UAV Multispectral Data as a Suitable Tool for Predicting Size and Yield of Vidalia Onions

Marcelo Barbosa[1], Regimar dos Santos[1], Lucas Sales[1], Ronega Vargas[1], Ted McAvoy[1], Angelos Deltsidis[1], Christoper Tyson[2], Aubrey Shirley[2], Luan Oliveira[1]

[1] Department of Horticulture, University of Georgia, Tifton, GA, USA.
[2] Vidalia Onion & Vegetable Research Center, University of Georgia, Tifton, GA, USA.

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.
Vidalia onions is a specialty crop cultivated solely within the southeastern region of Georgia. The key distinguishing characteristic of Vidalia onions is its high sugar content, making them highly prized and widely consumed. Ten thousand acres are grown with Vidalia Onions each year approximately, and the market value (~$150Mi/year) makes the crop very important for the State of Georgia. Traditionally, the planting, weeding, spraying, harvesting, and post-harvesting operations are usually done using manual methods. One of those practices involve classifying the onions size only after it has been harvested, which is a labor-intensive and time-consuming process. This approach also lacks the ability to prevent any factors that can compromise the quality of the onions (e.g., insect infestation, diseases, and stress). In light of these challenges, the present project had the main objective of using a multispectral imagery data approach to predict the size and yield of Vidalia onions. This approach is non-invasive and non-destructive. Data collection occurred from one month after transplanting through the harvest date, resulting in a temporal dataset covering eight different dates. The dataset consisted of four spectral bands (Green, Red, RedEdge, and NIR) and several vegetation indices. Subsequently, we conducted correlations and developed machine learning (ML) models to establish a more robust relationship with the crop. The results demonstrated significant and timely correlations. The ML models excelled in predicting the yield of Vidalia onions well in advance of the harvest date.

Keywords.
Multispectral Imagery, predictive models, machine learning, market class.

Introduction

Multispectral data have proven to be a valuable tool for supporting agricultural decision-making due to their ability to be acquired quickly, accurately, and representatively. Numerous studies have utilized multispectral data for various purposes, such as monitoring crop growth, detecting diseases, and predicting yield. Yield prediction, in particular, stands out as a crucial parameter, supporting in proactive decision-making and identifying yield even before harvest. This capability supports decision-making by identifying field viability and scheduling the optimal time and location for harvest. Then, given the significance of Vidalia onions and the lack of research on yield prediction for this specific crop, we propose to develop predictive models to predict onion yield both overall and by market classes.

Material and Methods

The study was conducted in two commercial Vidalia onion fields located near Glennville and Cobbtown, both cities in Georgia, USA. The onions were planted in December 2023, and data collection began in January 2024, continuing every other week until April 2024, when the onions were harvested. Data collection was carried out using unmanned aerial vehicle (UAV) flights. The UAV (DJI Mavic 3 Multispectral, Shenzhen, China) captured images in four spectral bands: Green, Red, RedEdge, and NIR. In total, we obtained six sets of images for each field. At the end of the season, we collected 50 samples from each field to constitute our dataset. For each sample, we measured the weight and size of 20 onions to later classify them into medium, jumbo, and colossal categories. Both fields were used as sources for our predictive model, with a 70% and 30% split for training and testing, respectively. Various algorithms were used to train the model, but we determined that best-performing algorithm for this study was the random forest. Finally, the predictive model was validated using the R^2, MAE and RMSE metrics.

Results

Clearly, using spectral data, such as spectral bands (4 - G, R, RE, and NIR) and vegetation indices (12 - SCI, NGRDI, TVI, RVI, PSRI, …), presents a timely opportunity for predicting onion parameters. However, simple methods like correlation analysis are not always feasible. Notably, the correlations between spectral data and onion parameters were weak (Figure 1). This supports the need for more complex statistical methods to optimize data analysis and accurately develop predictive models. Particularly, only indivudial correlations between spectral data and onion parameters were strong.

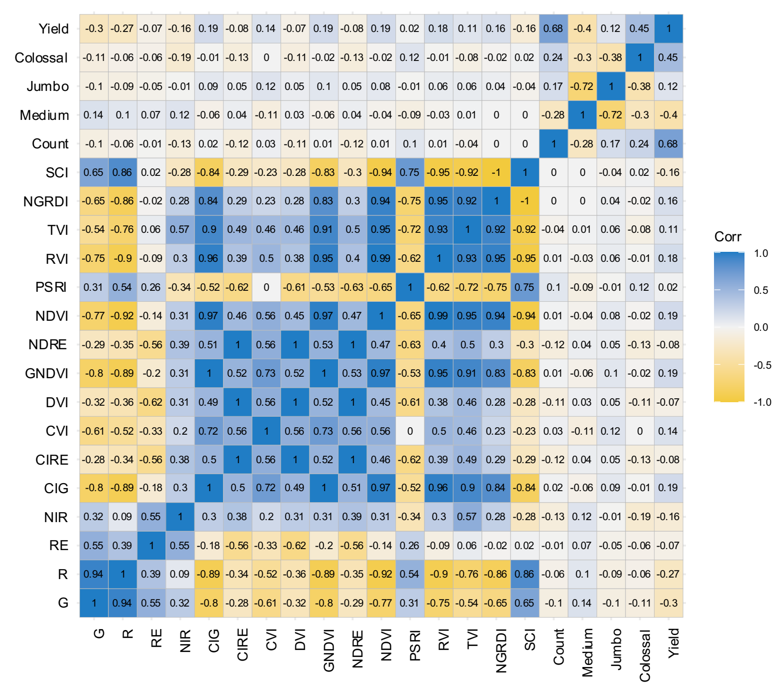

Figure 1. Pearson's correlation between predictive inputs and onions parameters.

Conversely, investing more time in developing robust algorithms could yield more realistic and accurate results. Such algorithms are well-suited for handling high-complexity datasets and overcoming non-linear issues. For instance, utilizing random forest models could improve the accuracy of onion yield predictions (Figure 2). Our findings indicate that predicting onion yield 90 days before harvest produces promising results (R^2 = 0.6, MAE = 3816 kg/ha, RMSE = 4732 kg/ha). Also, combining predictions from the best dates (90 and 15 days before harvest) enhances the prediction accuracy even further (R^2 = 0.7, MAE = 3787 kg/ha, RMSE = 4807 kg/ha).

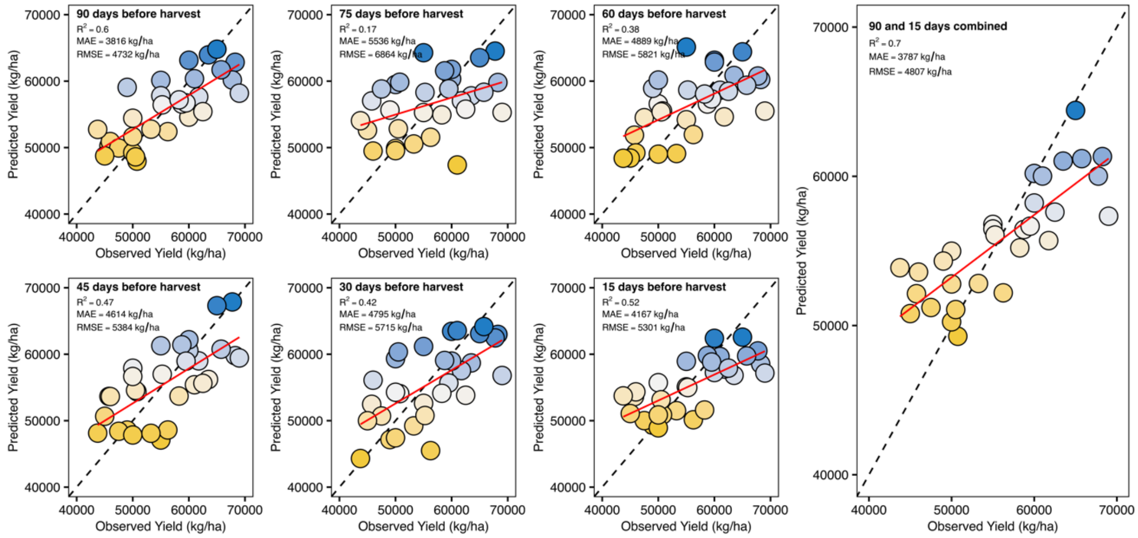

Figure 2. Random forest algorithm to predict onion yield. Predictions occurred 90, 75, 60, 45, 30, and 15 days before harvest. Points merging from yellow to blue represent low and high values, respectively.

On the other hand, predicting yield by market classes is more challenging than the overall yield (Figure 3). An interesting trend can be observed in the medium class, which appears to be more feasible to predict accurately. Conversely, further analysis will be conducted to improve the predictive ability for all classes. In short, all predictions achieved R² < 0.52, with MAE and RMSE values ranging between 7% and 17%. Similar to the overall yield prediction, the best results for market classes were obtained 90 days before harvest. Also, we combined the best data collection dates to develop a better model, but this approach did not produce satisfactory results.

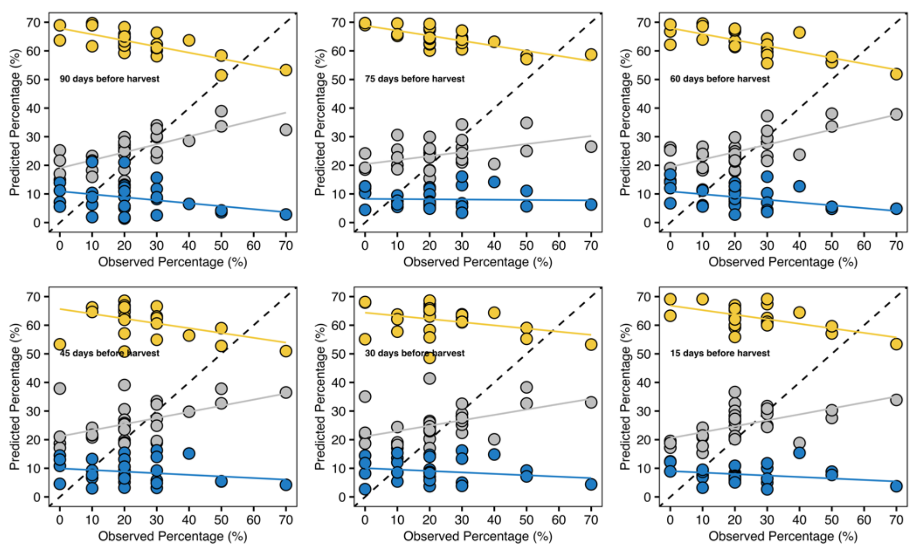

Figure 3. Random forest algorithm to predict onion yield by market classes. Predictions occurred 90, 75, 60, 45, 30, and 15 days before harvest. Points merging from yellow to blue represent low and high values, respectively.

Conclusion

Using multispectral data for predicting yield presents a timely opportunity to develop a non-invasive, non-destructive, and scalable framework. In this study, we were able to accurately predict the overall yield. Conversely, predicting yield by market class produced moderate performance, highlighting an area for further investigation. Therefore, this study supports the application of technological advancements in the field of onion cultivation.

Acknowledgements

We would like to express our profound gratitude to the Precision Horticulture Laboratory of the Department of Horticulture at the University of Georgia, Tifton Campus.