unlike the plot level experiment's results, results using farm data show that historical yield can predict yield very well throughout the season (Table 3). The added weather data can improve predictive power from 67% to 78%. However, adding EVI improves prediction accuracy only slightly. The results are similar for early yield prediction. More farm and satellite image data will be tested in future research.

Table 3. Performance of various prediction models using historical farm data

Prediction Model	R-squared	
	Whole-Season	Early Yield
Historical	67%	58%
Weather & Historical	78%	70%
Weather & EVI & His.	79%	70%

Contact

Dr. Amr Abd-Elrahman
UF/IFAS GCREC – Plant City
P: 813-757-2283
E: aamr@ufl.edu