Scientific Research Publishing

Employing Canopy Hyperspectral Narrowband Data and Random Forest Algorithm to Differentiate Palmer Amaranth from Colored Cotton

Reginald S. Fletcher[1], Rickie B. Turley[2]

[1]Crop Production Systems Research Unit, Agricultural Research Service, United States Department of Agriculture, Stoneville, USA
[2]Crop Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, Stoneville, USA
Email: reginald.fletcher@ars.usda.gov

Abstract

Palmer amaranth (*Amaranthus palmeri* S. Wats.) invasion negatively impacts cotton (*Gossypium hirsutum* L.) production systems throughout the United States. The objective of this study was to evaluate canopy hyperspectral narrowband data as input into the random forest machine learning algorithm to distinguish Palmer amaranth from cotton. The study focused on differentiating the Palmer amaranth from cotton near-isogenic lines with bronze, green, and yellow leaves. A spectroradiometer was used to acquire hyperspectral reflectance measurements of Palmer amaranth and cotton canopies for two separate dates, December 12, 2016, and May 14, 2017. Data were collected from plants that were grown in a greenhouse. The spectral data were aggregated to twenty-four hyperspectral narrowbands proposed for study of vegetation and agriculture crops. Those bands were tested by the conditional inference version of random forest (cforest) to differentiate the Palmer amaranth from cotton. Classifications were binary: Palmer amaranth and cotton bronze, Palmer amaranth and cotton green, and Palmer amaranth and cotton yellow. Classification accuracies were verified with overall, user's, and producer's accuracy. For the two dates combined, overall accuracy ranged from 77.8% to 88.9%. The highest overall accuracies were observed for the Palmer amaranth versus the cotton yellow classification (88.9%, December 12, 2016; 83.3%, May 14, 2017). Producer's and user's accuracies range was 66.7% to 94.4%. Errors were predominately attributed to cotton being misclassified as Palmer amaranth. The overall results indicated that cforest has moderate to strong potential for differentiating Palmer amaranth from cotton when it used hyperspectral narrowbands known to be useful for vegetation and agricultural surveys

as input variables. This research further supports using hyperspectral narrowband data and cforest as decision support tools in cotton production systems.

Keywords

1. Introduction

Palmer amaranth, an aggressive and invasive weed, negatively impacts cotton growth and productively throughout the United States. It grows at a rapid rate (*i.e.*, approximately 25 - 50 mm per day), produces several thousand seeds per plant, competes with cotton plants for sunlight and soil nutrients, and reduces cotton yield. Palmer amaranth populations are controlled by chemical and mechanical means. Its presence in fields is observed via site visits, which are time consuming when several hundred hectares need to be surveyed. To better implement control strategies for Palmer amaranth invasions in cotton production systems, agriculturalists need tools that can help them differentiate it from cotton. Researchers, consultants, and producers have shown interests in employing remote sensing technologies as weed detection and survey tools in crop production systems.

Remote sensing applications for weed survey have focused on comparing leaf and canopy light reflectance properties of weeds to crops with the intention to identify spectral bands showing statistically significant differences between them. Spectral profiles and statistical analyses were used to distinguish soybean (*Glycine max* L.) from pitted morning glory (*Ipomea lacunose* L.) [1], sunflower (*Helianthus annuus* L.) from corn caraway (*Ridfolia segetum* Moris.) [2], and Palmer amaranth and redroot pigweed from cotton [3]. Over the years, there has been a shift towards using multispectral and hyperspectral data as input into machine learners to differentiate crops from weeds. Machine learning involves training computer algorithms to discover patterns in datasets.

Hyperspectral and multispectral datasets were used with stepwise discriminate analysis and multilayer perceptron and radial base function neural networks to distinguish cruciferous weeds (*Sinapis* spp. and *Diplotaxis* spp.), wheat (*Triticum durum* L.), and broad bean (*Viciafaba* L.) [4]. The multilayer perceptron model was the most accurate classifier. Artificial neural networks coupled with spectral data distinguished grass weed species [wild oat (*Avena sterilis* L.), canary grass (*Phalaris* spp.) and ryegrass (*Lolium rigidum* L.)] from winter wheat [5] and performed better than maximum likelihood classification to distinguish redroot pigweed (*Amaranthus retroflexus* L.) and wild oats (*Avena fatua* L.) [6]. Linear discriminate analysis and leaf hyperspectral and multispectral reflectance data separated wheat and canola (*Brassica napus* L.) from common lambsquarter (*Chenopodium album* L.), green foxtail (*Setaria viridis* L.), redroot pigweed,

wild mustard (*Sinapis arvensis* L.), and wild oak (*Avena fatua* L.) [7] and soybeans from pitted morning glory (*Ipomoea lacunosa* L.) [1]. Random forest and multispectral data have shown promise for soybean Palmer amaranth separation [8] [9]. The previously mentioned studies were based on laboratory and field measurements. Researchers have had success in employing hyperspectral data obtained with aerial systems as input into support vector machines, artificial neural networks, and discriminant analyses to detect several grass [yellow nutsedge (*Cyperus esculenthus* L.), barnyardgrass (*Echinochloa crus galli* L. Beauv.), and crab grass (*Digitaria ischaemum* (Schreb.) Schreb. exMuhl.)] and broadleaved weeds [Canada thistle (*Cirsium arvensis* L. Scop.), sow thistle (*Sonchus oleraceus* L.), redroot pigweed, and common lambsquarter] in corn (*Zea mays* L.) plots [10] [11]. Overall, [12] suggested that 15 to 28 spectral bands are needed for vegetation study, characterization, and mapping.

As indicated earlier, machine learners coupled with spectral data have shown promise for weed-crop discrimination. The user has many learners to choose from. Random forest, a nonparametric machine learning method, uses a group of decision trees to estimate a value or assign an object to a class [13]. It is competitive with other machine learners such as support vector machines and neural networks and was ranked as one of the best classifiers [14] [15]. Minimum data preparation is needed by the algorithm. Random forests work well with large datasets and are not affected by outliers. The algorithm does not require a separate testing set because it uses bootstrap sampling for each tree, meaning each tree is built with 63% of the data and thus leaving 37% of the data for testing (*i.e.*, the "out-of-bag" data).

Crop plants with specific traits are being developed by plant breeders to meet the desired characteristics needed for crop production systems. Presently at experiments stations, researchers are evaluating cotton near-isogenic lines that have bronze, green, and yellow colored leaves with the intention of using them in cotton production systems. Reference [3] statistically compared the leaf spectral profile of the cotton lines to redroot pigweed and Palmer amaranth. Currently, no research is available that has explored application of canopy narrowband hyperspectral data and random forest machine learning algorithm for differentiating Palmer amaranth from cotton with different colored leaves. The objective of this study was to evaluate canopy hyperspectral narrowband data as input into random forest machine learning algorithm to distinguish Palmer amaranth from cotton. The study focused on using optimal hyperspectral narrowbands proposed for study of vegetation and agricultural crops [12] as input into the random forest algorithm. Additionally, the research concentrated on cotton near-isogenic lines that have bronze, green, or yellow leaf colors.

2. Materials and Methods

2.1. Experimental Setup

The study was conducted in a greenhouse located at the United States Depart-

ment of Agriculture (USDA), Agricultural Research Service (ARS), Jamie Whitten Delta States Research Center, Stoneville, MS (33.425261 latitude, −90.912740 longitude). Palmer amaranth and cotton near-isogenic line seeds (*i.e.*, cotton bronze, cotton green, and cotton yellow seed) were obtained from established seed banks maintained at USDA-ARS, Stoneville, MS. A detailed description of the cotton near-isogenic lines is provided in [3].

The experimental design was a randomized complete block design with 18 replications and 4 treatments per replication (*i.e.*, 1 cotton bronze plant, 1 cotton green plant, 1 cotton yellow plant, and 1 Palmer amaranth plant). Data were collected from two separate experiments. Planting dates were November 7, 2016 and April 18, 2017 for experiments one and two, respectively. For each experiment, several seeds of the respective plants were planted into 11-cm pots containing potting mix (Pro-Mix BX general professional growth medium, Premier Tech Horticulture, Quakertown, PA). Approximately 10 days after emergence, plants were thinned to 1 plant per pot. The plants were subjected to a 14-hr day length; the greenhouse temperature was maintained between 21.1˚C and 26.7˚C. Sodium vapor lamps were used as a supplemental light source at the beginning and ending of the day. The plants were fertilized weekly with a liquid fertilizer (approximately 4.9 ml of fertilizer to 3.8 L of water, Dyna-Gro All-Pro, Richmond, CA); water was added as needed.

2.2. Data Collection

Canopy reflectance measurements of Palmer amaranth and cotton plants were obtained with a hyperspectral spectroradiometer (Fieldspec 3 Full Range, ASD Inc. Boulder, CO) sensitive to a spectral range of 350 to 2500 nanometers (nm). Reflectance measurements were acquired in the vegetative growth stage (*i.e.*, Palmer amaranth 10 leaf stage and cotton 4 leaf stage). Herbicide management programs are more effective when weeds are treated in the vegetative phase; and most users want to kill or treat weeds prior to seeding.

Canopy reflectance measurements were acquired on December 12, 2016 and May 14, 2017. They were obtained outside of the greenhouse under sunlit conditions. Measurements were obtained ± 2 hours of solar noon. Prior to measurements, a black felt cloth was placed across the top of the pot to cover the potting mix, thus providing a uniform background. The black felt cloth's spectral reflectance was less than 2 percent in all regions of the spectrum measured by the spectroradiometer.

The spectroradiometer was calibrated at approximately 15-minute intervals with a white calibration panel. For canopy measurements, the spectroradiometer's fiber optic cable was held 15 cm from the top of the plant canopy, resulting in a ground field of view of 35 cm^2. For each plant canopy, the spectral reading was an average of 15 scans collected by the instrument.

2.3. Post Processing

The 1-nm hyperspectral reflectance data were post processed as follows [16]: (1)

splice correction of each spectra, (2) removal of water absorption bands and bands not commonly used by remote sensing systems (350 - 400 nm, 1330 - 1480 nm, 1780 - 1990 nm, 2400 - 2500 nm), and (3) smoothing (*i.e.*, removal of noise from the data) of the spectrum with Savitzky-Golay filter [17] (The filter removes noise from the spectral data while maintaining the shape of the spectrum). Splice correction eliminated gaps in the spectral data caused by the different detector arrays in the spectroradiometer. For the FieldSpec3 spectroradiometer, correction is needed at 1000 nm and 1800 nm. It was completed with the View SpecPro software (Version 6.2; ASD, Inc., Boulder, CO). The Hyperspectral Data Analysis in R (HSDAR, [18] [19]) package was employed to remove water absorption bands, to smooth spectra with the Savitzky-Golay filter (smoothing parameter; n = 25), and to create the optimal hyperspectral narrowband bands proposed for study of vegetation and agricultural crops [12]. The center wavelengths and bandwidths of the hyperspectral narrowbands are summarized in Table 1 [12].

Table 1. Hyperspectral narrowbands evaluated as input into cforest for Palmer amaranth and cotton discrimination. These bands have been identified as optimal for vegetation and agricultural studies [12].

Spectral Region	Central Wavelength (nm)	Bandwidth (nm)	Vegetation and Agricultural Applications
Blue	405	5	Nitrogen; senescing
	490	5	Carotenoid; light use efficiency; stress in vegetation
Green	515	5	Carotenoid; anthocyanins; nitrogen; vigor
	531	1	Light use efficiency; xanophyll cycle; stress in vegetation; pest and disease
	550	5	Chlorophyll
	570	5	Anthocyanins; chlorophyll; nitrogen
Red	682	5	Biophysical quantities; yield
Red-edge	705	5	Stress; chlorophyll
	720	5	Stress; chlorophyll
Near-infrared	855	20	Biophysical quantities and yield
	910	5	Moisture; biomass; and protein
	970	10	Water; moisture; and biomass
	1075	5	Biophysical and biochemical quantities
	1180	5	Water absorption band
	1245	5	Water sensitivity
Shortwave-infrared	1518	5	Moisture; biomass
	1650	5	Heavy metal stress; moisture sensitivity
	1725	5	Lignin; biomass; starch; moisture
	2025	5	Plant litter; lignin; cellulose; plant litter-soil differentiation
	2133	5	Plant litter; lignin; cellulose
	2205	5	Litter; lignin; cellulose; sugar; starch; protein; heavy metal stress
	2260	5	Moisture and biomass
	2295	5	Stress
	2359	5	Cellulose; protein; nitrogen

2.4. Classifications

The spectral data were highly correlated according to Pearson correlation analyses. Therefore, classifications were completed with the conditional inference version (cforest) of random forest. It produces more stable variable of importance readings than random forest, especially when variables are highly correlated [20]. Also, cforest is different than random forest in that it employs conditional inference trees as base learners and the conditional permutation scheme described by [20] [21] to rank variables. To complete the classifications, the decision trees (*i.e.*, forest) were developed using subsampling without replacement [22]. That procedure reduces the bias in variable importance rankings and is appropriate to use when potential predictor variables vary in their scale of measurement or their number of categories [22].

Cforest required adjustments of two parameters to develop the classification models, *ntree* and *mtry*. *Ntree* characterized the number of decision trees to use in each forest. *Mtry* represented the number of variables randomly chosen to derive tree split points.

Cforest models were developed and tested based on the following procedures. A set seed value of 100 was chosen; the set seed value represented the start point of a set of pseudo-random numbers. Using the same seed number, the analyst will be able to produce identical results. Then the model was evaluated using *ntree* values of 50, 75, and 100 to 1000 in increments of 100 (*i.e.*, 100, 200, 300, etc.). The *ntree* value in which the "out-of-bag" (samples not used in a tree) classification error stabilized was selected as the appropriate number of trees. Then, *mtry* values of 3, 5, 10, 15, 20, and 24 were evaluated with the selected optimal *ntree* value. The *mtry* value in which the "out-of-bag" error stabilized was chosen as the optimal model. To test the consistency in the variable importance scores, the classification was repeated with the optimal *ntree* and *mtry* values and a different starting seed [22]; 200 was selected as the starting seed value when checking consistency of the model. For each date, the classifications were binary: (1) Palmer amaranth versus cotton bronze, (2) Palmer amaranth versus cotton green, and (3) Palmer amaranth versus cotton yellow. Cforest classifications were conducted with the Party package of R 3.4.0 [18] [21] [22] [23].

2.5. Accuracy Assessment of Cforest Models

Accuracy of cforest models was determined with user's, producer's, and overall accuracies [24]. User's accuracy was tabulated by dividing the sum of the correctly classified plants for a specific class by the total number of plants determined by the classifier for that class [24] [25]. It described the reliability of the classification. Producer's accuracy was calculated by dividing the number of plants correctly classified into a category by the total number of reference plants in that category [24] [25]. It represented the accuracy in classifying the reference samples. Overall accuracy was determined by dividing the total number of correctly classified plants for all classes by the total number of plants. User's, pro-

ducer's, and overall accuracies were manually tabulated using error matrix results obtained with cforest.

3. Results

Cforest model parameters are summarized in Table 2 for each classification. The default *ntree* (500) and *mtry* (5) values were adequate to use for five out of six classifications. The only exception was the May 14, 2017, Palmer amaranth versus cotton bronze classification. For that classification, the best accuracies were obtained when all variables were considered for splitting a tree node.

Accuracy assessment results of the classifications are shown in Table 3. Overall accuracies were greater than 77% with the lowest overall accuracy achieved for Palmer amaranth versus cotton green and Palmer amaranth versus cotton bronze classifications for December 12, 2016 and May 14, 2017, respectively. The greatest overall accuracy was obtained for the Palmer amaranth versus cotton yellow classifications for both dates. User's and producer's accuracies ranged from 66.7% to 94.4% (Table 3). Overall, Palmer amaranth user's accuracy was less than its producer's accuracy; whereas, the opposite was observed for the cotton classes (Table 3). The highest user's and producer's accuracies for the Palmer amaranth class varied for the classifications (Table 3). For both dates, the user's accuracy of cotton yellow was greater than the user's accuracies of cotton bronze and cotton green. Producer's accuracy varied from one classification to the next for the cotton classes (Table 3). The user's accuracy of the Palmer amaranth class was less than the user's accuracy of the cotton classes; in contrast, the producer's accuracy of the Palmer amaranth class was greater than the producer's accuracy of the cotton classes (Table 3).

The variable importance rankings of the classifications for both dates are shown in Figure 1 and Figure 2. The results of the December 12, 2016 dataset indicated that 14 to 15 bands were important for differentiating Palmer amaranth from cotton. Those bands encompassed the blue, green, red, red-edge, near-infrared, and shortwave-infrared regions of the spectrum (Figure 1, Table 1). The top ranked bands were 705 nm (red-edge), 720 nm (red-edge), and 570

Table 2. Cforest model parameters (*i.e.*, *ntree* and *mtry*) used in Palmer amaranth versus cotton classifications.

Date	Comparison	*ntree*	*mtry*
December 12, 2016	PAL-CB[a]	500	5
	PAL-CG	500	5
	PAL-CY	500	5
May 14, 2017	PAL-CB	500	24
	PAL-CG	500	5
	PAL-CY	500	5

[a]CB: cotton bronze, CG: cotton green, CY: cotton yellow, PAL: Palmer amaranth, *ntree* = number of trees grown, *mtry* = number of predictors to use when splitting a node.

Table 3. Accuracy assessment results of Palmer amaranth versus cotton classifications based on cforest algorithm using hyperspectral narrowbands as input.

Date	Comparison	User's Accuracy	Producer's Accuracy	Overall Accuracy
December 12, 2016	PAL[a]	73.9%	94.4%	
	CB	92.3%	66.7%	
				80.6%
	PAL	72.7%	88.9%	
	CG	85.7%	66.7%	
				77.8%
	PAL	85.0%	94.4%	
	CY	93.8%	83.3%	
				88.9%
May 14, 2017	PAL	72.7%	88.9%	
	CB	85.7%	66.7%	
				77.8%
	PAL	78.9%	83.3%	
	CG	82.4%	77.8%	
				80.6%
	PAL	77.3%	94.4%	
	CY	92.9%	72.2%	
				83.3%

[a]CB: cotton bronze, CG: cotton green, CY: cotton yellow, and PAL: Palmer amaranth

nm (green) corresponding to Palmer amaranth versus cotton bronze, Palmer amaranth versus cotton green, and Palmer amaranth versus cotton yellow classifications, respectively.

For the May 14, 2017 dataset, 2, 11, and 15 bands were identified as important to Palmer amaranth versus cotton bronze, Palmer amaranth versus cotton green, and Palmer amaranth versus cotton yellow classification models, respectively (Figure 2). Those bands encompassed the blue, green, red, red-edge, near-infrared, and shortwave-infrared regions of the spectrum (Figure 2, Table 1). The top ranked bands were 705 nm (red-edge), 2025 nm (shortwave-infrared), and 705 nm (red-edge) for the Palmer amaranth versus cotton bronze, Palmer amaranth versus cotton green, and Palmer amaranth versus cotton yellow classifications, respectively.

4. Discussion

Hyperspectral narrowbands proposed for vegetation and agricultural surveys were evaluated as input into cforest classification algorithm to differentiate Palmer amaranth from cotton. The spectral data and machine learning algorithm were tested for three different scenarios and for two separate dates (Table 3).
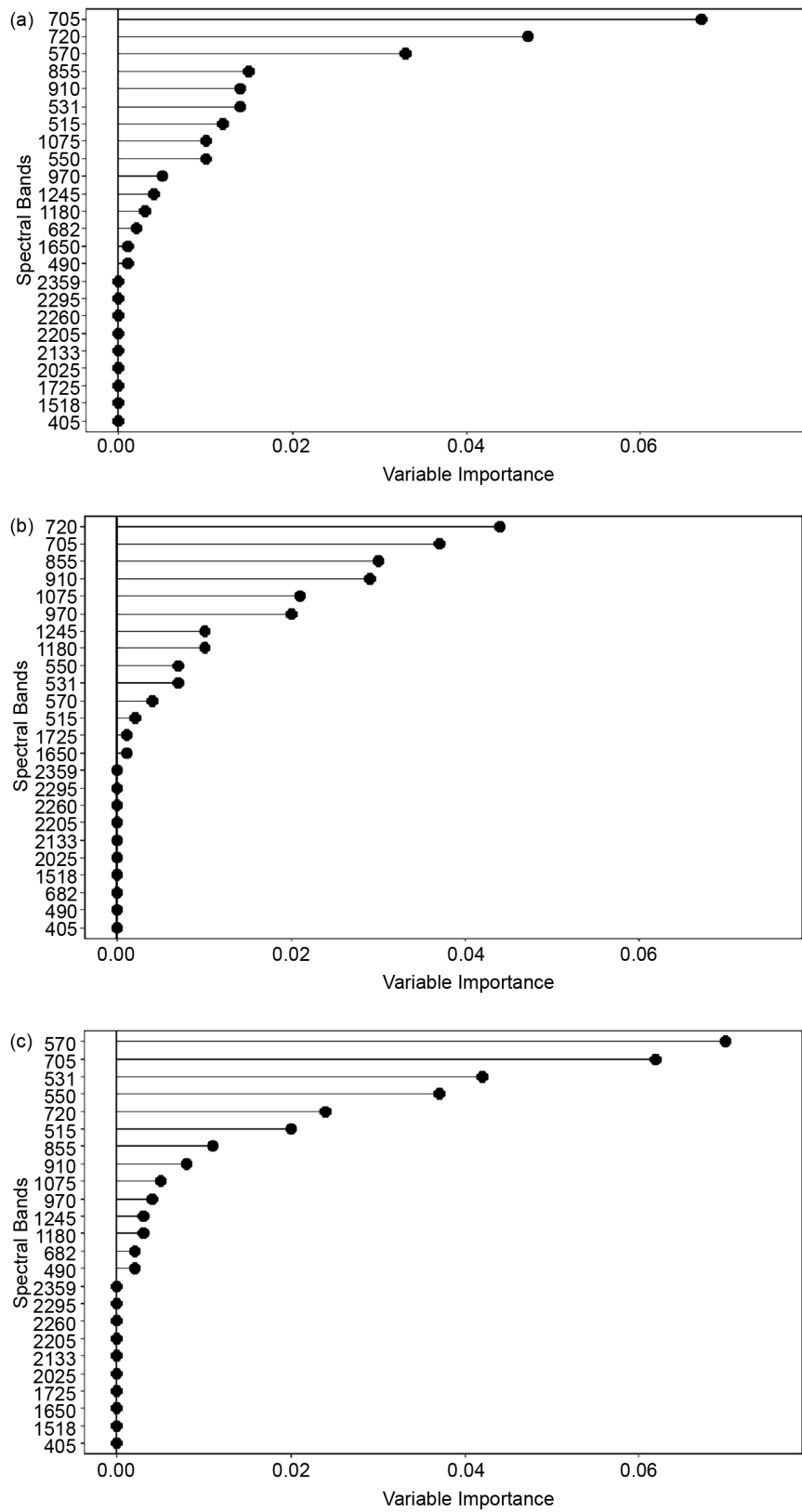
Figure 1. Variable importance rankings derived from cforest model for each classification, December 12, 2016: (a) Palmer amaranth versus cotton bronze; (b) Palmer amaranth versus cotton green; and (c) Palmer amaranth versus cotton yellow.
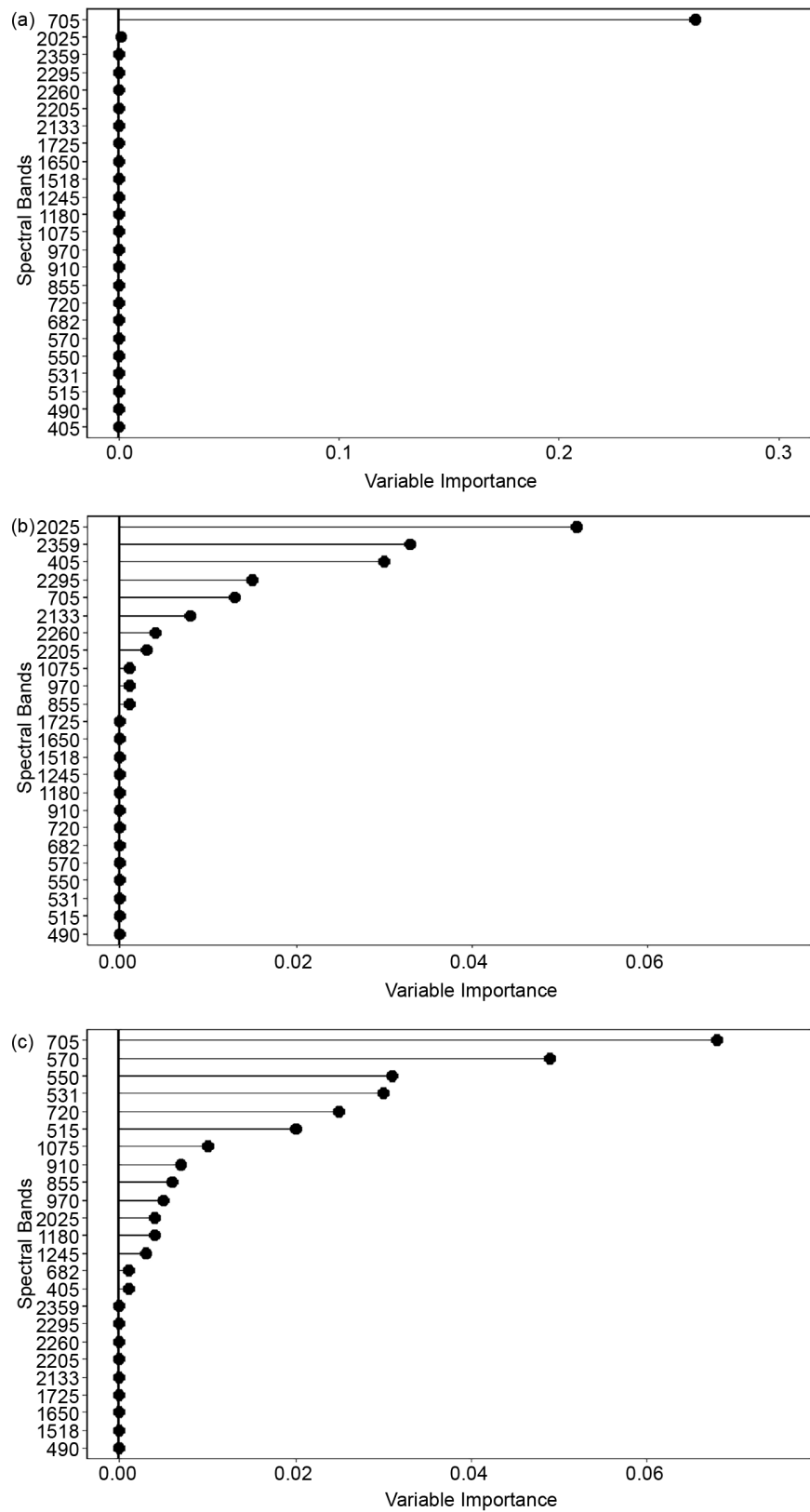
Figure 2. Variable importance rankings derived from cforest model for each classification, May 14, 2017: (a) Palmer amaranth versus cotton bronze; (b) Palmer amaranth versus cotton green; and (c) Palmer amaranth versus cotton yellow.

Overall, the best classification accuracies were achieved for the Palmer amaranth versus cotton yellow classifications with accuracies ranging from 77.3% to 94.4% (Table 3). For December 12, 2016, the second best classification accuracies were observed for the Palmer amaranth versus cotton bronze classification (Table 3). The second best overall accuracy for the May 14, 2017 classifications occurred for Palmer amaranth versus cotton green (Table 3). The user's and producer's accuracies rankings were not consistent for the Palmer amaranth and cotton bronze versus Palmer amaranth and cotton green classifications (Table 3). User's and producer's accuracies indicated that errors in the classifications were caused by cotton being misclassified as Palmer amaranth. From the weed management perspective, that misclassification error is not as bad as Palmer amaranth being misclassified as cotton.

The results suggested that there still exist areas for continued development of the models. Several research studies have found that narrowband vegetation indices performed well in characterizing agricultural crops [12] [26]. An alternative is to explore the narrowband vegetation indices as input variables for cforest algorithms. The analyst could also test other machine learning algorithms such as support vector machines or artificial neural networks. Overall, the results for all three classification scenarios were in agreement with other studies promoting hyperspectral narrowbands for vegetation and agricultural surveys [12] [26] [27] and showing potential of random forest for crop-weed discrimination [8] [9].

Variable importance rankings suggested that 15 spectral bands or less were needed for Palmer amaranth cotton discrimination (Figure 1, Figure 2). Other researchers have also shown that 15 to 30 spectral bands are needed for vegetation and crop surveys [12] [26]. Additionally, one or more of the red-edge bands (*i.e.*, 705 nm, 720 nm) were ranked in the top five variable importance for all classifications. The red-edge region (680 - 740 nm) of the optical spectrum is the transition zone between red and near-infrared reflectance of vegetation, and represents the change between chlorophyll absorption and light scattering caused by leaf internal structure [28]. Changes in both chlorophyll content and leaf structure are often reflected in the red-edge region of the spectrum. Shifts in the red-edge position are caused by chlorophyll content [29] [30], leaf area index [31], and plant biomass [32].

For this study, differences in leaf chlorophyll content and canopy architecture impacted red-edge bands variable importance ranking. The cotton canopies consisted of bronze, green, or yellow leaves, hence difference in chlorophyll content when compared with the green leaves of Palmer amaranth. Also, the cotton plants consisted of broad leaves and had a closed canopy compared with the narrower leaves and open canopy of the Palmer amaranth plants. The canopy architecture of Palmer amaranth would have more in-canopy shadowing compared to cotton, hence affecting its reflectance. Furthermore, the rankings indicated that visible, near-infrared, and shortwave-infrared bands contributed to the models derived for the classifications (Figure 1, Figure 2; Table 1), fur-

ther substantiating that differences in leaf pigment, canopy architecture, and leaf water content, respectively affected the classification results.

5. Conclusion

This study demonstrated that canopy hyperspectral narrowband data could be used in tandem with random forest machine learning algorithm to differentiate cotton from Palmer amaranth, an invasive weed of cotton production systems. This study not only focused on cotton with green leaves, but also evaluated the hyperspectral narrowbands and classification algorithm on cotton with bronze and yellow leaves. Out of the 24 bands evaluated in this study, 15 or less were important to Palmer amaranth-cotton discrimination. The next step will be to determine whether random forest (*i.e.*, cforest and other versions of the algorithm) could be used with airborne hyperspectral imaging data to differentiate Palmer amaranth from cotton with different colored leaves or not. When using airborne imaging systems, the user must consider that in-canopy shadowing, soil background, bi-directional reflectance, spatial resolution of the imagery, and radiometric resolution of the imaging system will influence the spectral response of the feature of interest. Overall, this research further supports using hyperspectral narrowband data and cforest as decision support tools for weed discrimination in cotton production systems.

Acknowledgements

References

[1] Koger, C.H., Shaw, D.R., Reddy, K.N. and Bruce, L.M. (2004) Detection of Pitted Morningglory (*Ipomoea lacunosa*) with Hyperspectral Remote Sensing. II. Effects of Vegetation Ground Cover and Reflectance Properties. *Weed Science*, **52**, 230-235. https://doi.org/10.1614/WS-03-083R1

[2] Peña-Barragãn, J.M., López-Granados, F., Jurado-Expósito, M. and García-Torres, L. (2006) Spectral Discrimination of *Ridolfia segetum* and Sunflower as Affected by Phenological Stage. *Weed Research,* **46**, 10-21. https://doi.org/10.1111/j.1365-3180.2006.00488.x

[3] Fletcher, R.S., Reddy, K.N. and Turley, R.B. (2016) Spectral Discrimination of Two Pigweeds from Cotton with Different Leaf Colors. *American Journal of Plant Sciences*, **7**, 2138-2150. https://doi.org/10.4236/ajps.2016.715190

[4] de Castro, A.I., Jurado-Exposito, M., Gomez-Casero, M.T. and Lopez-Granados, F. (2012) Applying Neural Networks to Hyperspectral and Multispectral Field Data for Discrimination of Cruciferous Weeds in Winter Crops. *The Scientific World Journal*, **2012**, Article ID: 630390. https://doi.org/10.1100/2012/630390

[5] López-Granados, F., Peña-Barragán, J.M., Jurado-Expósito M., Francisco-Fernández,

M., Cao, R., Alonso-Betanzos, A. and Fontenla-Romero, O. (2008) Multispectral Classification of Grass Weeds and Wheat (*Triticum durum*) Crop Using Linear and Nonparametric Functional Discriminant Analysis, and Neural Networks. *Weed Research*, **48**, 28-37. https://doi.org/10.1111/j.1365-3180.2008.00598.x

[6] Eddy, P., Smith, A., Hill, B., Peddle, D., Coburn, C. and Blackshaw, R. (2008) Hybrid Segmentation-Artificial Neural Network Classification of High Resolution Hyperspectral Imagery for Site-Specific Herbicide Management in Agriculture. *Photogrammetric Engineering and Remote Sensing*, **74**, 1249-1257. https://doi.org/10.14358/PERS.74.10.1249

[7] Smith, A.M. and Blackshaw, R.E. (2003) Weed-Crop Discrimination Using Remote Sensing: A Detached Leaf Experiment. *Weed Technology*, **17**, 811-820. https://doi.org/10.1614/WT02-179

[8] Fletcher, R.S. (2016) Using Vegetation Indices as Input into Random Forest for Soybean and Weed Classification. *American Journal of Plant Sciences*, **7**, 2186-2198. https://doi.org/10.4236/ajps.2016.715193

[9] Fletcher, R.S. and Reddy, K.N. (2016) Random Forest and Leaf Multispectral Reflectance Data to Differentiate Three Soybean Varities from two Pigweeds. *Computers and Electronics in Agriculture*, **128**, 199-206. https://doi.org/10.1016/j.compag.2016.09.004

[10] Karimi, Y., Prasher, S.O., McNairn, H., Bonnell, R.B., Dutilleul, P. and Goel, P.K. (2005) Classification Accuracy of Discriminant Analysis, Artificial Neural Networks, and Decision Trees for Weed and Nitrogen Stress Detection in Corn. *Transactions of the American Society of Agricultural Engineers*, **48**, 1261-1268.

[11] Karimi, Y., Prasher, S.O., Patel, R.M. and Kim, S.H. (2006) Application of Support Machine Technology for Weed and Nitrogen Stress Detection in Corn. *Computers and Electronics in Agriculture*, **51**, 99-109. https://doi.org/10.1016/j.compag.2005.12.001

[12] Thenkabail, P.S., Glumma, M.K., Teluguntla, P. and Mohammed, I.A. (2014) Hyperspectral Remote Sensing of Vegetation and Agricultural Crops. *Photogrammetric Engineering and Remote Sensing*, **80**, 697-709.

[13] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. https://doi.org/10.1023/A:1010933404324

[14] Fernández-Delgado, M., Cernadas, E., Barro, S. and Amorim, D. (2014) Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research*, **15**, 3133-3181.

[15] Belgiu, M. and Drăgut, L. (2016) Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, **114**, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

[16] Prasad, K.A., Gnanappazham, L., Selvam, V., Ramasubramanian, R. and Kar, C.S. (2015) Developing a Spectral Library of Mangrove Species of Indian East Coast Using Field Spectroscopy. *Geocarto International*, **30**, 580-599. https://doi.org/10.1080/10106049.2014.985743

[17] Savitzky, A. and Golay, M. (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, **36**, 1627-1639. https://doi.org/10.1021/ac60214a047

[18] Lehnert, L.W., Meyer, H. and Bendix, J. (2015) HSDAR: Manage, Analyse and Simulate Hyperspectral Data in R. R Package Version 0.3.0. https://cran.rproject.org/web/packages /hsdar/index.html

[19] R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[20] Strobl, C., Hothorn, S. and Zeileis, A. (2009) Party On! A New, Conditional Variable Importance Measure for Random Forests Available in the Party Package. Technical Report Number 050, Department of Statistics, University of Munich, Munich.

[21] Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T. and Zeileis, A. (2008) Conditional Variable Importance for Random Forests. *BMC Bioinformatics*, **9**, 307.
http://www.biomedcentral.com/1471-2105/9/307
https://doi.org/10.1186/1471-2105-9-307

[22] Strobl, C., Boulesteix, A.L., Zeileis, A. and Hothorn T. (2007) Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics*, **8**, 25. http://www.biomedcentral.com/1471-2105/8/25
https://doi.org/10.1186/1471-2105-8-25

[23] Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A. and Van Der Laan, M. (2006) Survival Ensembles. *Biostatistics*, **7**, 355-373.
https://doi.org/10.1093/biostatistics/kxj011

[24] Congalton, R. and Green, K. (2009) Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. 2nd Edition, CRC/Taylor and Francis, Boca Raton, 183 p.

[25] Adam, E., Deng, H., Odindi, J., Abdel-Rahman, E.M. and Mutanga, O. (2017) Detecting the Early Stage of Phaeosphaeria Leaf Spot Infestations in Maize Crop Using in situ Hyperspectral Data and Guided Regularized Random Forest Algorithm. *Journal of Spectroscopy*, **2017**, Article ID: 6961387.

[26] Thenkabail, P.S., Mariotto, I., Gumma, M.K., Middleton, E.M., Landis, A.D.R. and Huemmrich, F.K. (2013) Selection of Hyperspectral Narrowbands (HNBs) and Composition of Hyperspectral Two Band Vegetation Indices (HVIs) for Biophysical Characterization and Discrimination of Crop Types Using Field Reflectance and Hyperion/E0-1 Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **6**, 427-438.
https://doi.org/10.1109/JSTARS.2013.2252601

[27] Thenkabail, P.S., Enclona, E.A., Ashton, M.S. and Van Der Meer, B. (2004) Accuracy Assessments of Hyperspectral Waveband Performance for Vegetation Analysis Applications. *Remote Sensing of Environment*, **91**, 354-376.
https://doi.org/10.1016/j.rse.2004.03.013

[28] Dawson, T.P. and Curran, P.J. (1998) A New Technique for Interpolating the Reflectance Red Edge Position. *International Journal of Remote Sensing*, **19**, 2133-2139. https://doi.org/10.1080/014311698214910

[29] Munden, R., Curran, P.J. and Catt, J.A. (1994) The Relationship between Red Edge and Chlorophyll Concentration in Broadbalk Winter Wheat Experiment at Rothamsted. *International Journal of Remote Sensing*, **15**, 705-709.
https://doi.org/10.1080/01431169408954110

[30] Belanger, M.J., Miller, J.R. and Boyer, M.G. (1995) Comparative Relationships between Some Red Edge Parameters and Seasonal Leaf Chlorophyll Concentrations. *Canadian Journal of Remote Sensing*, **21**, 16-21.
https://doi.org/10.1080/07038992.1995.10874592

[31] Danson, F.M. and Plummer, S.E. (1995) Red-Edge Response to Forest Leaf Area Index. *International Journal of Remote Sensing*, **16**, 183-188.
https://doi.org/10.1080/01431169508954387

[32] Filella, I. and Penuelas J. (1994) The Red Edge Position and Shape as Indicators of Plant Chlorophyll Content, Biomass and Hydric Status. *International Journal of Remote Sensing*, **15**, 1459-1470. https://doi.org/10.1080/01431169408954177