Computers and Electronics in Agriculture 128 (2016) 199–206

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds

Reginald S. Fletcher *, Krishna N. Reddy

United States Department of Agriculture, Agricultural Research Service, Crop Production Systems Research Unit, P.O. Box 350, Stoneville, MS 38776, United States

ARTICLE INFO

Article history:
Received 21 September 2015
Received in revised form 31 August 2016
Accepted 6 September 2016
Available online 15 September 2016

Keywords:
Light reflectance
Machine learning
Palmer amaranth
Redroot pigweed

ABSTRACT

Accurate weed identification is a prerequisite for implementing site-specific weed management in crop production. Palmer amaranth (*Amaranthus palmeri* S. Wats.) and redroot pigweed (*Amaranthus retroflexus* L.) are two common pigweeds that reduce soybean [*Glycine max* (L.) Merr.] yields in the southeastern United States. The objective of this study was to evaluate leaf multispectral reflectance data as input into the random forest machine learning algorithm to differentiate three soybean varieties (Progeny 4928, Progeny 5160, and Progeny 5460) from Palmer amaranth and redroot pigweed. Leaf reflectance measurements of soybean, Palmer amaranth, and redroot pigweed plants grown in a greenhouse were collected with a plant contact probe attached to a hyperspectral spectroradiometer. Data were obtained at the vegetative growth stage of the plants on two dates, June 30, 2014, and September 17, 2014. The hyperspectral data were aggregated to sixteen multispectral bands (viz. coastal, blue, green, yellow, red, red-edge, near-infrared 1 and 2, and shortwave-infrared 1–8) mimicking those recorded by the WorldView-3 satellite sensor. Classifications were binary, meaning one soybean variety versus one weed tested per classification. Random forest classification accuracies were determined with a confusion matrix, incorporating user's, producer's, and overall accuracies and the kappa coefficient. User's, producer's, and overall accuracies of the soybean weed classifications ranged from 93.8% to 100%. Kappa results (values of 0.93–0.97) indicated an excellent agreement between the classes predicted by the models and the actual reference data. Shortwave-infrared bands were ranked the most important variables for distinguishing the pigweeds from the soybean varieties. These results suggest that random forest and leaf multispectral reflectance data could be used as tools to differentiate soybean from two pigweeds with a potential application of this technology in site-specific weed management programs.

1. Introduction

Amaranthus species, commonly known as pigweeds, reduce crop yield and quality throughout the United States and Canada. Their aggressive growth habit allows pigweeds to outgrow and out-compete crops for water and soil nutrients (Teyker et al., 1991; Blackshaw and Brandt, 2008). Pigweed populations ranging from one to three plants per three meter of row have caused significant yield losses in corn (*Zea mays* L.) and soybean [*Glycine max* (L.) (Merr.)] (Klingman and Oliver, 1994; Knezevic et al., 1994; Massinga et al., 2001). In agricultural systems, fertilizer added to the soil increases pigweeds biomass and seed production (Egley, 1986). Pigweeds produce numerous seeds that are viable for three to eight years depending on climate and burial depth in soil. They readily adapt to different crop production systems (including crop rotations) and control tactics, and evolved resistance to herbicides (Brainard et al., 2007; Volenberg et al., 2007; Fugate, 2009). Various management strategies are available for producers to use for combating pigweeds. To effectively implement weed management strategies, reduce the use of herbicides, and protect the environment, producers need effective ways to distinguish pigweeds from crops.

Automatic classification methods using plant light reflectance measurements as inputs have shown promise as tools to discriminate crops from weeds (Koger et al., 2003; Smith and Blackshaw, 2003; Yang et al., 2004; Gómez-Casero et al., 2010; Nieuwenhuizen et al., 2010; de Castro et al., 2012; Deng et al., 2014, 2016). Neto et al. (2006) used elliptic fourier and discriminant analyses and multispectral reflectance data to identify young soybean, sunflower (*Helianthus pumilus* L.), redroot pigweed (*Amaranthus retroflexus* L.), and velvetleaf (*Abutilon theophrasti* Medicus) plants based on leaf shape. Zhang et al. (2012) distinguished tomato (*Solanum lycopersicum* L.) plants from black

* Corresponding author.
E-mail address: reginald.fletcher@ars.usda.gov (R.S. Fletcher).

http://dx.doi.org/10.1016/j.compag.2016.09.004
0168-1699/Published by Elsevier B.V.

nightshade (*Solanum nigrum* L.) and redroot pigweed with canonical Bayesian classifiers and canopy hyperspectral reflectance measurments. Deng et al. (2016) showed that leaf hyperspectral reflectance data subjected to Bayesian discriminant analysis could be used to distinguish weeds from seedling cabbages (*Brassica oleracea* L.). They used principal component analysis to reduce the dataset to eight spectral bands prior to using the Bayesian discriminant analysis. Eddy et al. (2014) used artificial neural networks and hyperspectral imagery to distinguish wild oats (*Avena fatua* L.) and redroot pigweed from field pea (*Pisum sativum* L.), spring wheat (*Triticum aestivum* L.), and canola (*Brassica* spp. L.). They reduced a 61-band hyperspectral dataset to 7 bands and achieved classification accuracies ranging from 88% to 94%. The 7-band dataset results were nearly equivalent to the full 61-band hyperspectral accuracies ranging from 89% to 95%. Deng et al. (2014) showed that support vector machines, artificial neural networks, and decision tree algorithms could differentiate crops and weeds better in 350–760 nm wavelength range than in the 350–2500 nm range. Nevertheless, with the successes of automatic classification methods and plant light reflectance data for weed crop discrimination, gaps still exist in applications of those technologies for weed detection in agricultural fields. For example, which algorithm(s) work best, should multispectral or hyperspectral data be used as input for crop weed discrimination, and does crop variety affect the ability of the algorithm(s) to differentiate the crop from the weed? These are just a few of the many questions that remain unanswered.

Ensemble machine learning techniques employ a set of classifiers in the decision process, or a combination of different classifiers built on rule-based approaches including maximum voting, product, sum, Bayesian rule or on an iterative error minimization technique to assign an unknown sample to a class (Steele, 2000; Breiman, 2001; Mountrakis et al., 2009; Ghimire et al., 2010). They are gaining popularity as classifiers because the ensemble of classifiers performs well together, leading to ensemble methods being highly reliable classifiers and achieving accuracies equivalent to or better than other classifiers (Steele, 2000; Kotsiantis and Pintelas, 2004; Gislason et al., 2006; Sesnie et al., 2008; Ghimire et al., 2010).

Random forest (Breiman, 2001), an ensemble technique, has been applied extensively to multispectral and hyperspectral satellite imagery by researchers to produce land-cover maps (Pal, 2005; Lawrence et al., 2006; Chan and Paelinckx, 2008; Sesnie et al., 2008; Ghimire et al., 2010). It ranks as one of the best family of classifiers (Fernández-Delgado et al., 2014). Random forest popularity continues to grow because the algorithm is fully automated. Analysts can design powerful models with little experience in using the machine learner, and it is not necessary to have an independent external accuracy assessment dataset (Breiman, 2001; Lawrence et al., 2006). That characteristic of random forest is an advantage, especially in circumstances of small sample sizes.

The random forest algorithm also produces a variable importance ranking, meaning it ranks each variable's prominence in the classification model. This information is valuable to the user in selecting variables to build simpler, more readily interpretable models (Liaw and Wiener, 2002). Other machine learning algorithms do not have the ability to produce variable rankings. Quite often, another algorithm is used to create the variable ranking, and then the selected variables are evaluated by the preferred learner. For a more detailed description of the random forest, consult Breiman (2001).

Presently, no information is available on using multispectral reflectance data as input into the random forest machine learner for pigweed soybean discrimination. The objective of this investigation was to test the performance of random forest machine learner to discriminate soybean from Palmer amaranth and redroot pigweed, two common pigweeds that interfere in soybean production in the southeastern United States. Specifically, the study focused on using multispectral leaf reflectance data obtained within the visible, red-edge, near-infrared, and shortwave-infrared regions of the light spectrum as input variables into the random forest algorithm and on discriminating the two weeds from three soybean varieties.

2. Materials and methods

2.1. Plant descriptions

Three Progeny (P) brand LibertyLink (LL) soybean varieties (P4928LL, P5160LL, and P5460LL Progeny Ag Products, Wynne, Arkansas) and non-glyphosate resistant redroot pigweed and Palmer amaranth (Crop Production Systems Research Unit, USDA-ARS, Stoneville, MS) were grown for the study. Soybean P4928LL has an indeterminate growth habit (i.e., continuation of vegetative growth after flowering) with a maturity group of 4.9, and gray pubescence. Soybean P5160LL and P5460LL have a determinate growth habit (i.e., vegetative growth completed prior to flowering) and maturity group labels of 5.1 and 5.4, respectively. Soybean P5160LL and P5460LL have a tawny and a light tawny hair color; respectively. The soybeans were selected for their differences in growth habit, maturity level, and leaf pubescence.

2.2. Greenhouse experiment

Two greenhouse experiments were conducted at the Crop Production Systems Research Unit, United States Department of Agriculture, Agricultural Research Service, Stoneville, MS facility. On June 13, 2014, and August 28, 2014, soybean and weed seeds were planted in plugs containing commercial potting mix (Pro-Mix, Ultimate Potting Mix, Quakertown, Pennsylvania). Two weeks after germination, thirty plants of each soybean variety and weed species were transferred to individual 2 L pots filled with the commercial potting mix. Plants were watered on three- to four-day intervals as needed. The potting mix consisted of a slow release nitrogen, phosphorus, and potassium fertilizer. The greenhouse was maintained at 28/24 ± 3 °C day/night temperature with natural light supplemented by sodium vapor lamp to provide a 14-h photoperiod.

2.3. Data collection

Leaf reflectance measurements were obtained with a plant probe attached to the fiber optic of a full range hyperspectral spectroradiometer (FieldSpec 3, PANalytical Boulder, Boulder Co., USA). The plant probe was equipped with a light source, allowing the user to acquire reflectance measurements anytime during the day or night. The device measures a 1 cm area. A leaf clip (PANalytical Boulder, Boulder Co., USA) was attached to the contact probe, which had a trigger lock/release gripping system designed to hold the leaf in place without removing it from the plant or causing damage to the plant. The leaf clip was equipped with a two-sided rotating head, with one side having a black panel face and the other side having a white panel face. The black panel is ideal for reflectance measurements; the white panel is perfect for transmittance. The black panel was employed in this study.

The spectroradiometer obtained continuous spectra in the range of 350–2500 nm. Its sampling interval and spectral resolution were 1.4 nm and 3 nm, respectively, within the 350–1000 nm spectral range. The sampling interval and spectral resolution were 2 nm and 10 nm, respectively, within the 1000–2500 nm spectral range. The proprietary software operating the instrument resampled the reflectance data to 1 nm wavelengths.

Reflectance measurements were collected from the most recently matured leaf of each plant. Soybean has a trifoliolate leaf; therefore, the center leaflet (terminal leaflet) of the most recently matured leaf was chosen for data collection. At the selected sample spot of each plant leaf, reflectance measurements were an average of fifteen readings. Leaf reflectance measurements were obtained on June 30, 2014 (first experiment) and September 17, 2014 (second experiment), when plants were at 15 and 18 day-old stage, respectively. This growth stage simulates weeds in field conditions at the time of early postemergence application for weed control. Measurements were obtained for all plants during the vegetative growth stage. The spectroradiometer was calibrated with a white spectralon panel on 15 min. intervals.

2.4. Aggregation of spectral bands

The 1 nm spectral data were aggregated to sixteen multispectral bands mimicking those of the Worldview-3 satellite sensor: coastal (400–450 nm), blue (450–510 nm), green (510–580 nm), yellow (585–625 nm), red (630–690 nm), red-edge (705–745 nm), near-infrared 1 (770–895 nm), near-infrared 2 (860–1040 nm), shortwave-infrared 1 (1195–1225 nm), shortwave-infrared 2 (1550–1590 nm), shortwave-infrared 3 (1640–1680 nm), shortwave-infrared 4 (1710–1750 nm), shortwave-infrared 5 (2145–2185 nm), shortwave-infrared 6 (2185–2225 nm), shortwave-infrared 7 (2235–2285 nm), and shortwave-infrared 8 (2295–2365 nm) (Digital Globe, 2014). These bands were chosen because one or more of them were similar to spectral bands appearing on other ground-based and airborne multispectral remote sensing systems. Currently, the sixteen selected bands are the most comprehensive multispectral measurements available for land and water surveys.

2.5. Deriving and evaluating classification models

The conditional inference version of random forest (cforest) was used to derive models for classifying the June 30, 2014, and the September 17, 2014, data into groups. Spectral data obtained in the same region of the light spectrum are often highly correlated, which was the case for the spectral bands evaluated in this study. Several research studies have shown that biased variable importance rankings are tabulated by the original version of random forest if strong correlation exists among predictor variables (Strobl et al., 2009). Cforest implementation of random forest was designed to better handle correlation among variables, thus providing more accurate and unbiased rankings of the variable importance (Strobl et al., 2009). It employs conditional inference trees as base learners, compared to random forest, which uses classification and regression trees as base learners (Hothorn et al., 2006a; Strobl et al., 2009). Cforest utilizes subsampling without replacement for constructing unbiased decision trees for the forest; whereas, random forest uses bootstrap samples to construct its decision trees. Finally, the cforest algorithm uses the conditional permutation scheme described by Strobl et al. (2009) to determine the variable importance ranking.

Models were created for binary classifications, meaning one soybean variety versus one weed species. Before implementing the algorithm, the user has to set two parameters: (1) *mtry* - the number of randomly preselected variables used in each split and (2) *ntree* - the number of trees in the forest. The default values of five and five hundred were used as the starting point for *mtry* and *ntree*, respectively.

The variable importance ranking was tabulated for each random forest model per classification. Before accepting the variable ranking, the user customarily repeats the process employing the preselected *mtry* and *ntree* values and a different seed (i.e., starting point for random number sequence). Completing this test authenticates the model robustness and stability (Strobl et al., 2009). If the variable ranking order substantially changed from one run to the next, then the ntree value was increased by 1000, and the classification was repeated. This process was duplicated until a stable ranking was obtained.

Model accuracies were determined by creating and evaluating a confusion matrix, consisting of user's, producer's, and overall accuracies and the kappa coefficient on the "out-of-bag samples" (Congalton, 1991; Foody, 2002). The weeds were considered the class of interest, otherwise known as the event. Random forest model development and evaluation was determined with the party package (Hothorn et al., 2006b; Strobl et al., 2007, 2008) of the R software [R version 3.1.2 (October 31, 2014) - Pumpkin Helmet).

3. Results

3.1. Classification results

Error matrix results for each classification are shown in Tables 1–3. User's and producer's accuracies greater than 93% were achieved by the random forest algorithm using the multispectral data as input variables. The lowest per class accuracy was observed for the Palmer amaranth class in the Palmar amaranth versus soybean P5460LL classification for the June 30, 2014 data set. Generally, classification errors occurred when the pigweeds were misclassified as soybean. Overall accuracies and Kappa values were equal or greater than 96.7% and 0.93, respectively in both dates.

3.2. Model parameters

Table 4 summarizes the random forest model parameters used to distinguish the Palmer amaranth and redroot pigweed from the soybean varieties. The number of classification trees ranged from 500 to 5500. The number of classification trees was increased from the default value of 500–obtain stable variable importance rankings.

3.3. Variable importance

Figs. 1–3 illustrate the variable importance rankings of the classifications. Typically, ten of the sixteen spectral bands were ranked as important variables in discriminating the pigweeds from the soybeans. Shortwave-infrared bands were the most important to the models; whereas, near-infrared, red-edge, and visible (i.e., coastal, blue, green, yellow) band rankings were dependent on classification and date. The red band was not important to the classification models.

4. Discussion

Using leaf multispectral reflectance data as input, the random forest algorithm showed excellent potential as a tool to discriminate Palmer amaranth and redroot pigweed from soybean (Tables 1–3). Palmer amaranth or redroot pigweed was the class of interest, otherwise known as the event. The user's accuracy results indicated most errors were related to the pigweeds being misclassified as soybean (Tables 1–3). Kappa statistic values were equal or greater than 0.93, representing almost perfect agreement between the predicted data and the reference data (Landis and Koch, 1977). It appears that the pubescence color of the soybean leaves had no influence on classification accuracy of the algorithm.

Table 1
Error matrix of random forest using the 16-band multispectral leaf reflectance data for Palmer amaranth and redroot pigweed discrimination from soybean variety P4928LL.

		Reference			
Date	Prediction	Palmer amaranth	Soybean P4928LL	Total	User's Accuracy
6/30/2014	Palmer amaranth	29	0	29	100%
	Soybean P4928LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.7%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	
		Reference			
Date	Prediction	Redroot pigweed	Soybean P4928LL	Total	User's Accuracy
6/30/2014	Redroot pigweed	29	0	29	100%
	Soybean P4928LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.7%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	
		Reference			
Date	Prediction	Palmer amaranth	Soybean P4928LL	Total	User's Accuracy
9/17/2014	Palmer amaranth	29	0	29	100%
	Soybean P4928LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.7%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	
		Reference			
Date	Prediction	Redroot pigweed	Soybean P4928LL	Total	User's Accuracy
9/17/2014	Redroot pigweed	29	0	29	100%
	Soybean P4928LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.7%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	

Table 2
Error matrix of random forest using multispectral leaf reflectance data for Palmer amaranth and redroot pigweed discrimination from soybean P5160LL.

		Reference			
Date	Prediction	Palmer amaranth	Soybean P5160LL	Total	User's Accuracy
6/30/2014	Palmer amaranth	29	0	29	100%
	Soybean P5160LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.7%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	
		Reference			
Date	Prediction	Redroot pigweed	Soybean P5160LL	Total	User's Accuracy
6/30/2014	Redroot pigweed	29	0	29	100%
	Soybean P5160LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.7%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	
		Reference			
Date	Prediction	Palmer amaranth	Soybean P5160LL	Total	User's Accuracy
9/17/2014	Palmer amaranth	29	0	29	100%
	Soybean P5160LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.7%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	
		Reference			
Date	Prediction	Redroot pigweed	Soybean P5160LL	Total	User's Accuracy
9/17/2014	Redroot pigweed	29	0	29	100%
	Soybean P5160LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.7%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	

Table 3
Error matrix of random forest using multispectral leaf reflectance data for Palmer amaranth and redroot pigweed discrimination from soybean P5460LL.

Date	Prediction	Reference		Total	User's Accuracy
		Palmer amaranth	Soybean P5460LL		
6/30/2014	Palmer amaranth	28	0	28	100%
	Soybean P5460LL	2	30	32	93.8%
	Total	30	30		
	Producer's accuracy	93.3%	100%		
	Overall accuracy	96.7%	Cohen's kappa	0.93	
		Reference			
Date	Prediction	Redroot pigweed	Soybean P5460LL	Total	User's Accuracy
6/30/2014	Redroot pigweed	29	0	29	100%
	Soybean P5460LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.8%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	
		Reference			
Date	Prediction	Palmer amaranth	Soybean P5460LL	Total	User's Accuracy
9/17/2014	Palmer amaranth	29	1	30	96.7%
	Soybean P5460LL	1	29	30	96.7%
	Total	30	30		
	Producer's accuracy	96.7%	96.7%		
	Overall accuracy	96.7%	Cohen's kappa	0.93	
		Reference			
Date	Prediction	Redroot pigweed	Soybean P5460LL	Total	User's Accuracy
9/17/2014	Redroot pigweed	29	0	29	100%
	Soybean P5460LL	1	30	31	96.8%
	Total	30	30		
	Producer's accuracy	96.7%	100%		
	Overall accuracy	98.3%	Cohen's kappa	0.97	

Table 4
Random forest model parameters used to distinguish Palmer amaranth and redroot pigweed from soybean varieties.

Classification	*mtry*[a]	*ntree* (June 30, 2014)	*ntree* (September 17, 2014)
Palmer amaranth-soybean P4928LL	5	3500	4500
Redroot pigweed-soybean P4928LL	5	1500	500
Palmer amaranth-soybean P5160LL	5	1500	4500
Redroot pigweed-soybean P5160LL	5	500	1500
Palmer amaranth-soybean P5460LL	5	500	1500
Redroot pigweed-soybean P5460LL	5	1500	5500

[a] *mtry* = number of randomly preselected variables; *ntree* = number of trees used in the classification.

Shortwave-infrared wavelengths were ranked as the most important variables used by the models for pigweed soybean discrimination (Figs. 1–3). Water concentration in plant tissues affects their ability to reflect shortwave-infrared light (Gausman, 1985), indicating leaf succulence played a role in discriminating Palmer amaranth and redroot pigweed from soybeans. Our findings concurred with Gray et al. (2009) who observed that shortwave-infrared wavelengths were essential for differentiating soybean, soil, and six broadleaf weeds. Their study focused on plant canopy hyperspectral reflectance measurements and principal component analysis and linear discriminate analysis for crop weed discrimination; whereas, our study evaluated leaf multispectral data and the random forest machine learning algorithm to distinguish soybean and the two selected pigweeds.

Other studies have shown that visible (400–670 nm), red-edge (680–760 nm), and near-infrared light (770–1290 nm) reflectance of plant leaves and canopies were important variables for crop weed discrimination (de Castro et al., 2012; Shapira et al., 2013). Plant pigments including chlorophyll (i.e., a and b), carotenes, and xanthophylls, influence visible light reflectance and absorption of plant leaves and canopies (Gausman, 1985). The combination of chlorophyll absorption and strong scattering of light by the leaf internal cellular structure affect the red-edge (680–760 nm) reflectance of plant leaves and canopies (Ray et al., 1993). The internal leaf structure and multiple leaf layers influence the near-infrared light reflectance properties of plant leaves and canopies (Gausman, 1985). In our study, the near-infrared bands were predominantly more important to the classification models than the visible bands that often received variable importance scores close to zero (Figs. 1–3). These findings suggested that the leaf internal structure was somewhat important to the separation of the weed and soybean classes and that leaf pigments played a minute role in the classifications.

Excluding the June 30, 2014, redroot pigweed-soybean P5160LL and the Palmer amaranth-soybean P5460LL classifications and the September 17, 2014, redroot pigweed-soybean P4928LL classification, the number of decision trees was increased from the default value of 500 to stabilize the variable of importance rankings (Table 4). Others have recommended using large number of decision trees in random forest classifications to achieve stable variable importance rankings (Liaw and Wiener, 2002; Strobl et al., 2009). Classification accuracies were not affected when the tree numbers

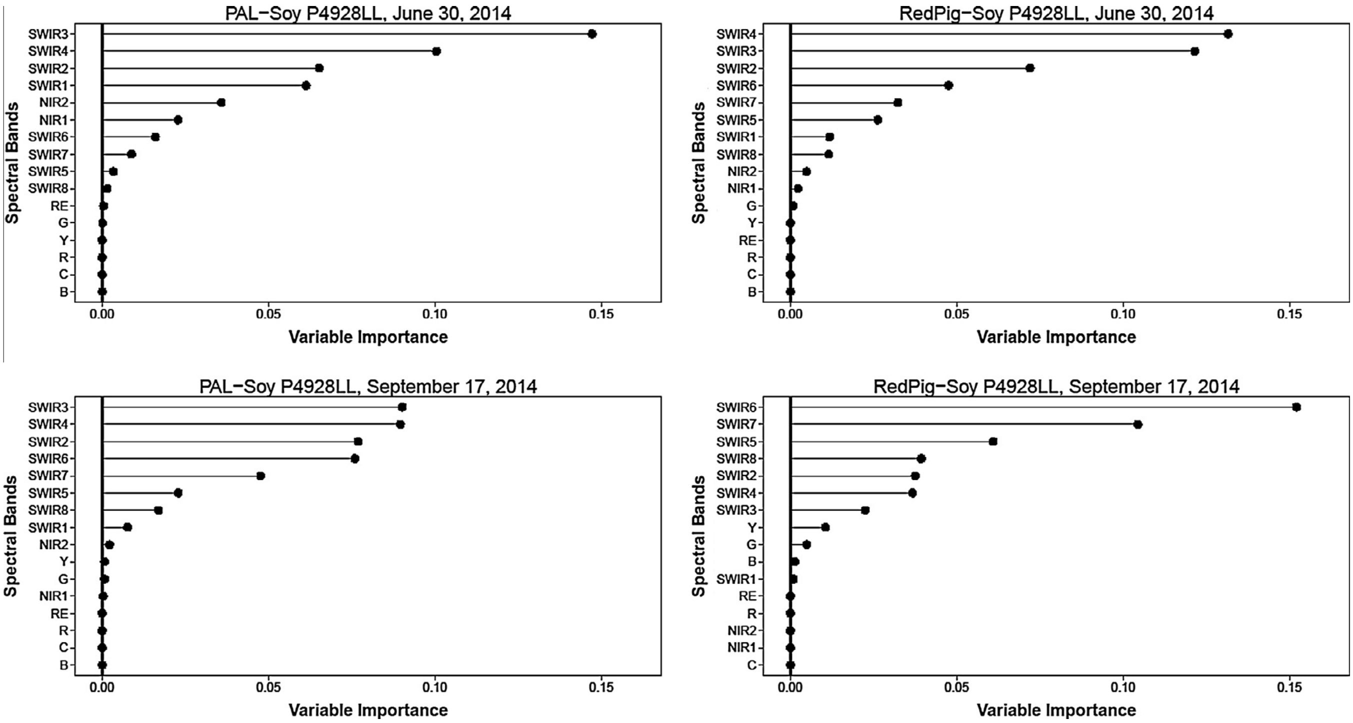

Fig. 1. Variable importance rankings of the spectral bands used by the random forest models in the pigweeds versus soybean classifications. PAL-Palmer amaranth, RedPig-Redroot pigweed, Soy P4928LL-Soybean variety P4928LL, SWIR-shortwave-infrared, NIR-near-infrared, RE-red-edge, G-green, Y-yellow, R-red, C-coastal, and B-blue.

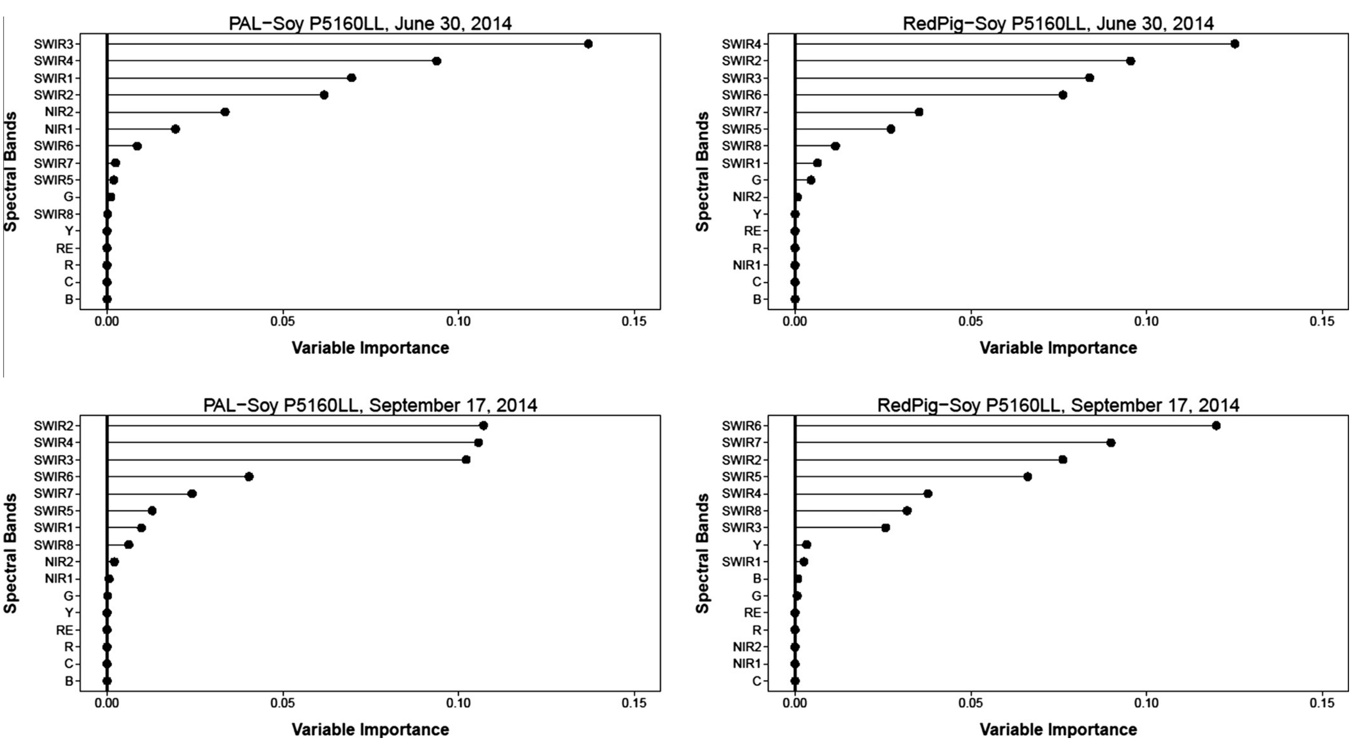

Fig. 2. Variable importance rankings of the spectral bands used by the random forest models in the weed versus soybean classifications. PAL-Palmer amaranth, RedPig-redroot pigweed, Soy P5160LL-soybean variety P5160LL, SWIR-shortwave-infrared, NIR-near-infrared, RE-red-edge, G-green, Y-yellow, R-red, C-coastal, and B-blue.

were increased, a characteristic commonly observed with random forest classifications (Rodriguez-Galiano et al., 2012).

To put this study into perspective, the authors stress that reflectance measurements were obtained from plant leaves and not plant canopies. The former represents pure spectra. At the canopy level, leaf angle to the sensor, leaf location in the canopy and inter-canopy shadowing will affect visible, red-edge, near-infrared, and shortwave-infrared regions of the light spectrum. Those aspects may alter to some degree classification accuracy of the random forest model and the variable importance ranking. Also, it is impor-

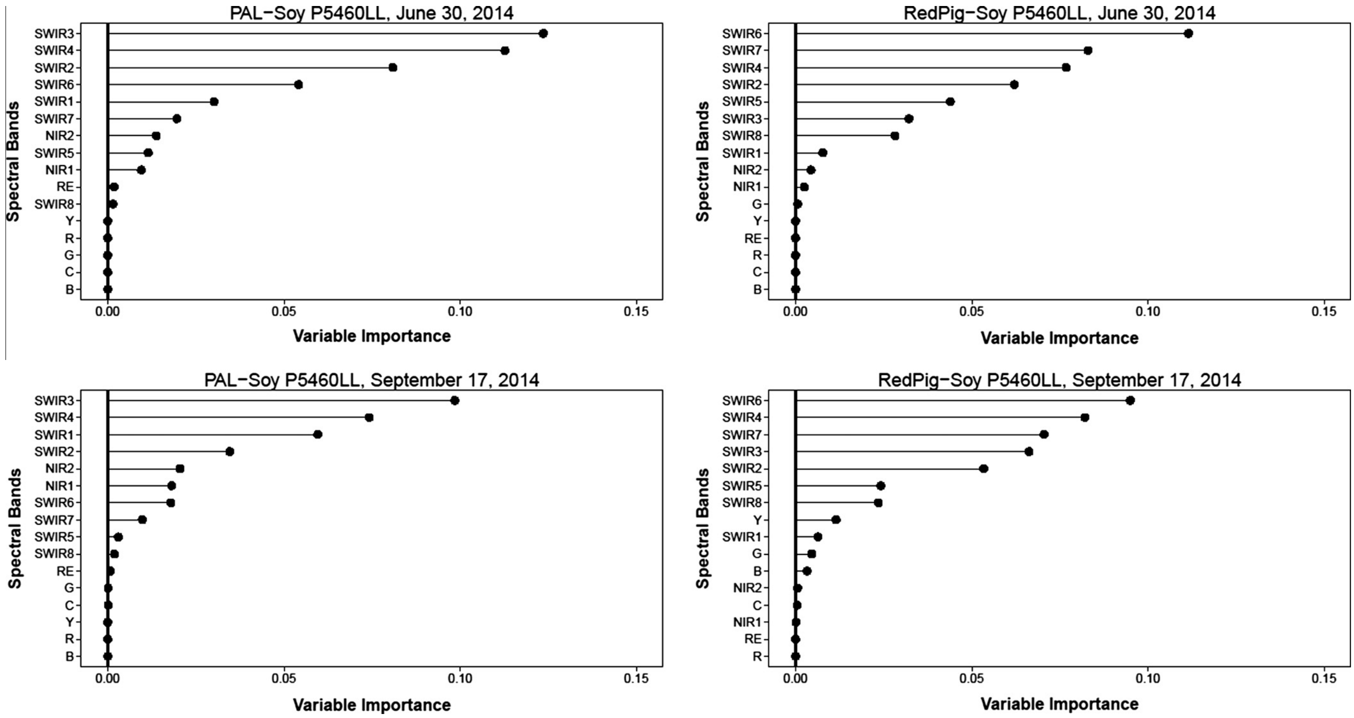

Fig. 3. Variable importance rankings of the spectral bands used by the random forest models in the weed versus soybean classifications. PAL-Palmer amaranth, RedPig-redroot pigweed, Soy P5460LL-soybean variety P5460LL, SWIR-shortwave-infrared, NIR-near-infrared, RE-red-edge, G-green, Y-yellow, R-red, C-coastal, and B-blue.

tant to note that this study focused on using random forest and multispectral data for soybean weed discrimination; however, there may be other algorithms that work as well with multispectral data for soybean pigweed discrimination. Overall, this study indicated that leaf multispectral reflectance data could be used by the random forest algorithm to differentiate Palmer amaranth and redroot pigweed from soybean. Additionally, the algorithm provided general information on which variables were most important in soybean and pigweed discrimination.

5. Conclusions

In this study, the random forest machine learner and leaf multispectral reflectance data were explored as tools to discriminate soybean from Palmer amaranth and redroot pigweed, two weeds causing yield reductions in soybean throughout the southeastern United States. The random forest algorithm distinguished the two weeds from three soybean varieties. Spectral bands sensitive to water concentration (i.e., shortwave-infrared bands) in plant tissues were essential to the models for weed soybean discrimination. Findings support further application of the random forest machine learner along with remotely-sensed multispectral data as tools for pigweed crop discrimination with future implications for site-specific management of weeds.

Acknowledgments

The authors thank Dr. Neal Teaster for providing pigweed seeds. Thanks also to Mr. Milton Gaston Jr., Mr. Arrington Smith, Ms. Keysha Hamilton, Mr. David Fisher, Ms. Raven Thompson, and Ms. Keyanna Nealon for their assistance in raising plants and data collection, and Dr. Ken Fisher and Dr. Webushet Tadessee for their critical review of the manuscript.

References

Blackshaw, R.E., Brandt, R.N., 2008. Nitrogen fertilizer rate effects on weed competitiveness is species dependent. Weed Sci. 56, 743–747. http://dx.doi.org/10.1614/WS-08-065.1.

Brainard, D.C., DiTommaso, A., Mohler, C.A., 2007. Intraspecific variation in seed characteristics of Powell amaranth (*Amaranthus powellii*) from habitats with contrasting crop rotation histories. Weed Sci. 55, 218–226. http://dx.doi.org/10.1614/WS-06-134.1.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. http://dx.doi.org/10.1023/A:101093 3404324.

de Castro, A.I., Jurado-Expósito, M., Gómez-Casero, M.T., López-Granados, F., 2012. Applying neural networks to hyperspectral and multispectral field data for discrimination of cruciferous weeds in winter crops. Sci. World J., 11 http://dx.doi.org/10.1100/2012/630390 630390.

Chan, J.C.W., Paelinckx, D., 2008. Evaluation of random forest and adaboost treebased ensemble classification and spectral band selection for ecotype mapping using airborne hyperspectral imagery. Remote Sens. Environ. 112 (6), 2999–3011. http://dx.doi.org/10.1016/j.rse.2008.02.011.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46. http://dx.doi.org/10.1016/0034-4257(91)90048-B.

Deng, W., Huang, Y., Zhao, C., Wang, X., 2014. Discrimination of crop and weeds on visible and visible/near infrared spectrums using support vector machine, artificial neural network, and decision tree. Sens. Transducers 26, 26–34.

Deng, W., Huang, Y., Zhao, C., Chen, L., Wang, X., 2016. Bayesian discriminant analysis of plant leaf hyperspectral reflectance for identification of weeds from cabbages. Afr. J. Agric. Res. 11 (7), 551–562.

Digital Globe, 2014. Worldview 3 data sheet. <https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/95/DG_WorldView3_DS_forWeb_0.pdf>.

Eddy, P.R., Smith, A.M., Hill, B.D., Peddle, D.R., Coburn, C.A., Blackshaw, R.E., 2014. Weed and crop discrimination using hyperspectral image data and reduced bandsets. Can. J. Remote Sens. 39 (6), 481–490. http://dx.doi.org/10.5589/m14-001.

Egley, G.H., 1986. Stimulation of weed seed germination in soil. Rev. Weed Sci. 2, 67–89.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15, 3133–3181.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201. http://dx.doi.org/10.1016/S0034-4257(01)00295-4.

Fugate, L., 2009. Pigweed causing farmers to rethink farming methods. University of Arkansas Division of Agriculture Cooperative Extension Service News – October 2009.

Gausman, H., 1985. Plant Leaf Optical Properties. Texas Tech Press, Lubbock, Texas.

Ghimire, B., Rogan, J., Miller, J., 2010. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. Remote Sens. Lett. 1, 45–54. http://dx.doi.org/10.1080/01431160903252327.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random forests for land cover classification. Pattern Recogn. Lett. 27 (4), 294–300. http://dx.doi.org/10.1016/j.patrec.2005. 08.011.

Gómez-Casero, M.T., Castillejo-González, I.L., García-Ferrer, A., 2010. Spectral discrimination of wild oat and canary grass in wheat fields for less herbicide application. Agron. Sustain Dev. 30 (3), 689–699. http://dx.doi.org/10.1051/agro/2009052.

Gray, C.J., Shaw, D.R., Bruce, L.M., 2009. Utility of hyperspectral reflectance for differentiating soybean (*Glycine max*) and six weed species. Weed Tech. 23, 108–119. http://dx.doi.org/10.1614/WT-07-117.1.

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., van der Laan, M., 2006a. Survival ensembles. Biostatistics 7 (3), 355–373. http://dx.doi.org/10.1093/biostatistics/kxj011.

Hothorn, T., Hornik, K., Zeileis, A., 2006b. Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph Stat. 15, 651–674. http://dx.doi.org/10.1198/1061 86006X133933.

Klingman, T.E., Oliver, L.R., 1994. Palmer amaranth (*Amaranthus palmeri*) interference in soybeans (*Glycine max*). Weed Sci. 42, 523–527 http://www.jstor.org/stable/4045448.

Knezevic, S.Z., Weise, S.F., Swanton, C.J., 1994. Interference of redroot pigweed (*Amaranthus retroflexus*) in corn (*Zea mays*). Weed Sci. 42, 568–573 http://www.jstor.org/stable/4045456.

Koger, C.H., Bruce, L.M., Shaw, D.R., Reddy, K.N., 2003. Wavelet analysis of hyperspectral reflectance data for detecting pitted morning glory (*Ipomoea lacunosa*) in soybean (*Glycine max*). Remote Sens. Environ. 86 (1), 108–119. http://dx.doi.org/10.1016/S0034-4257(03) 00071-3.

Kotsiantis, S., Pintelas, P., 2004. Combining bagging and boosting. IJCI 1 (4), 324–333.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174. http://dx.doi.org/10.2307/2529310.

Lawrence, R.L., Wood, S.D., Shelley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and breiman cutler classifications (random forest). Remote Sens. Environ. 100, 356–362. http://dx.doi.org/10.1016/j.rse.2005.10.014.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. RNews 2 (3), 18–22.

Massinga, R.A., Currie, R.S., Horak, M.J., Boyer Jr., J., 2001. Interference of Palmer amaranth in corn. Weed Sci. 49, 202–208. http://dx.doi.org/10.1614/0043-1745(2001)049%5B0202: IOPAIC%5D2.0.CO;2.

Mountrakis, G., Watts, R., Luo, L., Wang, J., 2009. Developing collaborative classifiers using an expert-ba sed model. Photogramm. Eng. Remote Sens. 75 (7), 831–844. http://dx.doi.org/10.14358/PERS.75.7.831.

Neto, J.C., Meyer, G.E., Jones, D.D., Samal, A.K., 2006. Plant species identification using elliptic Fourier leaf shape analysis. Comput. Electron. Agric. 50, 121–134. http://dx.doi.org/10.1016/j.compag.2005.09.004.

Nieuwenhuizen, A.T., Hofstee, J.W., van de Zande, J.C., Meuleman, J., van Henten, E.J., 2010. Classification of sugar beet and volunteer potato reflection spectra with a neural network and statistical discriminant analysis to select discriminative wavelengths. Comput. Electron. Agric. 73 (2), 146–153. http://dx.doi.org/10.1016/j.compag.2010.05.008.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26 (1), 217–222. http://dx.doi.org/10.1080/01431160412331269698.

Ray, T., Murray, B., Chehbouni, A., Njoku, E., 1993. The red edge in arid region vegetation: 340–1060 nm spectra. Summaries of the 4th Annual JPL Airborne Geoscience Workshop. Vol. 1: AVIRIS Workshop. Jet Propulsion Laboratory, Pasadena, CA.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104. http://dx.doi.org/10.1016/j.isprsjprs.2011. 11.002.

Sesnie, S., Gessler, P., Finegan, B., Thessler, S., 2008. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. Remote Sens. Environ. 112 (5), 2145–2159. http://dx.doi.org/10.1016/j.rse.2007.08.025.

Shapira, U., Herrmanna, I., Karnielia, A., Bonfilb, D.J., 2013. Field spectroscopy for weed detection in wheat and chickpea fields. Int. J. Remote Sens. 34 (17), 6094–6108. http://dx.doi.org/10.1080/01431161.2013.793860.

Smith, A.M., Blackshaw, R.E., 2003. Weed-crop discrimination using remote sensing: a detached leaf experiment. Weed Technol. 17 (4), 811–820. http://dx.doi.org/10.1614/WT02179.

Steele, B.M., 2000. Combining multiple classifiers: an application using spatial and remotely sensed information for land cover type mapping. Remote Sens. Environ. 74 (3), 545–556. http://dx.doi.org/10.1016/S0034-4257(00)00145-0.

Strobl, C., Boulesteix, A.L., Zeileis, A., 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.

Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9, 307 http://www.biomedcentral.com/1471-2105/9/307.

Strobl, C., Hothorn, S., 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.

Teyker, R.H., Hoelzer, H.D., Liebl, R.A., 1991. Maize and pigweed response to nitrogen supply and form. Plant Soil. 135, 287–292. http://dx.doi.org/10.1007/BF00010918.

Volenberg, D.S., Patzoldt, W.L., Hager, A.G., Tranel, P.J., 2007. Responses of contemporary and historical waterhemp (*Amaranthus tuberculatus*) accessions to glyphosate. Weed Sci. 55, 327–333. http://dx.doi.org/10.1614/WS-06-121.

Yang, C.C., Prasher, S.O., Goel, P.K., 2004. Differentiation of crop and weeds by decision-tree analysis of multi-spectral data. Trans. ASAE 47 (3), 873–879. http://dx.doi.org/10.13031/2013.16084.

Zhang, Y., Slaughter, D.C., Staab, E.S., 2012. Robust hyperspectral vision-based classification for multi-season weed mapping. ISPRS J. Photogramm. 69, 65–73. http://dx.doi.org/10.1016/j.isprsjprs.2012.02.006.