Available online at www.sciencedirect.com
SCIENCE DIRECT®
Remote Sensing of Environment 86 (2003) 108–119
Remote Sensing of Environment
www.elsevier.com/locate/rse

Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (*Ipomoea lacunosa*) in soybean (*Glycine max*)

Cliff H. Koger[a,*], Lori M. Bruce[b,1], David R. Shaw[a,2], Krishna N. Reddy[c,3]

[a] *Department of Plant and Soil Sciences, 117 Dorman Hall, Box 9555, Mississippi State University, Mississippi State, MS 39762, USA*
[b] *Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA*
[c] *USDA-ARS Southern Weed Science Research Unit, 141 Experiment Station Road, PO Box 350, Stoneville, MS 38776, USA*

Received 28 June 2002; received in revised form 28 February 2003; accepted 8 March 2003

Abstract

This research determined the potential for wavelet-based analysis of hyperspectral reflectance signals for detecting the presence of early season pitted morningglory when intermixed with soybean and soil. Ground-level hyperspectral reflectance signals were collected in a field experiment containing plots of soybean and plots containing soybean intermixed with pitted morningglory in a conventional tillage system. The collected hyperspectral signals contained mixed reflectances for vegetation and background soil in each plot. Pure reflectance signals were also collected for pitted morningglory, soybean, and bare soil so that synthetically mixed reflectance curves could be generated, evaluated, and the mixing proportions controlled. Wavelet detail coefficients were used as features in linear discriminant analysis for automated discrimination between the soil + soybean and the soil + soybean + pitted morningglory classes. A total of 36 different mother wavelets were investigated to determine the effect of mother wavelet selection on the ability to detect the presence of pitted morningglory. When the growth stage was two to four leaves, which is still controllable with herbicide, the weed could be detected with at least 87% accuracy, regardless of mother wavelet selection. Moreover, the Daubechies 3, Daubechies 5, and Coiflet 5 mother wavelets resulted in 100% classification accuracy. Most of the best wavelet coefficients, in terms of discriminating ability, were derived from the red-edge and the near-infrared regions of the spectrum. For comparison purposes, the raw spectral bands and principal components were evaluated as possible discriminating features. For the two-leaf to four-leaf weed growth stage, the two methods resulted in classification accuracies of 83% and 81%, respectively. The wavelet-based method was shown to be very promising in detecting the presence of early season pitted morningglory in mixed hyperspectral reflectances.

Keywords: Hyperspectral; Wavelet; Reflectance; Weeds; Discriminant; Soybean; Principle components

1. Introduction

Weeds typically are not distributed evenly over entire fields, but often aggregated into patches due to factors such as soil pH (Weaver & Hamill, 1985), nutrient levels (Banks, Santlemann, & Tucker, 1976), cation exchange capacity, and topography (Medlin et al., 2001). Traditionally, intensive scouting has been the only means of providing information concerning weed distributions. However, scouting is labor- and time-intensive, and even after determining where weed patches occur in fields, the entire field is often treated with herbicide to control weed patches. Applying herbicides where weeds do not occur can lead to excessive herbicide usage, costs associated with herbicide use, time required for application, soil compaction, and increased risk of herbicide movement to off-site areas (Cousens & Woolcock, 1987; Felton, Doss, Nash, & McCoy, 1991; Swanton & Weise, 1991; Thompson, Stafford, & Miller, 1991).

Rising environmental concerns about pesticide use and increasing production costs associated with their use has researchers investigating the potential use of remote sensing for detecting weed distributions in agricultural fields. Weed distribution maps, developed from remotely sensed imagery,

* Corresponding author. Current affiliation: USDA-ARS Southern Weed Science Research Unit, Stoneville, MS, USA. Fax: +1-662-686-5422.
E-mail addresses: ckoger@ars.usda.gov (C.H. Koger), bruce@ece.msstate.edu (L.M. Bruce), dshaw@pss.msstate.edu (D.R. Shaw), kreddy@ars.usda.gov (K.N. Reddy).
[1] Fax: +1-662-325-2298.
[2] Fax: +1-662-325-8742.
[3] Fax: +1-662-686-5422.

0034-4257/03/$ - see front matter © 2003 Elsevier Science Inc. All rights reserved.
doi:10.1016/S0034-4257(03)00071-3

could be used along with differential geographical positioning systems (DGPS) and site-specific herbicide applicators for targeting herbicide inputs only to areas containing weed densities above economic thresholds (Christensen, Walter, & Heisel, 1999; Thornton, Fawcett, Dent, & Perkins, 1990). The interest in multispectral remote sensing for characterizing agricultural objectives such as weed detection has been researched and used with some success to detect weeds in rangelands (Everitt, Escobar, Alaniz, Davis, & Richardson, 1996; Lass, Carson, & Callihan, 1996) and agricultural crops such as soybean and cotton (*Gossypium hirsutum* L.) (Medlin, Shaw, Gerard, & Lamastus, 2000; Richardson, Menges, & Nixon, 1985). However, for agricultural crop scenarios, soil background reflectance often interferes with early-season weed detection, which is when most weed control options are implemented (Barrentine, 1974). One reason for the interference from soil background reflectance is that most multispectral reflectance data contains three to seven broad-band measurements, which are often restricted to the visible and near-infrared portions of the electromagnetic spectrum. These portions of the spectrum are typically influenced by soil background at low vegetation cover (Elvidge & Mouat, 1985). Due to this interference, the potential use for other remote sensing sources such as hyperspectral data in detecting weeds has increased in recent years.

Most of the new hyperspectral sensors contain upwards of 200 spectral channels (Thenkabail, Smith, & Pauw, 2000). Some hand-held hyperspectral sensors have been developed that contain more than 1500 spectral channels. This increase in the amount of data, when compared to multispectral (three to seven bands), and the fact that hyperspectral data is often collected from several regions of the electromagnetic spectrum (ultraviolet, visible, near-, mid-, and far-infrared) has led to a variety of potential uses for hyperspectral data, such as estimating crop yield (Shibayama & Akiyama, 1991), chlorophyll content (Blackburn, 1998), and photosynthetic and crop leaf area index and biomass (Thenkabail et al., 2000). If hyperspectral data can be used to characterize physical and physiological crop factors, there may also be potential for narrow-band hyperspectral measurements collected across a wide array of the spectrum for weed detection in agricultural fields.

While hyperspectral data provide the opportunity for more detailed analysis of on-ground materials, the high-dimensional data generated by the hyperspectral sensors create a new challenge for conventional spectral data analysis techniques (Jiminez & Landgrebe, 1999). It has been proven that the high-dimensional data space is mostly empty, due to the fact that the volume of a hyperspectral dataset concentrates in the corners and the volume of a hypersphere concentrates in an outer shell (Jiminez & Landgrebe, 1998). Principally, two solutions exist: (i) provide larger sets of training data or (ii) reduce the dimensionality by extracting pertinent features from the hyperspectral signals. Oftentimes, solution (i) is not practical since one must take into account that as the number of dimensions increase, the sample size of the training data needs to increase exponentially in order to have reliable multivariate statistics for estimating specific parameters. Thus, we must consider solution (ii), creating a need for feature extraction methods that can reduce the data space dimensions without losing the original information that allows for the separation of classes.

One approach to feature extraction is to consider the hyperspectral signal amplitude in a given spectral channel as the feature. This leads to sorting through the various spectral channels to determine which channels are the best features for discriminating between classes. Another approach is to utilize more advanced digital signal processing (DSP) methodologies to extract pertinent features from the hyperspectral signal. For example, the discrete wavelet transform (DWT) is often used in DSP for analyzing the scale-position information of a signal. The DWT of a given signal results in a set of wavelet coefficients associated with a range of scales and positions. Each coefficient is directly related to the amount of energy in the signal at a particular position and scale. By extracting features from the wavelet coefficients, we can quantify the small-scale (or fine detail) behavior as well as the large-scale (or gross) behavior of the hyperspectral signal.

Wavelets have been previously utilized in various areas of remote sensing. Wavelets have been used to reveal relationships between landscape features and plant diversity indices at different imagery spatial resolutions (Brosofske, Chen, Crow, & Saunder, 1999), identify forest canopy structure (Bradshaw & Spies, 1992), and to evaluate the influence of cover crop residue on hyperspectral reflectance of soybean and weeds (Huang, Bruce, Koger, & Shaw, 2001). However, the potential for wavelet analysis of hyperspectral reflectance data for discriminating weeds from crop is not well documented. Also, little is known of the influence tillage practices, such as conventional till and no-till systems, have on these discriminant capabilities. With this in mind, the objectives of this research are (i) to evaluate the utility of the DWT for extracting pertinent features from hyperspectral signals and (ii) to investigate the effect of the choice of mother wavelet on the features efficacy. Both objectives are conditional on hyperspectral data being used to detect pitted morningglory in soybean.

2. Materials and methods

2.1. Study area

A field experiment was established in 2001 at the Plant Science Research Center, Starkville, MS (33°28′N, 88°47′W). The soil type is a Marietta fine sandy loam (fine-loamy, mixed, thermic, siliceous Aquic Fluventic Eutrochrept) with a pH of 6.0 and 1.4% organic matter. The experiment was arranged in a randomized complete

block with each treatment replicated four times in 4.5 × 12 m plots. The treatments were presence or absence of pitted morningglory, a troublesome vining-type row crop weed species (Sanders et al., 2001), with half of the plots containing pitted morningglory intermixed with soybean and the other half containing weed-free soybean. Existing vegetation in all plots was desiccated with paraquat at 1.1 kg/ha on April 19, 2001. Two days later, the entire experimental area was disked twice, and the glyphosate resistant soybean cultivar 'Asgrow 4702 RR' was planted in 57-cm rows in all plots on May 21. Immediately after planting, volunteer weeds in all plots were desiccated with paraquat (1.1 kg/ha). Then a 3.0 × 3.0-m plastic tarp was placed in the center of designated plots that were to contain pitted morningglory. A preemergence application of metolachlor at 2.3 kg/ha plus imazaquin at 0.14 kg/ha was applied to control all weeds in the areas not covered by tarps, including the "weed free" plots. Paraquat (1.1 kg/ha) was also applied to kill emerged weeds and cover crop re-growth. Tarps were then removed so that pitted morningglory seed could be planted in nine 1.0-m^2 quadrates in the center of each pitted morningglory plot. Once emerged, pitted morningglory populations were thinned so that each 1.0-m^2 quadrate contained four plants. This density was maintained throughout the course of the experiments by hand pulling excess pitted morningglory plants and other weeds as needed. All weeds in plots not containing pitted morningglory ("weed-free") and weeds outside of the center 3.0 × 3.0-m sampling area in the plots containing pitted morningglory were controlled with glyphosate at 1.1 kg/ha.

2.2. Data acquisition

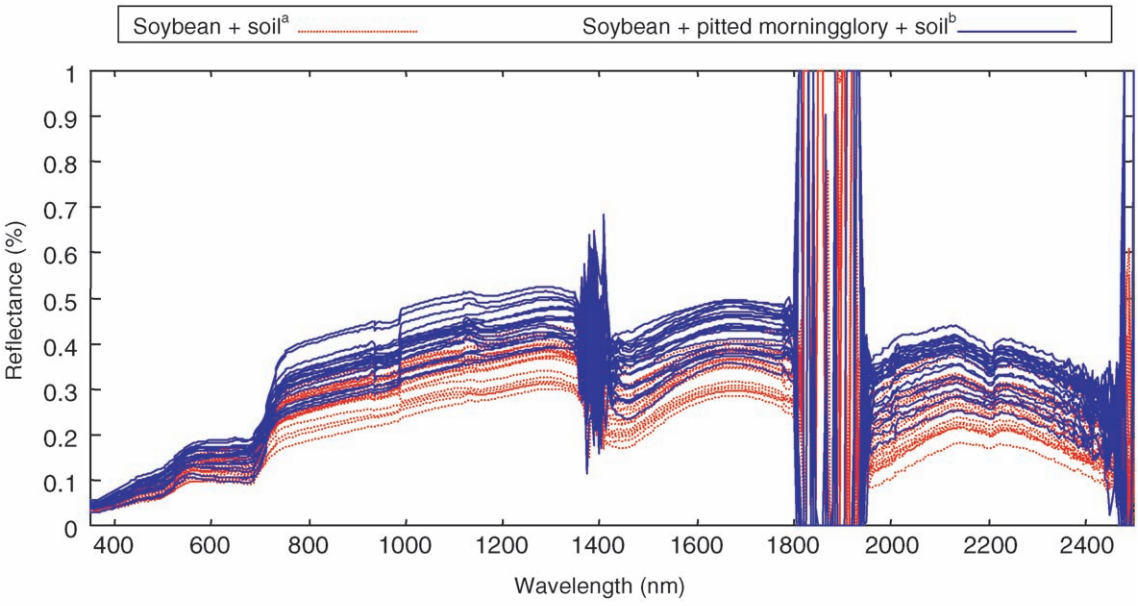
Beginning when pitted morningglory was in the cotyledon to two-leaf and soybean in the two-leaf to three-leaf growth stages (Fehr, Caviness, Burmood, & Pennington, 1971), hyperspectral reflectance measurements were collected using an ASD FieldSpec Pro FR portable spectroradiometer manufactured by Analytical Spectral Devices™. Reflectance is the ratio of energy reflected off the target (i.e. plant and background soil/residue) to energy incident on the target, which was measured using a $BaSO_4$ white reference. The hyperspectral reflectance measurements were collected between the spectral range of 350 and 2500 nm, with 3.0-nm spectral resolution at 700- and 30-nm resolution at 1400 and 2100 nm, respectively. This resulted in 2151 individual spectral bands for each hyperspectral reflectance curve, with a bandwidth of 1.4 nm between 350 and 1050 nm and 1.0 nm between 1000 to 2500 nm. Eight hyperspectral reflectance measurements were collected for soybean plus background soil in each weed-free plot. Eight measurements for pitted morningglory intermixed with soybean plus background soil were also collected from each plot containing pitted morningglory. Hyperspectral measurements were collected using a 23° field-of-view (FOV) optic. The sensor was held 122 cm directly above the object of interest (soybean or pitted morningglory intermixed with soybean). This resulted in approximately 0.25-m spatial resolution for each hyperspectral measurement. This was done to ensure that background reflectance of soil was included in each soybean or soybean intermixed with pitted morningglory measurement. Hyperspectral reflectance data were collected once per week until pitted morningglory plants were beyond the six-leaf to nine-leaf growth stage. Plant height, growth stage (Fehr et al., 1971), and ground cover estimates for soybean and pitted morningglory at time of hyperspectral data acquisition are listed in Table 1. These measurements resulted in a database of 32 hyperspectral signatures for each class (soybean + soil and soybean + pitted morningglory + soil) for each growth stage. As an example, Fig. 1 shows the hyperspectral signals measured during the two-leaf to four-leaf growth stage.

In situ hyperspectral reflectance measurements were also collected for canopies of soybean and pitted morningglory plants at the two-leaf to four-leaf pitted morningglory growth stage in the conventional tillage plots using an 8° FOV optic held 6 cm directly above each object of interest (soybean or pitted morningglory leaf or bare soil). Four reflectance measurements were collected from the uppermost leaflet of the youngest, fully expanded, trifoliolate soybean leaf from each plot. Four measurements were also taken from the youngest, fully expanded pitted morningglory leaf in each soybean plus pitted morningglory plot. Four bare soil reflectance measurements (two from weed-free soybean and two from soybean plus pitted morningglory plots) were also collected. In total, 16 reflectance

Table 1
Pitted morningglory and soybean growth stage, height, and ground cover estimates[a]

Soybean plus pitted morningglory plots						Weed-free soybean plots		
Pitted morningglory			Soybean			Soybean		
Growth stage (# leaves)	Height (cm)	Ground cover (%)	Growth stage[b] (# leaves)	Height (cm)	Ground cover (%)	Growth stage[b] (# leaves)	Height (cm)	Ground cover (%)
Coty.-2	5–8	25	V2–V3	5–15	28	V2–V3	7–14	32
2–4	5–10	30	V3–V4	9–16	30	V3–V4	12–16	34
4–6	8–12	38	V4–V6	22–29	36	V4–V6	23–27	39
6–9	8–14	45	V5–V7	23–35	44	V5–V7	25–35	46

[a] Coty., cotyledon.
[b] Soybean were in (V) vegetative stage of development.



[a] Soybean + soil mixed-pixel signatures contain 34% soybean and 66% soil reflectance.

[b] Soybean + pitted morningglory + soil contain 30% soybean, 30% pitted morningglory, and 40% soil reflectance.

Fig. 1. Soybean + soil and soybean + pitted morningglory + soil mixed-pixel reflectance signals containing unequal proportions of vegetation and soil reflectance (naturally mixed).

measurements for soybean leaves, 16 for pitted morningglory leaves, and 16 for bare soil were collected for each plot. Because of these measurements, the authors were able to create a database of synthetically mixed reflectance signals, where there were 120 hyperspectral signatures in each class (soybean + soil and soybean + pitted morningglory + soil).

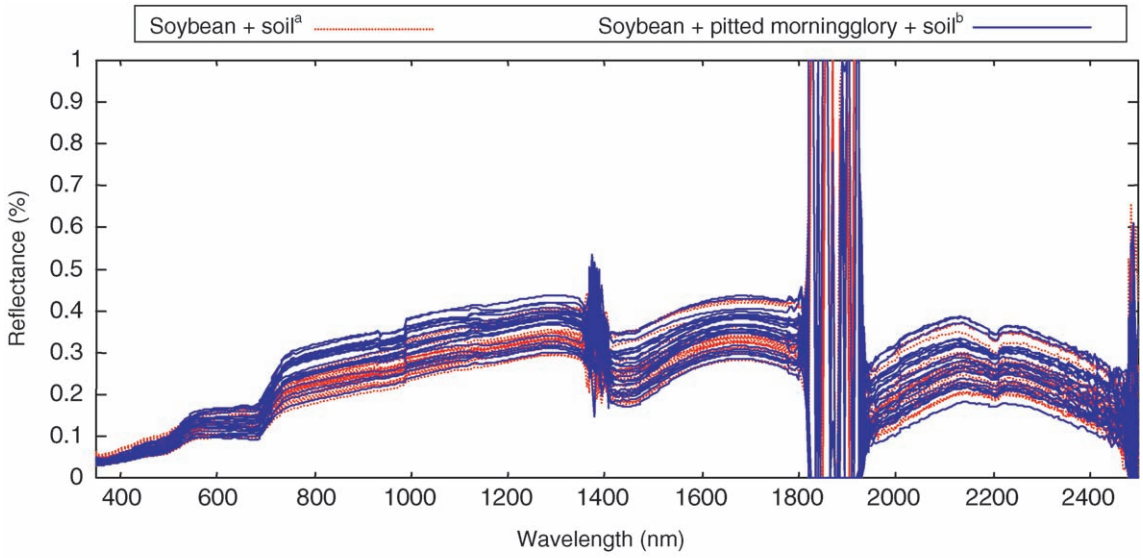
2.3. Development of mixed reflectance signals

When the hyperspectral data were collected for the mixed signatures, using the 23° FOV optic, the data could be divided into two main classes: (a) soybean + pitted morningglory + soil and (b) soybean + soil. Note that these are commonly referred to as mixed pixels, since they correspond to pixels from a hyperspectral image cube where more than one endmember was present in the field-of-view. Regardless of growth stage, the mixed pixels for these two classes consisted of differing percentages of vegetative groundcover. For example, for the two-leaf to four-leaf pitted morningglory growth stage in Table 1, the percent vegetative groundcover is listed as (a) 30% soybean and 30% pitted morningglory for the soybean + pitted morningglory + soil class and (b) 34% soybean for the soybean + soil class. Thus, the percentage of total vegetative groundcover varied between 60% and 34% for the two classes. An automated classification system was designed and tested for the purpose of discriminating between the two classes: (a) soybean + pitted morningglory + soil and (b) soybean + soil. However, there was concern whether the system was detecting the actual presence of pitted morningglory or simply detecting a higher percentage of vegetative groundcover. Therefore, the unmixed hyperspectral signals (data collected with the 8° FOV optic) were used to investigate this issue.

The 8° FOV data represented pure pixels of each of the groundcover components: (a) soybean only, (b) pitted morningglory only, and (c) soil only. The pure pixel hyperspectral signals were used to synthesize mixed-pixel data. Several linearly mixed datasets were constructed, and the mixing abundances are listed in Table 2. In each mixing case, the pure hyperspectral signals used for the two classes were mutually exclusive, so that each pure signal was used

Table 2
Proportions (%) of soybean, pitted morningglory, and soil reflectance for synthetic mixed-pixel hyperspectral signals containing unequal and equal amounts of vegetation reflectance

Mixed-pixel reflectance component	Reflectance curves containing unequal proportions of vegetation and soil		Reflectance curves containing equal proportions of vegetation and soil	
	Soybean + soil	Soybean + pitted morningglory + soil	Soybean + soil	Soybean + pitted morningglory + soil
Soybean	34	30	34	17
Pitted morningglory	0	30	0	17
Soil	66	40	66	66



[a] Soybean + soil mixed-pixel reflectance curves contain 30% soybean and 70% soil reflectance.

[b] Soybean + pitted morningglory + soil contain 15% soybean, 15% pitted morningglory, and 70% soil reflectance.

Fig. 2. Soybean + soil and soybean + pitted morningglory + soil mixed-pixel reflectance signals containing equal proportions of vegetation and soil reflectance (synthetically mixed).

only once in developing synthetic mixed-pixel reflectance curves. Fig. 2 shows an example case of the synthetically mixed hyperspectral signals. Note that for each case listed in Table 2, the total percentage of vegetation was held constant. For example, consider the second dataset where the two classes of mixed signals had mixing proportions of 70/30% and 70/10/20%. In each class, the proportion of vegetation was 30%. Hence, if the automated classification system detected any difference between the two classes, the differences resulted from the variation in vegetation type, not variation in percentage of vegetation groundcover.

2.4. Hyperspectral signal analysis

Several automated classification systems were implemented and tested for the purpose of automatically discriminating between the hyperspectral curves in the soybean + soil class and the soybean + pitted morningglory + soil class. The difference in the various classification systems was the method of feature extraction. Three main feature extraction methods were used to reduce the dimensionality of the hyperspectral signals: (i) selecting a reduced set of spectral bands (without using any transformation), (ii) selecting a reduced set of coefficients after applying principal component analysis (PCA), and (iii) selecting a reduced set of coefficients after applying a discrete wavelet transform (DWT) to the signal.

For the PCA approach, case (ii), the principal component transform was applied to each hyperspectral signal, and the transform was based on the total covariance matrix. That is, it is the unsupervised PCA method. Oftentimes, simply selecting the first few components is not an effective method of dimensionality reduction when the goal is target detection. Therefore, the components were sorted, and an optimum subset of components was selected. Selection was based on receiver operating characteristics (ROC) curves as is described in more detail in the following paragraphs. For case (i), the spectral bands of the hyperspectral signals are used directly with no transformations. Again, ROC curves were used to sort through the spectral bands and select an optimum subset. For case (iii), the DWT was applied to each hyperspectral signal, and the resulting coefficients were sorted using ROC curves to obtain an optimum subset.

The DWT was implemented using a dyadic filter tree, as shown in Fig. 3. The input to the filterbank, $f(\lambda)$, is the hyperspectral signal, and the signal is passed through a series of low-pass filters (LPF) and high-pass filters (HPF). After each filter, the signal is downsampled by a factor of two ($\downarrow 2$). The result of the filter bank, $\vec{W}$, is a set of wavelet coefficients. The filters are dependent on the selection of the

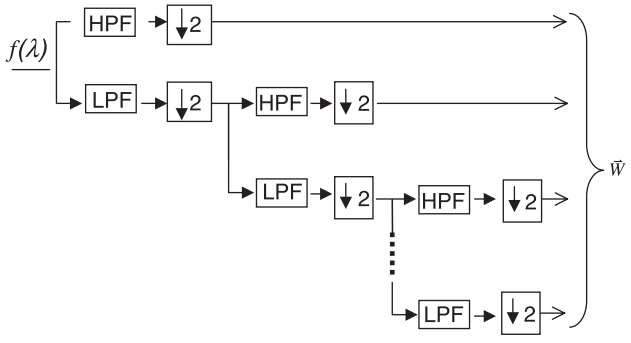


Fig. 3. Dyadic filter tree implementation of the discrete wavelet transform.

mother wavelet function, $\psi(\lambda)$, used in the transform. As a result, the wavelet coefficients vary according to the selection of the mother wavelet. The authors investigated a variety of mother wavelets in order to determine their effect on the wavelet coefficients' class discriminating ability.

In order to be used in the DWT filter bank, a mother wavelet function must satisfy a mathematical criteria known as the mutiresolution analysis (MRA) property (Mallat, 1989). Thirty-six different mother wavelets were investigated for identifying pertinent features capable of being used for discriminating the two classes. These wavelet functions can be organized into five classes: Haar, Daubechies, Symlet, Coiflet, and Biorthogonal (Grabs, 1995). Each class is different with respect to symmetry, regularity, support, and orthogonality. The Haar wavelet is arguably the simplest type of mother wavelet. It is asymmetric, has compact support, and has $1/\lambda$ decay in frequency. It is important to note that the detail coefficients resulting from the use of the Haar mother wavelet are equivalent to a first-order approximation of the first derivative of the hyperspectral signal. This is interesting considering the common use of derivative analysis in spectroscopy, such as the Savitzky–Golay method (Bruce & Li, 2001). However, the Haar DWT has the advantage that scale is a factor in the transform. That is, the coefficients that result from the Haar DWT correspond to systematically applying various width first-order derivatives to the hyperspectral signal in such a way that there exists no redundancy in the output. Furthermore, other mother wavelets were also investigated to determine if the choice of mother wavelet greatly affects the results. For the Daubechies-n wavelet, n specifies the order of the mother wavelet, corresponds to the regularity of the mother wavelet, and is related to the number of coefficients necessary to represent the associated low-pass and high-pass filters in the dyadic filter tree implementation. The Daubechies wavelet has compact support, and the support length is $2n-1$. Symlet wavelets are a variation on the Daubechies wavelets in order to obtain a more symmetric mother wavelet function. For the Coiflet-n wavelet, n specifies the order of the mother wavelet. The Coiflet wavelet has $2n$ moments equal to zero. This is the highest number of vanishing moments for an n-order wavelet. The Coiflet wavelet also has compact support, and the support length is $6n-1$. The Haar, Daubechies, Symlet, and Coiflet wavelets all result in an orthogonal decomposition of the hyperspectral signal. Finally, the biorthogonal wavelets have compact support but do not provide an orthogonal decomposition of the hyperspectral signal. The low-pass and high-pass filters associated with the biorthogonal wavelets are symmetric. Differences in support, regularity, and symmetry of the mother wavelets affect the resulting wavelet coefficients and, hence, the resulting features used for discriminating between ground cover classes. One goal of this experiment is to investigate whether the selection of mother wavelet greatly affects the discriminating capability and to investigate which of these mother wavelets perform best.

The automated classification systems were designed and trained using receiver operating characteristics (ROC) curves and Fisher's linear discriminant analysis (LDA). For each system, $\vec{H}$ represents the set of potential hyperspectral features for a given system (Fig. 4). For case (i), $\vec{H}=f(\lambda)$; that is, the set of potential features were the amplitudes of the hyperspectral curve in each of the 2151 spectral bands. For case (ii), $\vec{H}$ is the set of coefficients resulting from applying the principal component transform to the hyperspectral signal. For case (iii), $\vec{H}=\vec{W}$; that is, the set of potential features were the wavelet coefficients resulting from the DWT.

For each case, the potential features were ranked according to their area, A_Z, under the ROC curve (Hanley & McNeil, 1982). For a given feature, A_Z ranges between 0.5 and 1.0. A value of 0.5 indicates the feature has no ability to separate the classes (soybean + soil and soybean + pitted morningglory + soil), and a value of 1.0 indicates the feature has perfect ability to separate the classes. For case (i), A_Z was computed for each of the 2151 spectral bands. This resulted in a vector, $\vec{A_Z}$, containing 2151 areas. For cases (ii) and (iii), A_Z was computed for each of the PCA coefficients and DWT coefficients, respectively. In each case, the potential features were then sorted, such that the optimum combination of 10 spectral bands was selected to form a reduced feature set, $\vec{F}$. An index vector, $\vec{I}$, was also created to keep track of which features were selected for the reduced feature set. The index vector was used when the system was later evaluated on test data.

The reduced feature set was used as an input to an automated statistical classifier. The first stage of the classifier was Fisher's linear discriminant analysis (LDA). LDA produces a new feature vector, $\vec{G}$, via an optimal linear combination of the elements in $\vec{F}$. That is, $\vec{G}=\mathbf{W}\vec{F}$, where $\mathbf{W}$ is a linear combination weight matrix. For an M-class

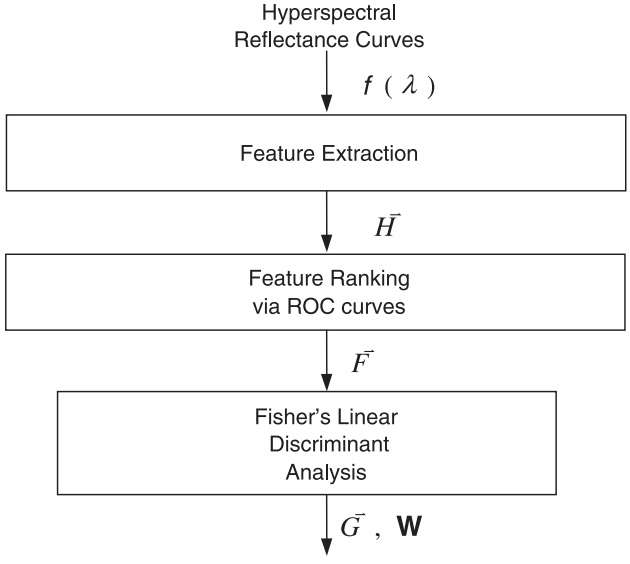


Fig. 4. Block diagram of training phase of automated classification system.

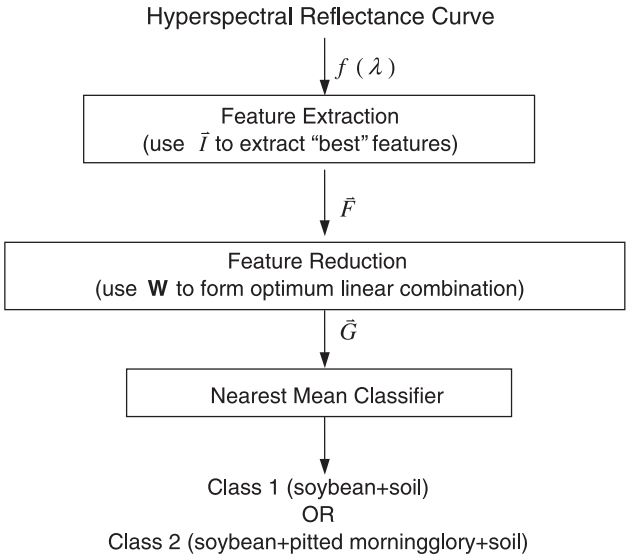


Fig. 5. Block diagram of automated classification system.

problem, **W** is a $((M-1)\times J)$ weight matrix, $\overrightarrow{F}$ is a $(J\times 1)$ vector, and $\overrightarrow{G}$ is a $((M-1)\times 1)$ vector. The optimal weights, **W**, are calculated using the training data (Webb, 1999). Note that for this experiment, the number of classes was $M=2$, since the classes were soybean + soil and soybean + pitted morningglory + soil. Thus, the final feature vector, $\overrightarrow{G}$, had a scalar matrix dimensionality of 1×1. That is, the 2151 dimensionality of the hyperspectral signal was reduced down to a one-dimensional value. The weight matrix, **W**, was used later when the system was evaluated on test data.

Fig. 5 shows the stages of the final automated classification system. The system was tested with Fisher's linear discriminant analysis using the unbiased leave-one-out, or cross-validation, method (Johnson & Wichern, 1992). The classification accuracy, A_{cc}, was computed to evaluate the system performance. A 95% confidence interval, I_{95}, was also calculated to monitor the reliability of classification results and account for the limited size of the training and testing data (Fisher & Belle, 1993). Synthetic mixed-pixel development, PCA, DWT, ROC, and automated classification system analysis procedures were all conducted with Matlab[4] software developed by the authors.

3. Results and discussion

3.1. Varying feature extraction method

For each pitted morningglory growth stage, the accuracies of the automated classification systems for the cases when the optimum subsets of (i) spectral bands, (ii) principal components, and (iii) Haar wavelet detail coefficients are used as features are listed in Table 3. The classification results are reported in terms of specificity (class not containing pitted morningglory reflectance), sensitivity (class containing pitted morningglory reflectance), and overall classification accuracy, where accuracy is defined as percent correctly classified.

The PCA approach resulted in the lowest classification accuracies at three of the four pitted morningglory growth stages (Table 3). The system's overall classification accuracy ranged from 72% to 83%. While PCA is commonly used in hyperspectral data analysis, it is not a particularly useful feature extraction method when the difference in reflectance for different vegetation types does not differ substantially, especially when the within-class variances dominate the between-class variances. Typically, the untransformed spectral bands outperfromed the PCA components. When using the spectral bands as features, the system's overall classification accuracy ranged from 80% to 87%. However, the DWT approach performed the best at every growth stage, with the system's overall classification accuracy ranging from 90% to 100%. Recall that the Haar mother wavelet is arguably the simplest of all mother wavelets to implement; that is why the Haar mother wavelet was used in the first stage of analysis. Based on these results, the wavelet-based approach shows great promise for discriminating weed-free soybean and soybean intermixed with pitted morningglory, even at early growth stages when pitted morningglory is most easily controlled with herbicide.

The classification system's specificity is generally higher than the sensitivity. Typically, this is not a desired result, since the specificity indicates the system's ability to detect the absence of the target (pitted morningglory) and the sensitivity indicates the system's ability to detect the target. A low sensitivity indicates that the system will have a high number of target misses. In this application, the end user would more likely prefer a false alarm rather than a target miss, since false alarms would cause the end user to use excessive amounts of herbicide. However, target misses would cause areas that contain weeds to not be treated. This is a trend worth noting, even though the system results in specificities and sensitivities that are both quite high.

When spectral band features were used, each spectral band was evaluated using the area under the ROC curve, and an optimum subset of 10 spectral bands was selected as a feature vector (Table 4). Fig. 6 shows a plot of the ROC areas (A_Z's) for the four-leaf to six-leaf pitted morningglory growth stage. The figure also shows a randomly selected soil + soybean hyperspectral signal, for comparison purposes. Notice that the ROC areas are largest in the NIR region. However, simply selecting a subset of these bands would be suboptimum because of the redundancy in these closely neighboring bands. When the optimum subset of 10 bands was actually derived, the spectral bands came from across the spectrum (Table 4).

[4] Matlab, The Mathworks, 3 Apple Drive, Natick, MA 01760-2098, USA.

Table 3
Classification accuracy (%) of automated systems for the three feature extraction methods at each growth stage

Growth stage[a]		Feature extraction method	Classification accuracy		Overall classification accuracy
Pitted morning glory	Soybean		Specificity, soybean + soil	Sensitivity, soybean + pitted MG[b] + soil	
Coty.[b]-2	V2–V3	Principal components	72	73	73
		spectral bands	85	75	80
		Haar wavelet details	95	85	90
2–4	V3–V4	Principal components	85	80	83
		spectral bands	88	75	81
		Haar wavelet details	93	97	95
4–6	V4–V6	Principal components	73	83	78
		spectral bands	87	87	87
		Haar wavelet details	100	100	100
6–9	V5–V7	Principal components	75	69	72
		spectral bands	88	84	86
		Haar wavelet details	92	90	91

[a] Soybean were in (V) vegetative stage of development.
[b] Coty., cotyledon; MG, morningglory.

Likewise, when the wavelet features were used, each detail coefficient was evaluated using the area under the ROC curve, and an optimum subset of 10 coefficients was selected as a feature vector. Simply plotting the area under the ROC curve for the wavelet detail coefficients was not as informative since the wavelet coefficients correspond not only to spectral band position but also to scale. In order to obtain a visual representation of "where" in the spectral curve the best wavelet coefficients reside, the 10 coefficients were used to reconstruct the reflectance curve. That is, an inverse DWT was computed using only the 10 best detail coefficients, the ones selected to be in the best subset and make up the feature vector $\vec{F}$ (as shown in Fig. 4). Fig. 7 shows the reconstruction of a randomly selected soil + soybean hyperspectral signal, as well as the original signal for comparison purposes. Notice that the best wavelet coefficients primarily correspond to small details near the red-edge and larger scale details in the infrared region of the spectrum. Since the Haar mother wavelet was used in this portion of the experiment, the reconstruction has a distinct square wave behavior. Recall that wavelet coefficients are a function of scale and position (fine detail versus global behavior at various locations in the hyperspectral signal). Since a Haar wavelet detail coefficient corresponds to a first-order approximation of the first derivative of the analyzed signal, the best detail coefficients directly correspond to the fine detail and large-scale slopes of the reflectance curve at various locations. Several of the best detail coefficients are located at $\approx$ 700 to 750 nm. These are very small-scale coefficients and are related to the slope of the red-edge region. Another one of the best detail coefficients is located at $\approx$ 1050 nm. It is a larger scale coefficient and is related to the slope of the reflectance curve in the infrared region. The fact that some of the best detail coefficients stem from larger scales in the DWT indicates that a more global view of the signal can be more useful than simply observing the reflectance at finely resolved spectral bands. This might be obvious when the reflectance signals of the tested classes are dramatically different, such as soil versus vegetation. However, this is not intuitive when the reflectance signals of the classes (soybean + soil and soybean + pitted morningglory + soil) are subtly different, as with these data.

Table 4
Spectral bands chosen with ROC curve analysis at each pitted morning glory growth stage[a]

Pitted morning glory growth stage (# leaves)	Spectral band[b] (nm)
Coty.-2	521, 601, 673, 798, 906, 914, 1161, 1523, 1541, 2107
2–4	572, 606, 782, 791, 816, 907, 1173, 1498, 1512, 2087
4–6	583, 627, 629, 801, 843, 867, 952, 1587, 2197, 2201
6–9	621, 689, 787, 801, 822, 914, 977, 1522, 1531, 2083

[a] Coty., cotyledon.
[b] Each spectral band 1.4 nm in width.

3.2. Varying mother wavelet

Table 5 shows the classification results when varying the mother wavelet. These results demonstrate that the choice of mother wavelet can greatly affect the efficacy of the wavelet-based features. If we set a threshold of 90% for the system accuracy and demand the system perform above that threshold regardless of plant growth stage, then only a few of the mother wavelets would qualify. These are the Haar, db5, db10, bior2.2, bior2.4, bior2.6, bior2.8, bior6.8, sym2, and sym7. Note however, that the very simple Haar mother wavelet is in this select group of superior performing mother wavelets.

3.3. Effect of plant growth stage

Next, consider in detail the effect of pitted morningglory growth stage on the automated system's classification accuracy. The overall classification accuracy for each of the 36

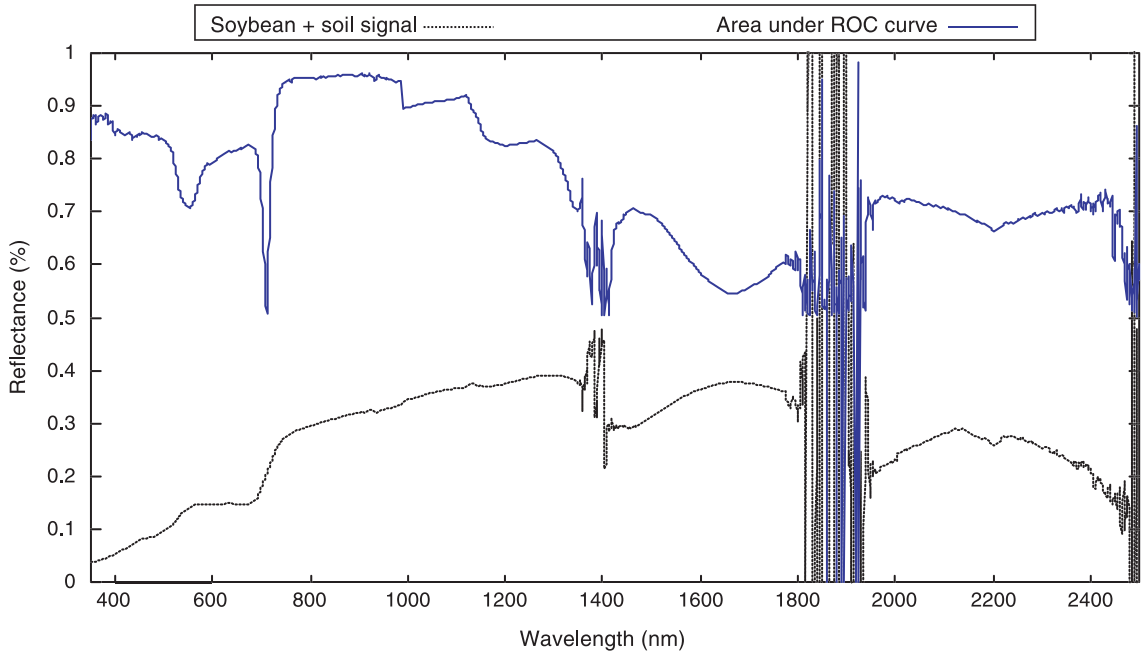


Fig. 6. Randomly selected soybean + soil hyperspectral signal and area under ROC curve at the four-leaf to six-leaf pitted morningglory growth stage.

mother wavelets at each pitted morningglory growth stage is shown in Fig. 8. In general, the best classification accuracies occurred when pitted morningglory was in the four-leaf to six-leaf growth stage, with 98% to 100% classification accuracy for many of the tested mother wavelets. While classification accuracy was generally higher at the four-leaf to six-leaf growth stage, in many cases classification accuracy at the two-leaf to four-leaf stage is almost as good if not better. In terms of practical implementation, it is important to classify weeds when the weeds are very small, as this is the time frame in which they are most easily controlled with herbicide. Also, one should consider the length of time over which the system provides acceptable accuracies. Depending on the source of the hyperspectral imagery, the end user may or may not be able to specify the exact date of the data collection or know the exact weed growth stage at the time of data collection. In this case, we would want to design the classification system to be as robust as possible. For example, consider the case where the Daubechies-5 (db5) mother wavelet was used. Classification accuracy was 95%, 94%, 92%, and 92% at the cotyledon-two leaf, two–four leaf, four–six leaf, and six–nine leaf growth stages, respectively. Likewise, when the biorthogonal-6.8 (bior6.8) mother wavelet was used, the accuracies

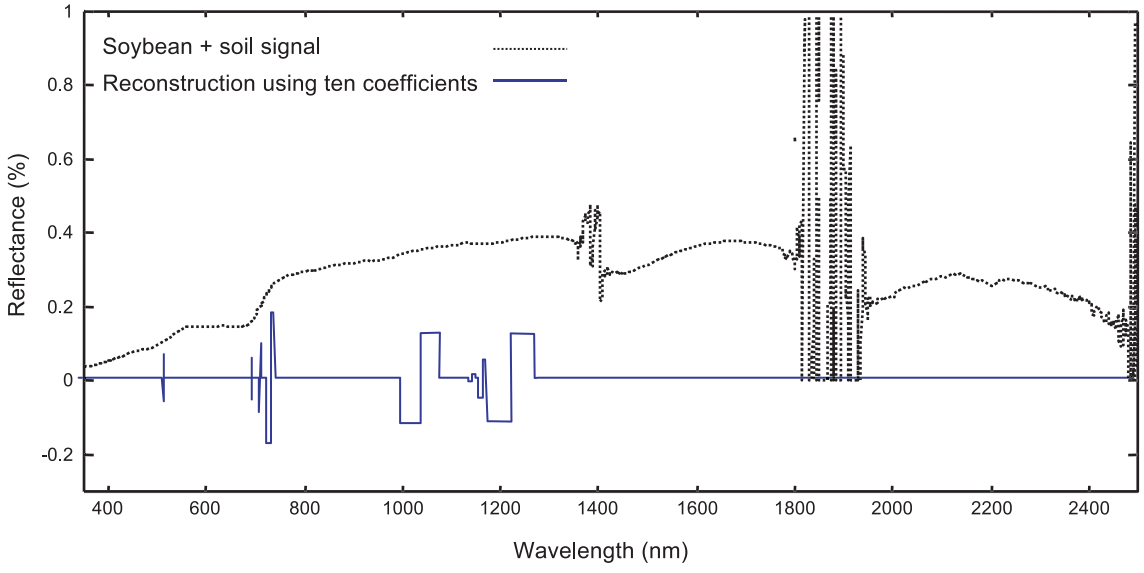


Fig. 7. Randomly selected soybean + soil hyperspectral signal and reconstruction of the signal using only best 10 wavelet detail coefficients.

Table 5
Classification accuracy of the soybean + soil and soybean + pitted morningglory + soil classes when the hyperspectral signals are synthetically mixed with varying mixing proportions

Mixing proportions (%) soybean + soil class		Mixing proportions (%) soybean + pitted morningglory + soil class			Overall classification accuracy (%)[a]
Soil	Soybean[b]	Soil	Soybean[b]	Pitted morning-glory[c]	
90	10	90	5	5	73
80	40	80	5	15	88
80	40	80	10	10	83
80	40	80	15	5	80
70	30	70	10	20	97
70	30	70	15	15	94
70	30	70	20	10	91
60	40	60	10	30	100
60	40	60	20	20	100
60	40	60	30	10	97
50	50	50	10	40	100
50	50	50	20	30	100
50	50	50	30	20	97
50	50	50	40	10	97
40	60	40	10	50	100
40	60	40	20	40	100
40	60	40	30	30	100
40	60	40	40	20	100
40	60	40	50	10	97

[a] Haar DWT feature extraction.
[b] V3–V4 growth stage.
[c] Two-leaf to four-leaf growth stage.

were 98%, 100%, 100%, and 97%. In these two cases, the system would perform well regardless of the pitted morningglory's growth stage

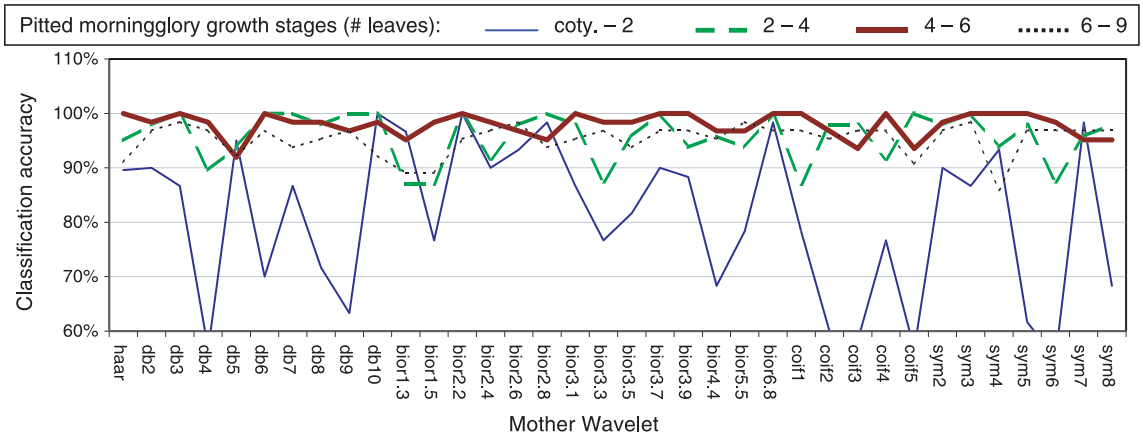
3.4. Synthetically mixed reflectance curves

Finally, in order to account for varying mixtures of soybean, pitted morningglory, and soil, let us consider the results of the system when analyzing the synthetically mixed hyperspectral signals. Table 2 shows the 22 different mixing proportion cases that were investigated. In each case, the percentage of vegetation was the same for the two classes (soybean + soil and soybean + pitted morning glory + soil). These datasets were tested using the DWT approach to feature selection and the Haar mother wavelet. When the two classes had 90% soil, the system's overall accuracy was only 73%, but this is to be expected considering the small amount of vegetation in the scene. When the amount of soil was decreased from 90% to 80%, the system's performance significantly increased, resulting in accuracies from 80% to 88%. And again when the proportion of soil was decreased from 80% to 70%, the systems performance significantly increased, resulting in accuracies from 91% to 97%. Once the amount of soil was $\leq 60\%$, the system's accuracy was between 97% and 100%. Notice that for a given proportion of soil in the scene, as the proportions of soybean and pitted morningglory was varied the accuracies changed. In each case, as the amount of pitted morningglory decreased, the system's accuracy decreased. This is intuitive since the scene containing pitted morningglory is becoming more similar to the scene not containing pitted morningglory.

Note that in general in our experiments, the synthetically mixed curves resulted in slightly higher classification accuracies (Table 2) than with the naturally mixed signatures (Table 3). This could be explained by the fact that the naturally mixed curves probably contain some nonlinear mixing effects, which could cause more natural intraclass variance.

4. Conclusions

Wavelets show promise for being able to select pertinent features (detail coefficients) that can be used in discriminating weed-free crop from crop intermixed with pitted morningglory.



[a] Abbreviations: db., daubechies; bior., biorthogonal; coif., coiflet; sym., symelt.

Fig. 8. Overall classification accuracy of automated systems for each mother wavelet and growth stage.

When applying the Haar DWT to naturally mixed hyperspectral signatures, the overall classification accuracy was 90% to 100%, depending on pitted morningglory growth stage. These results were significantly higher than when spectral bands or principal components were used as classification variables. The ability to discriminate weed-free soybean from soybean intermixed with pitted morningglory fluctuated according to which mother wavelet was used. Several mother wavelets were capable of discriminating weed-free soybean from soybean intermixed with two-leaf to four-leaf pitted morningglory 100% of the time. In fact, when the biorthogonal-6.8 (bior6.8) mother wavelet was used, the system could discriminate weed-free soybean from soybean intermixed with pitted morningglory $\geq 97\%$ of the time regardless of the pitted morningglory's growth stage. However, a DWT system that utilizes a Haar mother wavelet can be very efficiently implemented with lower computational costs than a DWT system that utilizes the biorthogonal-6.8 mother wavelet. The increased accuracy may not warrant the added computational expense of the biorthogonal-6.8 mother.

In order to investigate the effect of vegetation type versus magnitude of vegetation cover, several databases of synthetically mixed hyperspectral signatures were constructed. The Haar DWT-based system also performed well on the synthetically mixed signals. As the percentage of vegetation ground cover increased, the system's performance increased. In order to attain a classification accuracy of $\geq 80\%$, at least 20% of the ground cover needed to be vegetation. And in order to attain a classification accuracy of $\geq 90\%$, at least 30% of the ground cover needed to be vegetation. As the percentage of pitted morningglory increased, relative to the percentage of soybean, the system's classification accuracy increased. These results are quite promising when considering the kind of treatment that could be applied to the crop when the pitted morningglory's growth stage (two-leaf to four-leaves) is still controllable with herbicide.

Acknowledgements

The authors would like to thank Louis Wasson, Research Assistant I, Engineering Research Center, and appropriate undergraduate and graduate research assistants, Mississippi State University for assistance in data collection and in operation of the spectroradiometer.

References

Banks, P. A., Santlemann, P. W., & Tucker, B. B. (1976). Influence of long-term soil fertility treatments on weed species in winter wheat. *Agronomy Journal*, *68*, 825–827.

Barrentine, W. L. (1974). Common cocklebur competition in soybeans. *Weed Science*, *22*, 600–603.

Blackburn, G. A. (1998). Quantifying chlorophylls and carotenoids at leaf and canopy scales: an evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, *66*, 273–285.

Bradshaw, G. A., & Spies, T. A. (1992). Characterizing canopy gap structure in forests using wavelet analysis. *Journal of Ecology*, *80*, 205–215.

Brosofske, K. D., Chen, J., Crow, T. A., & Saunders, S. C. (1999). Vegetation responses to landscape structure at multiple scales across a northern Wisconsin, USA, pine barrens landscape. *Plant Ecology*, *143*, 203–218.

Bruce, L., & Li, J. (2001). Wavelet for computationally efficient hyperspectral derivative analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 1540–1546.

Christensen, S., Walter, A. M., & Heisel, T. (1999). The patch treatment of weeds in cereals. *Proceedings of Brighton Crop Protection Conference of Weeds* (pp. 591–600). Farnham, UK: British Crop Protection Council.

Cousens, R. D., & Woolcock, J. L. (1987). Spatial dynamics of weeds: an overview. *Proceedings of Brighton Crop Protection Conference of Weeds* (pp. 613–618). Farham, UK: British Crop Protection Council.

Elvidge, C. D., & Mouat, D. A. (1985). Influence of rock–soil spectral variation on the assessment of green biomass. *Remote Sensing of Environment*, *17*, 265–269.

Everitt, J. H., Escobar, D. E., Alaniz, M. A., Davis, M. R., & Richardson, J. V. (1996). Using spatial information technologies to map Chinese tamarask (*Tamarix chinensis*) infestations. *Weed Science*, *44*, 194–201.

Fehr, W. R., Caviness, C. E., Burmood, D. T., & Pennington, J. S. (1971). Stage of development descriptions for soybeans, *Glycine max* (L.) Merr. *Crop Science*, *11*, 929–931.

Felton, W. L., Doss, A. F., Nash, P. G., & McCloy, K. R. (1991). To selectively spot spray weeds. *American Society of Agricultural Engineers Symposium*, *11–91*, 427–432.

Fisher, L. D., & Belle, G. V. (1993). *Biostatistics: a methodology for the health sciences* (p. 857). Wiley-Interscience, Wiley, New York.

Grabs, A. (1995). An introduction to wavelets. *IEEE Computational Science & Engineering*, *2*, 50–61.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Diagnostic Radiology*, *143*, 29–36.

Huang, Y., Bruce, L. M., Koger, C. H., & Shaw, D. R. (2001). Analysis of the effects of cover crop residue on hyperspectral reflectance discrimination of soybean and weds via Haar transform. *Proceedings of IGARSS* (pp. 1276–1278). New York, NY: IEEE.

Jiminez, L., & Landgrebe, D. (1998). Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, *28*, 39–54.

Jiminez, L., & Landgrebe, D. (1999). Hyperspectral data analysis and supervised feature reduction via project pursuit. *IEEE Transactions on Geoscience and Remote Sensing*, *37*, 2653–2667.

Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis* (pp. 522–525). New York: Prentice-Hall.

Lass, L. W., Carson, H. W., & Callihan, R. H. (1996). Detection of yellow starthistle (*Centaurea solstitialis*) and common St. Johnswort (*Hypericum perforatum*) with multispectral digital imagery. *Weed Technology*, *10*, 466–474.

Mallat, S. (1989). A theory for multi-resolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 674–693.

Medlin, C. R., Shaw, D. R., Cox, M. S., Gerard, P. D., Abshire, M. J., & Wardlaw, M. C. (2001). Using soil parameters to predict weed infestations in soybean. *Weed Science*, *49*, 367–374.

Medlin, C. R., Shaw, D. R., Gerard, P. D., & Lamastus, F. E. (2000). Using remote sensing to detect weed infestations in *Glycine max*. *Weed Science*, *48*, 393–398.

Richardson, A. J., Menges, R. M., & Nixon, P. R. (1985). Distinguishing weed from crop plants using video remote sensing. *Photogrammetric Engineering and Remote Sensing*, *51*, 1785–1790.

Sanders, J. C., Monks, C. D., Patterson, M. G., Delaney, D. P., Moore, D. P., & Wells, L. W. (2001). Effectiveness of ammonium thiosulfate to en-

hance weed control and reduce cotton (*Gossypium hirsutum*) injury. *Weed Technology*, *15*, 236–241.

Shibayama, M., & Akiyama, T. (1991). Estimating grain yield of maturing rice canopies using high spectral resolution reflectance measurements. *Remote Sensing of Environment*, *36*, 45–53.

Swanton, C. J., & Weise, S. F. (1991). Integrated weed management: the rationale and approach. *Weed Technology*, *5*, 657–663.

Thenkabail, P. S., Smith, R. B., & Pauw, E. D. (2000). Hyperspectral vegetation indices and their relationship with agricultural crop characteristics. *Remote Sensing of Environment*, *71*, 158–182.

Thompson, J. F., Stafford, J. V., & Miller, P. C. H. (1991). Potential for automatic weed detection and selective herbicide application. *Crop Protection*, *10*, 254–259.

Thornton, P. K., Fawcett, R. H., Dent, J. B., & Perkins, T. J. (1990). Spatial weed distribution and economic thresholds for weed control. *Crop Protection*, *9*, 337–342.

Weaver, S. E., & Hamill, A. S. (1985). Effects of soil pH on competitive ability and leaf nutrient content of corn (*Zea mays* L.) and three weed species. *Weed Science*, *33*, 447–451.

Webb, A. (1999). *Statistical pattern recognition* (pp. 30–36). New York: Arnold and Oxford Univ. Press.