Snap Bean Flowering Detection from UAS Imaging Spectroscopy

E. W. Hughes [1], S. J. Pethybridge [2], J. R. Kikkert [3], C. Salvaggio [1], J. van Aardt [1]

[1] Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY

[2] The New York State Agricultural Experiment Station, Cornell University, Geneva, NY

[3] Cornell Vegetable Program, Cornell Cooperative Extension, Canandaigua, NY

A paper from the Proceedings of the

14th International Conference on Precision Agriculture

June 24 – June 27, 2018; Montreal, Quebec, Canada

Abstract. Sclerotinia sclerotiorum *(white mold) is a fungus that infects the flowers of snap beans and causes a reduction in the number of pods, and subsequent yields, due to premature pod abscission. Snap bean fields typically are treated with prophylactic fungicide applications to control white mold, once 10% of the plants have at least one flower. The holistic goal of this research is to develop spatially-explicit white mold risk models, based on inputs from remote sensing systems aboard unmanned aerial systems (UAS). The objectives of this study are to i) identify spectral signatures for the onset of flowering towards optimal timing of the fungicide application and ii) investigate spectral characteristics of white mold onset in snap bean, for eventual inclusion in the risk models. This paper concentrates on the first objective. The study area was located at Cornell University, Geneva, NY, USA. A DJI Matrice-600 UAS, boasting a high spatial resolution color (RGB) camera, a Headwall Photonics Nano-imaging spectrometer (272 bands; 400-1000 nm), and a Velodyne VLP-16 light detection and ranging (LiDAR) system, was utilized to collect the data. High frequency flights were flown around days when various portions of the snap bean fields were beginning to flower. The imaging spectroscopy data were first ortho-rectified and then mosaicked using GPS/IMU (inertial measurement unit) information from the UAS. The imagery was calibrated into reflectance data using the empirical line calibration method, based on in-field black/white calibration panels. Samples of flowering and non-flowering snap bean spectra were selected from the hyperspectral imagery, followed by single feature linear discriminant analysis to determine which ratio indices, normalized difference indices, and wavelengths critical for discriminating between flowering and non-flowering plants. Next, the features with the highest c-index trained linear discriminant, logistic regression, and support vector machine models. Results showed that the linear discriminant model had the highest test accuracy of 93%, 95%, and 92% for 20, 10, and 3 features, respectively. These results are promising for eventual implementation in disease risk models.*

Keywords. *Unmanned Aerial Systems, White Mold, Snap Beans, Hyperspectral Data Analysis.*

Introduction

Snap beans are the fifth largest vegetable crop in the United States in terms of acres planted (USDA 2012), with harvests valued at 416 million dollars in recent years (USDA 2015). New York State ranks second for both processing and fresh market snap beans in terms of acres planted (USDA 2014). The most important disease affecting snap bean production is white mold caused by the fungus, *Sclerotinia sclerotiorum*. The disease causes reductions in the number of pods, from premature pod abscission and contamination from mycelia. White mold control is centered on the prophylactic application of fungicides initiated when 10% of snap bean plants have at least one flower. Protection of flowers is important because *S. sclerotiorum* ascospores can only infect the flowers. Research by Lehner *et al*. (2017) has identified suboptimal timing of fungicides as the primary reason for poor disease control. Tools to assist in detection of early flowering may therefore assist farmers in improving timing of fungicides and optimizing already available tools for optimal disease control.

Multivariate analysis and machine learning have potential to partition imagery into classes using spectroscopy. For example, Rumpf *et al*. (2010) trained a support vector machine (SVM) to classify Cercospora leaf spot, powdery mildew and leaf rust on sugar beets. An SVM is a machine learning technique that finds a hyperplane that maximizes the separation between classes. Classification results varied from 85.7 to 96.5% depending on the disease. Delalieux *et al.* (2007), on the other hand, used logistic regression on a per-wavelength basis to identify wavelengths that best discriminated between diseased and healthy apple trees (Braeburn cultivar) for apple scab (*Venturia inaequalis*). Logistic regression is a direct probability model that estimates the probability of a binary response based on one or more features. The authors found solid predictability (c-values > 0.8) when classifying diseased plants based on such supervised classification techniques. Studies performed by Rumpf and Delalieux concentrated on classifying diseased plants, but the discriminating methods used can be applied to any problem that needs to separate classes, like flowering and non-flowering snap beans.

Research by Rumpf and Delalieux made extensive use of ground-based spectroradiometers for spectral data collection. Until recently, airborne hyperspectral platforms have been unable to collect hyperspectral imagery at a sufficiently fine spatial resolution to effectively classify spectra of diseased plants. However, with the advent of unmanned aerial systems (UAS), mounted with hyperspectral platforms now boast centimeter-scale ground sampling distances (GSD). This capability, past research efforts, and the need for optimized, variable rate fungicide applications have all laid the foundation for studies such as this one.

The overall objectives of our research are to i) identify spectral signatures for the onset of flowering to optimally time the application of fungicide, ii) investigate spectral characteristics of white mold onset in snap beans, and iii) to link the location of white mold with metrics like ratio indices (RI), normalized difference indices (NDI), leaf area index (LAI), row and plant spacing, and digital elevation models to create a spatially-explicit probabilistic risk model for the appearance of white mold in a snap bean field. This paper will focus on applying proven multivariate and machine learning approaches to the first objective of this research.

Methods

Data Collection

A DJI Matrice-600 UAS, with a high spatial resolution color (RGB) camera, a Headwall Photonics Nano imaging spectrometer (272 bands; 400 to 1000 nm), and a Velodyne VLP-16 light detection and ranging (LiDAR) system, was used to collect snap bean spectra during flowering onset at the New York State Agricultural Experiment Station, Cornell University, Geneva, NY, USA. Snap bean plants were planted at three different periods in the growing season, 'early' (June 3, 2017), 'mid (June 14, 2017), and 'late' (June 28, 2017), which allowed for different crop development stages to be captured in single UAS flights. Twenty-two flights were conducted throughout the season on ten different days with GSDs that ranged between 1.25 to 3 cm. By July 19, 2017 the portion of the field that was planted first (early plant) was 99% flowering, while the second and third (mid and late planting dates) were not flowering. Data from the first flight (July 19, 2017) with a GSD of 3 cm were used to train the snap bean flowering models.

Data Analysis

White and black calibration panels were placed in the field during the various flight missions. Spectra of the calibration panels were collected using a SVC hand-held spectrometer (334-2508 nm; 2 nm resampled bandwidths). These panels were used to calibrate hyperspectral data from the UAS into reflectance using the empirical line method of calibration (ELM) (Smith 2010). Calibrating into reflectance is critical, since it largely removes the illumination dependence from the image. The ELM takes the UAS radiance spectra of the panels and calculates gain factors that force the radiance spectra to approximate the reference reflectance spectra, collected from the panels on the ground using the hand-held spectroradiometer.

$$\beta = \frac{SVC_{white\ Ref.} - SVC_{Black\ Ref.}}{UAS_{White\ Rad.} - UAS_{Black\ Rad.}} \tag{1}$$

where, β is a vector that contains the gain factors for each wavelength to turn any radiance spectrum in the image into reflectance. Next, the gain factors were applied to the spectra of every pixel to convert all the radiance spectra into reflectance.

$$UAS_{Ref.} = \beta * UAS_{Rad.} \tag{2}$$

Plots within the field were monitored at two to three-day intervals for crop development and flowering. Regions of interest (ROIs) were created for plots that contained flowering and non-flowering plants. Spectral samples were collected from the UAS reflectance imagery on dates when 99% of the plants within the plot were flowering.

The labeled spectral samples (100 flowering, 134 non-flowering) were then divided evenly into training and testing datasets (50 testing/ 50 training and 67 testing/ 67 training for flowering/ non-flowering, respectively) and processed by performing a per-feature (e.g., wavelength, ratio, or normalized difference index) linear discriminant analysis (LDA). LDA is a statistical classification and dimensionality reduction method that finds a linear boundary between classes by using Bayes rule to fit class conditional densities. Within LDA, each class is assumed to share the same covariance matrix and is fitted with a Gaussian density (Randles 2012). Single spectral features/metrics were tested on an individual basis by the LDA model. The features that were used in the LDA include RI, NDI and the reflectance values at the wavelengths. RI simply imply the division of two reflectance values.

$$RI = \frac{R_{\lambda_n}}{R\ \lambda_m} \tag{3}$$

Normalized difference indices refer to a normalization of the spectral differences by their sum, similar to the popular normalized difference vegetation index (NDVI) (Delalieux *et al*. 2009).

$$NDI = \frac{R_{\lambda_n} - R_{\lambda_m}}{R_{\lambda_n} + R_{\lambda_m}} \tag{4}$$

Table 1 shows the RI and NDI features used for the LDA. The wavelengths used for RI and NDI calculations are similar to those listed by Delalieux *et al.* (2009), specifically because these features correlated with chlorophyll content and water stress.

Finally, the discriminating performance of each spectral feature was evaluated via the c-index, which corresponds to the area under the receiver operator characteristic (ROC) curve. A c-index of 0.5 implies that the discrimination is random, while a c-index of 1.0 infers perfect discrimination between the training and testing datasets. A c-index of 0.8 or greater is an acceptable discriminating model (Delalieux *et al*., 2007).

Table 1. Wavelengths used to create ratio indices and normalized difference indices used as features in the per-feature linear discriminant analysis (Delalieux *et al.,* 2009).

Ratio Indices	
Numerator Wavelength (nm; R_{λ_n})	Denominator Wavelength (nm; R_{λ_m})
430	680
440	690
550	800
605	760
672	550
675	700
695	420
695	670
695	760
710	760
740	720
750	550
750	705
750	710
800	550
800	635
800	680
Normalized Difference Vegetative Indices	
First Wavelength (nm; R_{λ_n})	Second Wavelength (nm; R_{λ_m})
415	435
680	430
750	660
750	705
750	445
800	635
800	680

Next, features with the largest c-index were combined to train and test a LDA model using least squares, a logistic regression, and a SVM with a radial basis function kernel (Melgani 2004). The model with the highest mean accuracy was applied to regions of the hyperspectral imagery that were flowering and non-flowering, in order to spatially visualize the model probabilities.

Results and Discussion

The per-feature LDA identified three ratio index features and two NDI features with c-indices at least 0.8 (Fig. 1). The wavelengths with the highest c-index values (725.8 nm, 759.2 nm, and 721.3 nm), are within the red portion of the spectrum.

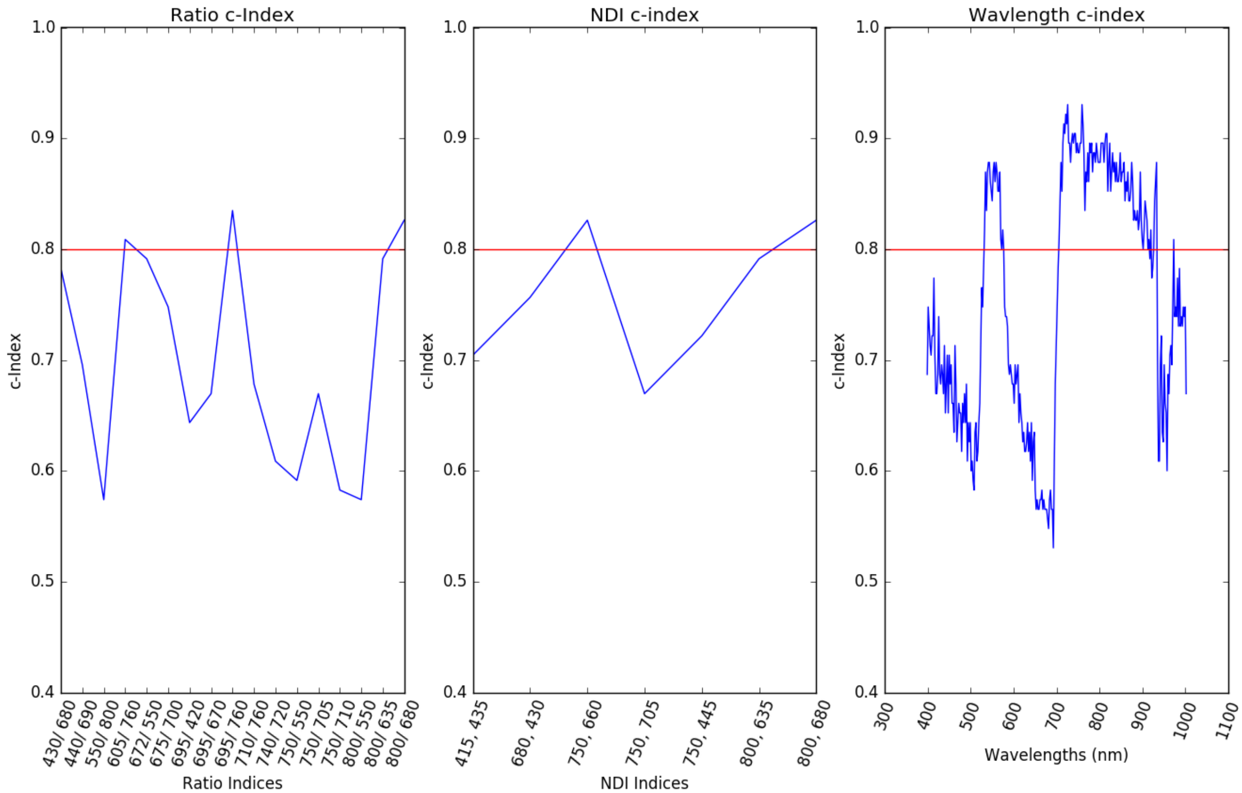

Figure 1. c-index values, which represent the area under the receiver operating characteristic curve, for each single spectral fearture that was used in the linear discrimiant analysis. Features with high c-index values have the ability to separate the flowering/ non-flowering classes and will be used toward reducing the data dimensionality of selected spectral features to train and test the linear discriminant, logistic regression, and support vector machine models in this study.

The top three, 10, and 20 spectral features, i.e., those with the highest c-index values, were used to train and test three different models. The LDA model approach had the highest mean accuracy of 95%, with 10 spectral features. Even with three features, the LDA model was still able to achieve a mean accuracy of 92% (Table 2).

The 10-feature LDA model was then applied to two ROIs, one flowering and one non-flowering, from the snap bean fields of interest. The flowering probability of each pixel is displayed to create a "heat map", which eventually may be used for management interventions. Figure 2 below confirms that the probability map for the image with flowering plants shows high probability-of-flowering where one would expect, i.e., at individual plant locations, and < 50% probabilities for bare soil areas. Figure 2 also shows that the model for flowering, when applied to an image with non-flowering plants, exhibits probabilities < 50%, with only erroneous outliers. This serves as visual confirmation of the efficacy of the approach, where the oversampled spectral data was distilled to key indicators, quantitatively evaluated flowering accuracies, and mapped these to potential management-ready heat maps for use in crop operations.

Table 2. Following the per-feature linear discriminant analysis, the top three, 10, and 20 3 c-index values were used to identify key spectral features with which to train and test linear discriminant, logistic regression, and support vector machine models (mean accuracies for the various models are shown here).

Model	Mean Accuracy (%)		
	Three features	10 features	20 features
Linear Discriminant	92	95	93
Logistic Regression	67	73	86
Support Vector Machine	84	85	80

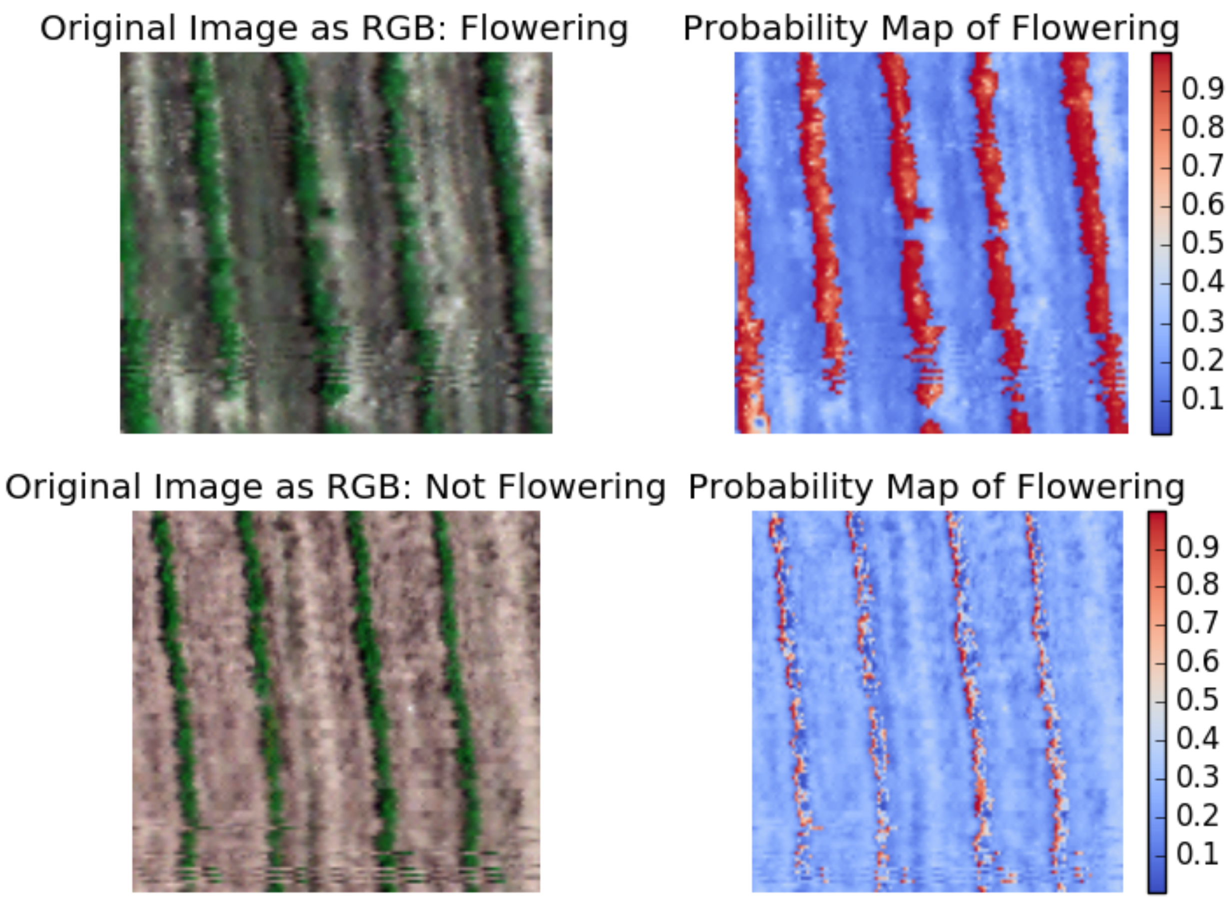

Figure 2. Two-dimensional spatial probability maps from the trained support vector machine model; red pixels indicate a higher probability that the pixel contains a flowering snap bean plant.

Conclusions

This research represents a first step toward developing extensive white mold risk models based on UAS sensing of both spectral and structural features. The first step involved an evaluation of spectral metrics, from ratio and NDI metrics to individual reflectance wavelengths, to detect flowering. This detection is critical to ensure appropriate timing of fungicides to protect flowers from ascosporic infection. The single feature LDA approach was able to identify key spectral metrics for flowering detection, e.g., with key wavelengths located in the red portion of the reflectance spectrum. The LDA model with least squares returned a classification accuracy of

92%, 95%, and 93% for three, 10, and 20 spectral features, respectively. This high classification accuracy for even three features has exciting implications for the commercial deployment for UAS in precision agriculture and for disease management. Instead of utilizing large UAS with heavy payloads and expensive imaging spectrometers, it may be possible to use silicon range detectors with bandpass filters to obtain the necessary data for flowering detection. This would enable more operational (smaller and cheaper) UAS solutions for crop management. Future work on this project will include investigating spectral features that will aid in the prediction of flowering onset, i.e., a temporal analysis approach.

Acknowledgements

We would like to thank Mr. Tim Bauch for his assistance with UAS data collection and processing, as well as Ms. Nina Raqueno for her help with data pre-processing. This research is supported USDA-NIFA Critical Agriculture in Research and Extension Award Number 2017-68008-26207.

References

Delalieux S., Somers B., Verstraeten W. W., Aardt J. van, Keulemans W., and Coppin P. (2009). Hyperspectral indices to diagnose leaf biotic stress of apple plants, considering leaf phenology. *International Journal of Remote Sensing*.

Delalieux S., Aardt J. van, Keulemans W., Schrevens E., and Coppin P. (2007). Detection of biotic stress (venturia inaequalis) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications. *European Journal of Agronomy*.

Lehner, M., Kikkert, J. R., Gugino, B., and Pethybridge, S. J. (2017). Fungicide sensitivity and efficacy of boscalid, fluazinam and thiophanate-methyl for white mold control in snap bean. Plant Disease 101:1253-1258.

Melgani F., Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines. (2004) *IEEE Transactions on Geoscience and Remote Sensing*.

Randles R., Broffitt J., Ramberg J., Hogg R. (2012). Generalized Linear and Quadratic Discriminant Functions Using Robust Estimates, *Journal of the American Statistical Association*, 73:363, 564-568,DOI: 10.1080/01621459.1978.10480055

Rumpf T., Mahlein A., Steiner U., Oerke E., Dehne H., and Plmera L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*.

Smith G., Milton E. (2010) The use of the empirical line method to calibrate remotely sensed data to reflectance, *International Journal of Remote Sensing*, 20:13, 2653-2662, DOI: 10.1080/014311699211994

USDA National Agricultural Statistics Service. (2012). Census of Agriculture. *U.S. Dep. Agric. Nat. Agric. Stat. Serv. Online.*

USDA National Agricultural Statistics Service. (2014). The 2014-2015 New York Annual Bulletin, *U.S. Dep. Agric. Nat. Agric. Stat. Serv. Online.*

USDA National Agricultural Statistics Service. (2015). National Statistics for Beans, *U.S. Dep. Agric. Nat. Agric. Stat. Serv. Online*.