Statistical Modeling of Strawberry Fruit Yield Using Weather and Image Information (Year 2)

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

Summary

Field trials conducted over two consecutive seasons (2017-2018 and 2018-2019) were used to develop strawberry yield predictive models for 'Florida Radiance' and 'Florida Beauty'. Model drivers such as canopy properties (i.e. area, volume, average height, and standard deviation of height), weather data, and flower and fruit counts (only 2017-2018 dataset) were identified. Three yield predictive models for next-harvest, next-week, and three-weeks ahead were developed from the drivers. All models utilizing previous harvest, time trend, flower and fruit counts, canopy variables, and weather data provided a goodness of fit exceeding 96%. Dropping the flower and fruit counts from the model (2017-2018) resulted in a 94% goodness of fit for next harvest yield.

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 predictive models.

Methods

Analysis was performed on two datasets collected during the 2017-2018 and 2018-2019 seasons. Each dataset was collected from twelve plots of 'Florida Radiance' and 'Florida Beauty' at the GCREC. Five beds were imaged (about 1000 images) approximately twice a week (~28 acquisition sessions per season, i.e. early November to late February). The images were analyzed to extract canopy geometrical properties such as canopy area, volume, average height, and standard deviation of height for each plot as well as visually identified flower and fruit counts (only the 2017-2018 season). Field observations were also collected from six plants per plot to provide actual flower and fruit counts and to document the physiological fruit development cycle. Weather data were acquired from the Florida Automated Weather Network. Statistical modeling was conducted to (1) model image-derived vs field-observed fruit and flowers, and (2) predict first-season yield using fruit/flower counts, canopy variables, and weather data.

Results

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

Models to estimate fruit counts from image counts, time trend, and canopy variables were developed using the 2017-2018 dataset (Figure 1). Model goodness-of-fit was 91.7% when canopy variables were not used. Adding canopy variables improved the goodness-of-fit to 94.1%. Similarly, the goodness-of-fit of flower counts was 82.6% with canopy variables and 83.8% after adding canopy variables to the model (Figure 2).

Fruit Count Prediction Model				
Predictors	Flower imagery count	Time	Canopy variables	Goodness of Fit
Model 1	✓	✓		91.7%
Model 2	✓	✓	✓	94.1%

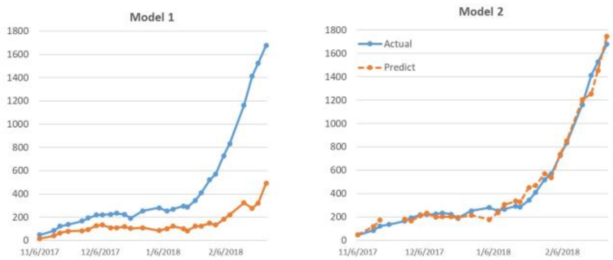

Figure 1. Fruit Count Estimation Model

Flower Count Prediction Model				
Predictors:	Flower imagery count	Time	Canopy variables	Goodness of Fit
Model 1	✓	✓		82.6%
Model 2	✓	✓	✓	83.8%

Figure 2. Flower Count Estimation Model

2. Plot-level yield predictive models

Plot-level yield predictive models were developed using linear regression relating strawberry yield (harvest weight) to various predictive factors, including within-season previous yield, weather, image-derived flower and fruit counts, and canopy variables. These factors were used to predict the current yield for the two cultivars used in the study. Stepwise regression was used to fit regression models and choose predictive variables that provided the highest predictive power. The predictive power was assessed by the goodness-of-fit statistics, which measure how close the predicted values are to the actual values.

Figures 3, 4, and 5 show the results of the yield prediction at next-harvest, next-week, and three-week ahead intervals. The figures also tabulate the goodness-of-fit associated with using different combinations of drivers. The best models were achieved when lagged harvest yields, time trend, lagged image derived flower and fruit counts, canopy variables, and weather data were used.

Predictors	Last four harvested yields	Last one harvested yield	Time	Last three manual fruit counts	Last four imagery fruit counts	Canopy variables	Weather	Goodness of Fit
Model 1	✓		✓					82.7%
Model 2		✓	✓	✓				88.4%
Model 3		✓	✓		✓	✓		94.1%
Model 4		✓	✓		✓	✓	✓	96.2%

Figure 3. Next-Harvest Yield Predictive Model

Predictors	Last four weekly yields	Time	Last two manual fruit counts	Last two imagery fruit and flower counts	Canopy variables	Weather	Goodness of Fit
Model 1	✓	✓					90.1%
Model 2	✓	✓	✓				94.2%
Model 3	✓	✓		✓	✓		96.5%
Model 4	✓	✓		✓	✓	✓	97.2%

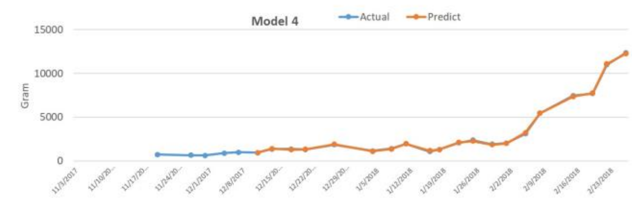

Figure 4. Next-Week Yield Predictive Models

Predictors	Last four weekly yields	Time	Last two manual flower counts	Last one manual fruit count	Last four imagery flower counts	Last two imagery fruit counts	Canopy variables	Weather	Goodness of Fit
Model 1	✓	✓							92.0%
Model 2	✓	✓	✓	✓					95.4%
Model 3	✓	✓			✓	✓	✓		97.8%
Model 4	✓	✓			✓	✓	✓	✓	98.0%

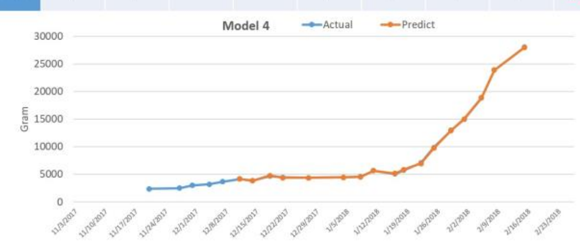

Figure 5. Three-Week Ahead Yield Predictive Models

Conclusion

Adding image-derived flower and fruit counts, canopy geometrical properties and weather data significantly improved yield predictive models. This applies to variable prediction intervals that range from next harvest (2-3 days out) to three weeks ahead. Our results also show that it is possible to achieve accurate flower and fruit counts using image-derived counts, time trend, and canopy variables. Developing variety-dependent and early season predictive models is recommended for future research.

Contact

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