

# HYPERSPECTRAL IMAGE DATA MINING FOR BAND SELECTION IN AGRICULTURAL APPLICATIONS

S. G. Bajwa, P. Bajcsy, P. Groves, L. F. Tian

**ABSTRACT.** *Hyperspectral remote sensing produces large volumes of data, quite often requiring hundreds of megabytes to gigabytes of memory storage for a small geographical area for one-time data collection. Although the high spectral resolution of hyperspectral data is quite useful for capturing and discriminating subtle differences in geospatial characteristics of the target, it contains redundant information at the band level. The objective of this study was to identify those bands that contain the most information needed for characterizing a specific geospatial feature with minimal redundancy. Band selection is performed with both unsupervised and supervised approaches. Five methods (three unsupervised and two supervised) are proposed and compared to identify hyperspectral image bands to characterize soil electrical conductivity and canopy coverage in agricultural fields. The unsupervised approach includes information entropy measure and first and second derivatives along the spectral axis. The supervised approach selects hyperspectral bands based on supplemental ground truth data using principal component analysis (PCA) and artificial neural network (ANN) based models. Each hyperspectral image band was ranked using all five methods. Twenty best bands were selected by each method with the focus on soil and plant canopy characterization in precision agriculture. The results showed that each of these methods may be appropriate for different applications. The entropy measure and PCA were quite useful for selecting bands with the most information content, while derivative methods could be used for identifying absorption features. ANN measure was the most useful in selecting bands specific to a target characteristic with minimum information redundancy. The results also indicated that a combination of wavebands with different bandwidths will allow use of fewer than 20 bands used in this study to represent the information contained in the top 20 bands, thus reducing image data dimensionality and volume considerably.*

**Keywords.** *Band selection, Data mining, Hyperspectral, Precision agriculture, Remote sensing.*

Precision farming practices are conceptually based on within-field variability information. Modern sensing technologies including remote sensing for information gathering results in large volumes of raw data. Remote sensing can provide high-resolution data on geospatial variability in yield-limiting soil and crop variables. It can be used for mapping soil characteristics (Ahn et al., 1999; Bajwa et al., 2001), leaf area index (LAI), crop development (Ghinelli and Bennet, 1998), canopy coverage, pest infestation, plant water content (Bach and Mauser, 1995), and crop stresses (Gopalapillai and Tian, 1999; Bajwa

and Tian, 2001). Lack of data mining tools and inability of agriculture producers and consultants to extract useful information from large volumes of raw data on yield-limiting factors is a major hurdle in the widespread application of remote sensing at production level.

The benefits of hyperspectral sensors come from relatively high spatial and spectral resolutions (Birk and McCord, 1994). The high spectral resolution aids in discovering subtle differences in narrow-band reflectance caused by different vegetation and soil characteristics that are not discernible with multispectral data. However, much of the information in hyperspectral data is redundant. Individual hyperspectral bands are spatially and spectrally correlated (fig. 1). Removal of redundancy and identification of the most significant hyperspectral bands for characterizing a specific geospatial phenomenon is of utmost interest to several application domains. Most of the commercially available remote sensing algorithms and software packages were developed for multispectral data processing, mainly for commercially available image data (Karimi, 1998). There are not many analytical tools developed for hyperspectral data processing. The analytical tools developed for multispectral data fail to make use of added information in hyperspectral data (Schowengerdt, 1997). Although high dimensionality in hyperspectral images results in increased class separability in classification problems, it leads to high errors in parameter estimation (Hsieh and Landgrebe, 1998a). There is ongoing research to modify parametric classification methods to fully utilize added information in hyperspectral data without the loss of accuracy in parameter

---

Article was submitted for review in June 2003; approved for publication by the Information & Electrical Technologies Division of ASAE in April 2004.

The use of trade names, proprietary products, or specific equipment is only meant to provide specific information to the reader, and does not constitute a guarantee or warranty by the University of Arkansas or the University of Illinois and does not imply the approval of the named product to the exclusion of other products that may be suitable.

The authors are **Sreekala G Bajwa, ASAE Member Engineer**, Assistant Professor, Department of Biological and Agricultural Engineering, University of Arkansas, Fayetteville, Arkansas; and **Peter Bajcsy**, Associate Research Scientist, National Center for Supercomputing Applications, **Peter Groves**, Graduate Student, Department of Computer Sciences, and **Lei F. Tian, ASAE Member Engineer**, Associate Professor, Department of Biological and Agricultural Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois. **Corresponding author:** Sreekala Bajwa, Assistant Professor, Department of Biological and Agricultural Engineering, University of Arkansas, 203 Engineering Hall, Fayetteville, AR 72701; phone: 479-575-2878; fax: 479-575-2846; e-mail: sgbajwa@uark.edu.

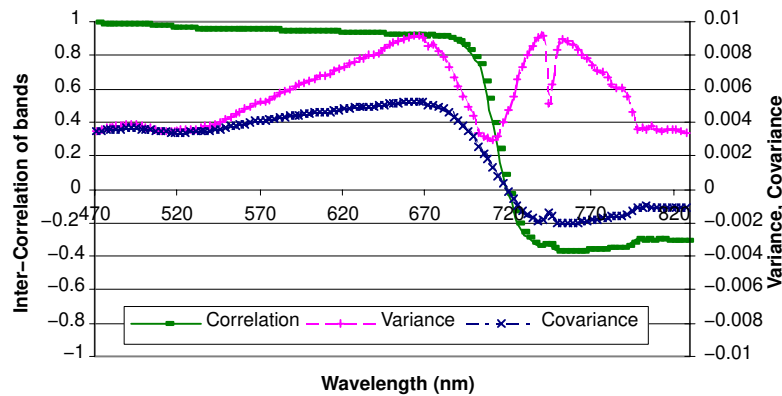


Figure 1. Interrelationship between image bands of a hyperspectral image. Correlation and covariance of band 1 with respect to all 120 bands, and band 1 variance, are plotted against band wavelength for field 2.

estimation (Hsieh and Landgrebe, 1998b; Tu, 2000). An alternate method that would extract information pertinent to one or more target properties without assuming a data distribution would be of great interest to applications in precision agriculture, forestry, and environmental monitoring.

The concept of signature bands or bands that are most responsive to a specific target characteristics is similar to the spectral fingerprints used in geological and mining applications (Schowengerdt, 1997). Unlike the space-scale aspect of spectral fingerprints, the signature bands utilize the absorbance-reflectance characteristics of individual wavebands or waveband regions. The experimental application of hyperspectral data in precision farming is focused on mapping narrow-band target reflectance to ground measurements used for crop characterizations. Signature band identification will be useful not only for geospatial characterization of specific target properties but also for application-dependent hyperspectral data compression and sensor characterization.

Our goal is to evaluate five methods for band selection from hyperspectral data with or without knowledge of the application domain. Signature bands can be extracted by using two fundamental approaches: unsupervised and supervised. The unsupervised approaches are application independent. They include information entropy measure and first and second derivatives along the spectral axis. The supervised approaches select hyperspectral bands based on supplemental ground data using principal component analysis (PCA) and an artificial neural network (ANN) based model.

We focus on methods for hyperspectral band selection that lead to data compression and spectral signature characterization. From a compression standpoint, these methods differ from pixel-based compression techniques such as that defined by the Joint Photographic Expert's Group (JPEG) (Taubman and Marcellin, 2000). The proposed methods compress data at the band level rather than at the pixel level as done by JPEG (Taubman and Marcellin, 2000). From a spectral signature characterization viewpoint, we propose a variety of methods that are appropriate for representation (information entropy, PCA) or for class discrimination (first and second derivative along spectral axis, ANN). While concise image representation and class discrimination methods have different objectives, a tradeoff of both objectives can provide satisfactory compression and classification results for an application. For example, an orthogonal subspace projection approach to classify images consider-

ably reduced data dimensionality while detecting spectral signatures of interest (Harsanyi and Chang, 1994). Another example would be a statistical separation method based on spatial autocorrelation that optimized the selection of bands with and without *a priori* knowledge of the scene (Petrie and Heasler, 1998).

The three major objectives of our study were to:

- Research and develop three unsupervised methods for rank ordering of bands from hyperspectral imagery. The methods are based on evaluation of each band separately with three different criteria. The first method is based on an information entropy criterion. The second and third methods are based on a residual criterion derived from the first and second derivatives along spectral axis with two and three adjacent bands, respectively.
- Design and implement two supervised methods based on ANN and PCA that use ground truth data for band selection.
- Compare the results obtained from the unsupervised and supervised methods.

## METHODOLOGY

### EXPERIMENTAL DESIGN

Remote sensing data were acquired from three agricultural fields located in the Midwest U.S. These three fields included field 1 and field 2 located in Missouri, and field 3 located in Illinois (table 1). Hyperspectral image data were collected from fields 1 and 2 on April 26, 2000, and from field 3 on June 8, 2000. The supplemental ground data for fields 1 and 2 included two measures of apparent soil electrical conductivity ( $EC_a$ , in mS/m) measured with two different devices. The first device was a Geonics EM-38RT (Geomatrix Earth Science, Ltd., Hockliffe, U.K.), which measured  $EC_a$  in the top 120 cm of soil using a non-invasive method called electromagnetic (EM) induction. The  $EC_a$  data collected with the Geonics EM-38RT are referred to as EM data in this article. The second device was a VERIS 3100 soil mapping system, which is a direct-contact soil EC meter. The  $EC_a$  data collected by the VERIS mapping system are referred to as VERIS data. The VERIS shallow (VERIS-s) represented the  $EC_a$  for the top 33 cm soil, and the VERIS deep (VERIS-d) represented the  $EC_a$  for the top 100 cm soil.

**Table 1. Descriptions of the three fields used in the study.**

Field	Area (ha)	Crop	Image Date
Field 1, Missouri	36	Soybean	April 26, 2000
Field 2, Missouri	15	Corn	April 26, 2000
Field 3, Illinois	16	Soybean	June 8, 2000

The EC<sub>a</sub> is a good indicator of yield-limiting soil physical and chemical properties such as soil texture, Ca, Mg, K, and CEC in some soils, and soil water content (Fritz et al., 1999; Kitchen et al., 2000). The soil EC measurements have long been used to identify contrasting soil properties in the geological and environmental domains (Lund et al., 1999).

The ground data for field 3 included canopy coverage, measured as a fraction of ground area covered by crop canopy. The canopy coverage was measured for every 16 cm square area in the field. The canopy coverage was collected on June 8, 2000, with a vehicle-mounted vision system (Tian et al., 1999). Ground truth data were collected with the objective of characterizing specific geospatial variability within a field at a maximum possible resolution. The resolution of the VERIS and EM data was approximately 10 m between adjacent passes. A summary of the ground truth database used in this study is given in table 2. Both EM and VERIS data showed very high variability within each field.

**IMAGE ACQUISITION AND PREPROCESSING**

The image data were collected from an aerial platform with a NASA RDACS/H-3, 120-channel prism-grating, pushbroom hyperspectral sensor. Each image had 2500 rows, 640 columns, and 120 bands per pixel. The 120 bands corresponded to the visible to near-infrared range of 471 to 828 nm, recorded at a spectral resolution of 3 nm. The hyperspectral sensor was mounted on a fixed-wing aircraft and was flown over the fields for data collection. The images were collected from altitudes of approximately 1200 m and 2250 m. The ground spatial interval (GSI) of the images was approximately 1 m for fields 1 and 2, and 2 m for field 3. The image preprocessing included four steps: (1) image calibration for sensor noise, (2) correction for geometric distortion caused by platform motion, (3) image georegistration, and (4) image calibration for variable illumination (fig. 2). The sensor-based noise varied with weather conditions. On each day of data collection, an image was taken with the lens covered to obtain the noise data for the sensor. This image was called the dark image. To minimize sensor noise, the raw image was first transformed with the help of the dark image into a minimum noise fraction (MNF) image (Green et al., 1988). The MNF transformation is similar to a principal

**Table 2. Geographic database of the experimental fields listing the statistics on apparent soil electrical conductivity measurements (EC<sub>a</sub>) made with EM and VERIS instruments (in mS/m) and canopy coverage as a ratio of canopy area to ground area.**

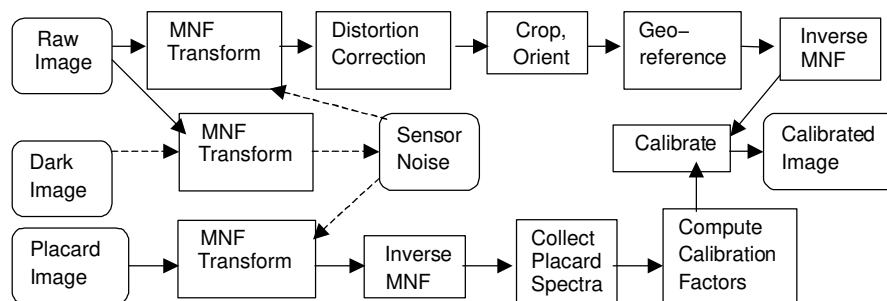
Field		Mean	Variance	Min.	Max.
Field	Variable <sup>[a]</sup>				
Field 1 (Missouri)	EC <sub>a</sub> -EM	30.71	14.08	15.85	77.74
	EC <sub>a</sub> -VERIS-s	9.74	9.89	1.60	34.70
	EC <sub>a</sub> -VERIS-d	19.61	72.24	2.30	53.80
Field 2 (Missouri)	EC <sub>a</sub> -EM	34.79	40.83	0	60.33
	EC <sub>a</sub> -VERIS-s	15.21	56.74	1.50	62.30
	EC <sub>a</sub> -VERIS-d	23.71	133.51	2.60	77.00
Field 3 (Illinois)	Canopy coverage	0.203	0.017	0	1

<sup>[a]</sup> EM refers to EC<sub>a</sub> measured with a Geonics EM38RT, which measures up to 120 cm depth in soil. VERIS-s refers to EC<sub>a</sub> measured up to 33 cm depth, and VERIS-d refers to EC<sub>a</sub> measured up to 100 cm depth with VERIS equipment.

component transform. The bands of the MNF image that contained significant information were transformed back to image space using inverse MNF transform.

The hyperspectral images acquired with the airborne pushbroom scanner showed varying degrees of geometric distortion caused by platform motion under atmospheric turbulence. Roll, pitch, and yaw of the aircraft led to significant geometric errors in pushbroom scanner data. Among the platform-based errors, the aircraft roll was the major source of image distortion, followed by aircraft pitch. The images after correcting for sensor noise were preprocessed to eliminate platform-based geometric distortion. Figure 3a illustrates image distortion caused by aircraft roll. The image shows gradual shifting of linear features, resulting in a distorted field boundary. To eliminate roll, first a known linear feature along the in-track direction was identified in the image. Then each row of the image was shifted in the cross-track direction to straighten the known linear feature. After correcting the image for geometric error caused by platform roll, the image layout (fig. 3b) appeared closer to the actual field layout. The geometric distortion due to pushbroom scanner geometry was considered negligible because of the low altitude of the airborne sensor (a maximum of 7500 ft) and the small area covered in each image.

The distortion-corrected images were then georeferenced using a triangulation method with bilinear resampling (fig. 3). Fifteen to twenty ground control points (GCP) distributed inside and outside the field boundary were used for georeferencing each image. Markers or white-painted boards placed in and around the field along with static features such as pumps, crossroads, and field corners were



**Figure 2. Flowchart showing the steps in image preprocessing for correction, georeferencing, and calibration (“MNF transform” refers to minimum noise fraction transformation).**

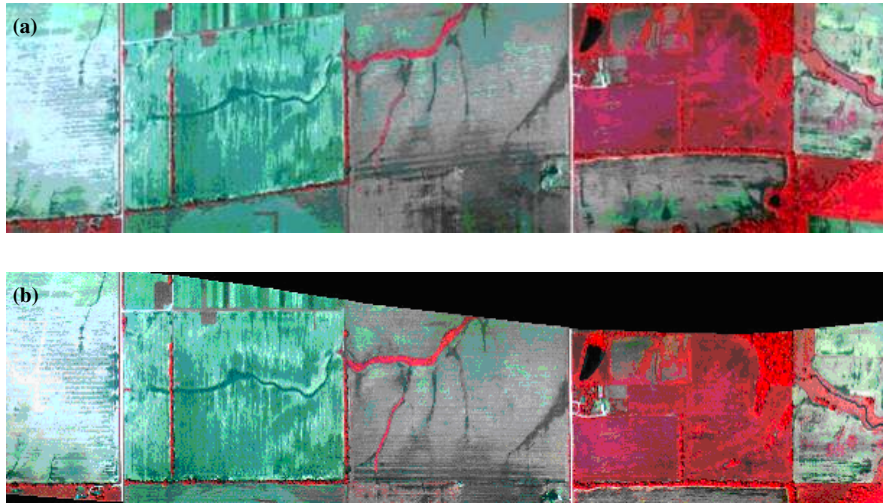


Figure 3. A raw image of field 1: (a) platform-based image distortion, and (b) after correcting for error caused by platform roll.

used as GCPs. All the images were corrected and calibrated using ENVI software package (Research Systems, Inc., Boulder, Colo.). The georeferenced images were corrected for variable illumination using an empirical line method (Smith and Milton, 1999). The raw images come as digital numbers (DN) or gray-level values. The empirical line calibration procedure transformed the DNs to apparent reflectance of the target, which does not change significantly with illumination conditions. For applying the empirical line method of calibration, markers or tarps (also called placards) of eight different reflectance levels of 2%, 4%, 8%, 16%, 32%, 48%, 64%, and 83% were placed on the field at the time of imaging. Hyperspectral images of placards were taken at the same resolution as the hyperspectral images of the field on each collection day. The empirical line method assumes that the calibrated values are linearly related to the uncalibrated values. This linear relationship between the known reflectance and the observed DNs of the placards was calculated by linear regression and applied to the whole image. The image acquisition time was limited to between 10:00 a.m. and 2:00 p.m. to minimize the errors due to the changes in sun angle.

#### UNSUPERVISED METHODS FOR BAND SELECTION

The three unsupervised methods of hyperspectral band selection did not require image preprocessing since they are application-independent techniques based on spatial and spectral similarity of bands. The two supervised methods (PCA and ANN) required image preprocessing into apparent reflectance since they used a model of target characteristics on hyperspectral data to establish the band order. The first and second derivatives and the ANN methods were implemented in the software environment called Data to Knowledge (D2K) developed under the Image to Knowledge (I2K) tools at the National Center for Supercomputing Applications (NCSA, 2000).

##### Method 1: Information Entropy

Information entropy is based on evaluating each band separately using the entropy measure ( $H$ ) defined in equation 1:

$$H = -\sum_{i=1}^m p_i \ln p_i \quad (1)$$

where  $p$  is the probability of occurrence of a DN in a hyperspectral band, and  $m$  is the number of distinct DNs in that band. The probabilities are estimated by computing a histogram of DNs. Generally, if the entropy value ( $H$ ) is high, then the amount of information in the data is large. Thus, the bands are ranked in descending order from the band with the highest entropy value (large amount of information) to the band with the smallest entropy value (small amount of information).

##### Method 2: First Spectral Derivative

The bandwidth of each band can be a variable in hyperspectral sensor design. The first spectral derivative method explores the bandwidth variable as a function of added information. It is apparent that if two adjacent bands do not differ much, then the underlying geospatial phenomenon can be characterized with only one band. The mathematical description of first spectral derivative is illustrated in equation 2:

$$D_1 = \frac{\partial I(x, \lambda)}{\partial \lambda} \quad (2)$$

where  $I$  represents the hyperspectral value,  $x$  is a spatial location, and  $\lambda$  is the band characteristic or central wavelength. Thus, if  $D_1$  is equal to zero, then one of the bands is redundant. In general, adjacent bands that differ a lot should be preserved for characterization, while similar adjacent bands can be reduced. In addition to evaluating band characteristics, this method is instrumental in bandwidth design of a remote hyperspectral sensor.

##### Method 3: Second Spectral Derivative

Similar to method 2, the second spectral derivative explores the bandwidth variable in hyperspectral imagery as a function of added information. Contrary to method 2, this approach identifies bands that can be represented by a linear combination of adjacent bands. Thus, if two adjacent bands can linearly interpolate the third band, then the third band is redundant. The larger the deviation from a linear model, the higher the information value of the band. The mathematical description of this method is shown in equation 3:

$$D_2 = \frac{\partial^2 I(x, \lambda)}{\partial \lambda^2} \quad (3)$$

where  $D_2$  represents the measure of linear deviation,  $I$  is a hyperspectral value,  $x$  is a spatial location, and  $\lambda$  is the band wavelength.

#### SUPERVISED METHODS FOR BAND SELECTION

##### Method 4: Artificial Neural Network (ANN)

Artificial neural networks are used in a wide variety of applications and in many disciplines because of their robustness in making predictions based on training examples. They are particularly applicable in agricultural problems for modeling complex relationships, where stochastic factors play major roles (Mukherjee, 1997). An example of ANN application in precision agriculture is to model the relationship between yield-limiting factors estimated with remote sensing data and yield measurements (Goh, 1995; Gopalapillai and Tian, 1999). Stochastic factors cause significant spatial variations in crop yield (20% to 80%, according to Bakhsh et al., 1998). Therefore, remote sensing is a valuable tool to monitor the in-season effect of stochastic factors on a near-real-time basis.

In order to train the ANN algorithm, training data was converted into ASCII format and then normalized. A plain text file was created with entries containing geographical position, field data (EM, VERIS, and canopy), and the reflectance data from each of the 120 bands of spectral data. Each column of the table (every vector of ground truth data and each spectral band) was independently normalized and scaled to a range of 0.1 to 0.9 for the benefit of the ANN. Although a neural network can normally output values between  $-1$  and  $1$ , values near zero can create problems with the learning and optimization processes. A multilayer feed-forward ANN was used for the training (Russel and Norvig, 1995). A genetic algorithm was used to optimize ANN topology (Goldberg, 1989, 1999). A non-standard activation function called Elliot's Proposed Activation Function (Elliot, 1993) was used to obtain the final results.

The hyperspectral bands were ranked based on their sensitivity to the ANN output value. A band with a high sensitivity had a high rank. The ANN processed a subset of the training data set. Every band input varied between 0.1 and 0.9 with a 0.05 increment for each band. After all the training examples had been explored, the mean score for each band was calculated. The rank of the best 60 bands based on the mean range of predictions for an ANN of 120 bands was retained for further examination. Similarly, the best 30 from 60, best 15 from 30, and the best 10 from 15 bands were identified and retained. The ANN parameters for each set included five nodes per hidden layer, one hidden layer, and a learning rate of 0.02. The number of iterations was changed with the number of input bands to compensate for the fact that an ANN takes more time and learns more from data when there are more bands per iteration. The iteration counts were 4000, 4000, 6000, 8000, 10000, and 12000 for 120, 60, 30, 15, 10, and 5 bands, respectively.

##### Method 5: Principal Component Analysis (PCA)

Multivariate analysis using PCA was conducted on 120 bands of the image to obtain the most significant bands characterizing spatial variability in a specific target characteristics represented in the field data. The PCA transformed the auto-correlated hyperspectral image bands to uncorrelated principal components based on the band covariance matrix. A correlation analysis was performed between the

principal components or bands of the transformed image and the ground truth data. The most significant bands were identified from their corresponding eigenvectors in the principal component that showed maximum correlation with the field data. The eigenvalue represents the degree of variance represented by each principal component. The maximum variance in the image is carried by the first principal component image, and the variance decreases for higher-order principal components. Therefore, the first few principal components are expected to represent the global variability in the image scene, and the latter principal components are expected to represent information on local variability, such as the variability in the target characteristics explored in this study.

## RESULTS AND DISCUSSION

The results are grouped based on the underlying two approaches to band selection (unsupervised and supervised). Discussion on spectral signature extraction and hyperspectral data compression is provided to illustrate benefits of band selection methods in precision agriculture. Our goal was to evaluate all methods based on the ground truth data collected at a few spatial locations in the three experimental fields. We selected subregions of the raw hyperspectral images corresponding to the bounding boxes of ground truth measurement points so that the results from all five methods for band selection were comparable. In all presented results, the band

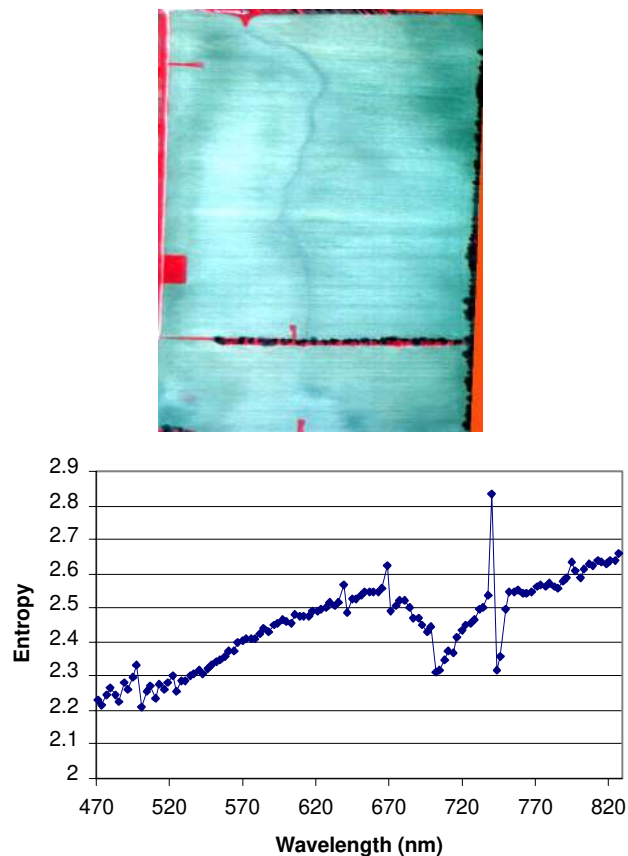


Figure 4. Hyperspectral image of bare soil in field 1 (top) and its band entropy measure (bottom) plotted as a function of the band wavelength for field 1. Entropy is unitless. Larger entropy measures indicate high information content.

**Table 3. Top 20 bands selected from hyperspectral image of bare soils by entropy, first derivative, and second derivative measures. The image was collected on April 26, 2000, from field 1 and field 2.**

Method 1: Entropy				Method 2: First Derivative (F.D.)				Method 3: Second Derivative (S.D.)			
Field 1		Field 2		Field 1		Field 2		Field 1		Field 2	
Band	Entropy	Band	Entropy	Band	F.D.	Band	F.D.	Band	S.D.	Band	S.D.
741	2.939	741	2.83	741	10.37	744	8.72	741	18.92	741	16.80
828	2.695	828	2.66	738	8.071	741	8.08	744	10.02	738	7.57
825	2.681	825	2.64	669	2.915	672	2.34	738	7.729	744	7.29
822	2.678	822	2.64	747	2.397	750	1.77	669	4.613	669	3.84
795	2.674	813	2.64	666	1.571	795	1.63	672	3.573	747	3.20
807	2.668	795	2.63	498	1.547	669	1.50	747	3.098	672	2.74
804	2.667	816	2.63	639	1.535	642	1.47	699	2.665	639	2.25
819	2.665	819	2.63	792	1.498	747	1.43	642	2.533	642	2.24
813	2.663	807	2.63	699	1.348	501	1.37	501	2.370	699	2.22
810	2.661	810	2.62	696	1.247	699	1.29	639	2.350	501	2.02
816	2.654	669	2.62	750	1.013	804	1.14	498	2.188	498	1.88
669	2.651	804	2.61	801	0.928	753	1.02	795	1.782	795	1.63
798	2.650	798	2.61	642	0.913	702	0.94	510	1.683	801	1.62
792	2.632	792	2.59	510	0.830	639	0.78	702	1.600	666	1.25
801	2.628	801	2.59	507	0.821	645	0.77	507	1.476	510	1.24
789	2.622	789	2.58	711	0.777	792	0.66	801	1.461	696	1.16
786	2.615	780	2.57	501	0.742	504	0.65	750	1.420	507	1.15
783	2.608	774	2.57	636	0.726	714	0.63	666	1.379	714	1.04
771	2.607	639	2.57	789	0.653	510	0.62	696	1.253	702	1.04
768	2.605	783	2.56	504	0.635	513	0.61	714	1.129	804	1.00

**Table 4. Top 20 bands selected by unsupervised methods 1 (entropy), 2 (first derivative), and 3 (second derivative) from hyperspectral image of field 3 with partial crop coverage. The image was collected on June 8, 2000.**

Method 1		Method 2		Method 3	
Band	Entropy	Band	F.D.	Band	S.D.
669	2.31	744	1.54	744	1.68
663	2.31	822	0.70	741	1.25
666	2.30	756	0.62	819	0.94
660	2.30	795	0.58	822	0.78
657	2.29	813	0.53	777	0.74
672	2.29	777	0.53	750	0.73
654	2.29	768	0.46	756	0.65
651	2.27	717	0.45	699	0.64
648	2.27	690	0.43	762	0.61
675	2.27	699	0.38	753	0.60
639	2.25	681	0.36	747	0.57
678	2.25	804	0.34	816	0.56
645	2.25	693	0.34	702	0.49
681	2.24	798	0.33	780	0.44
642	2.24	816	0.32	768	0.40
684	2.23	684	0.32	519	0.37
630	2.22	765	0.32	810	0.35
633	2.22	504	0.31	516	0.34
636	2.22	522	0.30	714	0.34
627	2.20	480	0.28	717	0.34

index or band number from an interval [1,120] was converted to the central wavelength ( $\lambda_b$ ) of a given band ( $b$ ) based on equation 4:

$$\lambda_b = 471 + 3 \times (b - 1) \text{ [nm]} \quad (4)$$

#### UNSUPERVISED BAND SELECTION

The entropy of the soil image with no vegetative cover showed highest values mostly in the near-infrared region of 780–828 nm (fig. 4 and table 3). Information entropy is a

measure based on the variance of DN within each band. In general, near-infrared bands of soil images showed larger variability within each bands. Considering that entropy measure is an unsupervised method, the dominating target characteristic will cause variability in spectral bands. We expected to see more variability in color bands since clay characteristics and mineral contents are expressed more in color bands than in NIR bands. However, crop residue may be one reason for the strong band variance in the NIR bands of a soil image. The partially vegetated image showed highest entropy values in the red region of 627–684 nm (table 4). These red bands signify plant pigment absorption, mainly due to chlorophyll.

The first and second derivative measures compared the spectral derivatives between adjacent bands to select wavebands. Derivative measures showed several significant bands in the 690–705 nm range for both soil and partial canopy image. The top 20 soil image bands from derivative measures also included several bands in the 500–510 nm, 735–750 nm, and 800–805 nm ranges (fig. 5). The top 20 bands for partial canopy image included 714–717 nm, 740–755 nm, 810–825 nm, and a few green bands (fig. 6). This means that there is significant amount of information in the green, far red, red edge, and NIR regions that is relevant to soil characteristics and canopy cover. Visible light bands responded to soil characteristics very well. Soil organic matter and residue may have contributed to the variations in the NIR bands in soil images. Both color and NIR bands responded well to partially vegetated image. These results were consistent with theoretical expectations and published research findings that NIR, far red, and red edge areas of the electromagnetic spectrum carry valuable information on crop and soil characteristics (Carter, 1998; Thenkabail et al., 2002). The first and second derivative criteria compare only adjacent bands for information redundancy. Although derivative measures are better than entropy measure for reducing redundancy, they are probably not the best methods because they compare only adjacent bands.

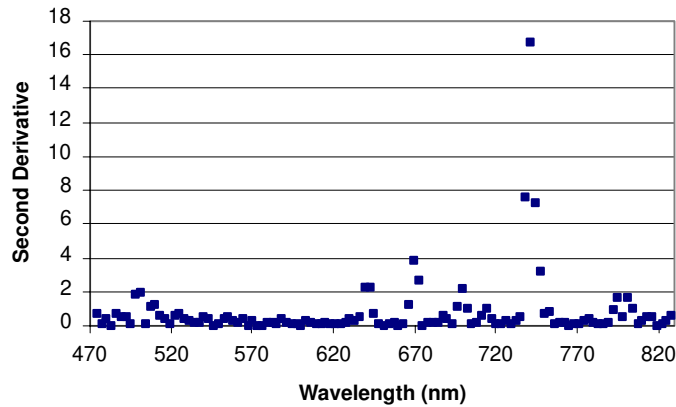


Figure 5. Hyperspectral image of bare soil of field 2 (top) and the second derivative measure (bottom) used for identifying bands with non-redundant information, plotted against band wavelength.

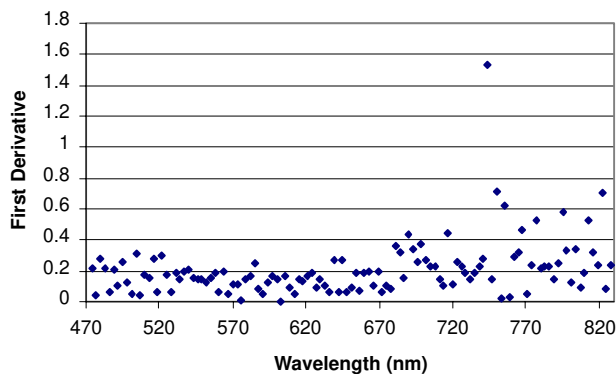
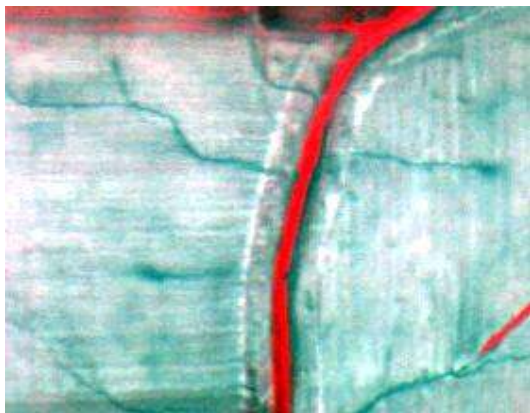


Figure 6. Hyperspectral subimage of bare soil of field 3 (top) and the first derivative measure (bottom) used for identifying bands with non-redundant information plotted against band wavelength. Larger values for first derivative measure (unitless) show that the two bands have non-redundant information.

### SUPERVISED BAND SELECTION *Artificial Neural Network (ANN)*

The outcome of sensitivity analysis of bands with an ANN model is shown in figure 7 as a mean prediction range against the band wavelength. The band sensitivity differed significantly among the bands, varying from 0 to 0.5. The spectral bands at chlorophyll absorption and red shift showed the maximum sensitivity (fig. 7). The top 20 bands selected by ANN included one or more bands from NIR (740–750 nm), red edge (710–740), green (530–550 nm), and several red bands for the partial canopy image (table 5). In addition to the above bands, several red and far red (690–710 nm) bands were highly significant in characterizing canopy. The NIR and red regions are understandably more responsive to partially vegetated fields because of the role of plant pigments in attenuating visible light bands and of biomass (cell structure) in attenuating NIR wavelengths.

The soil image showed a mixed set of individual narrow bands from the visible and red edge regions. The red edge (740–747 nm), green (540–546 nm), and scattered red bands in the range of 640–675 nm were most sensitive to  $EC_a$  (table 5). The absence of chlorophyll absorption bands in the far red of the soil image and the difference in the selected red bands between the soil and canopy images are noteworthy here. Soil clay particles, organic matter, residues, and mineral content such as iron, calcium, and magnesium are expected to generate a different response of soil to light spectra than that of the canopy. The significant differences in band sensitivity to soil  $EC_a$  and canopy density provide validity to the concept of signature bands, which are narrow wavebands that are most responsive to a target characteristic. Some of the adjacent bands showed similar responses, such as bands in the 774–777 nm ranges for representing canopy density and bands 741–747 nm range for representing soil  $EC_a$ , suggesting that broad bands in the respective range can be substituted for individual narrow bands.

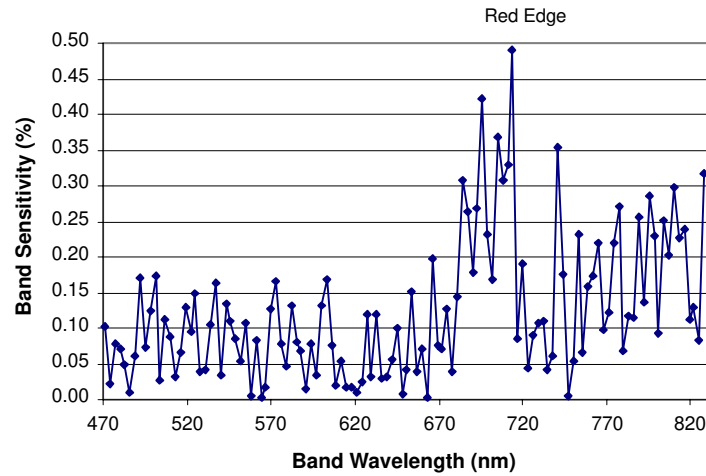


Figure 7. Band sensitivity of hyperspectral image data estimated with a supervised neural network model, plotted against wavelength for a partially vegetated field (field 3).

**Principal Component Analysis (PCA)**

The individual bands of a hyperspectral image are spatially and spectrally correlated. PCA transforms the image bands into orthogonal, and hence uncorrelated, principal components. The hyperspectral images were transformed using multivariate analysis into 15 principal component bands, since higher-order principal components carried very little information. Most of the variability in the images was represented by the first ten principal components. The first three to four principal components showed a gradual trend in the contribution of individual bands represented by their eigenvector (fig. 8a). However, in principal components 5 to 10, a few bands dominated in their contribution to each principal component (fig. 8b).

During our investigation of principal components with respect to ground truth data, different higher-order principal components showed significant correlation to soil apparent electrical conductivity and crop canopy. Each of the target characteristics showed varying degrees of correlation with different principal components. Similarly, the band responsiveness to a particular geospatial characteristic varied from field to field. For the bare soil image of field 1, principal components 4, 5, and 7 showed the highest correlation with the three measures of soil electrical conductivity (fig. 9). VERIS deep showed maximum correlation of 0.39 with PC4, EM showed maximum correlation of 0.49 with PC5, and VERIS shallow showed a maximum correlation of 0.37 with PC7. Based on the eigenvectors, the major contributors for these three principal components were spectral bands in the

**Table 5. Twenty most responsive wavelength bands obtained from supervised ANN training and PCA on field 2 image with EM data, and field 3 image with canopy density data. Sensitivity of individual bands to the neural net model is taken as the neural net measure. Eigenvector of the principal component that is most correlated to the field variable is taken as the principal component measure. The three measures of principal component are: eigenvectors of PC1, PC1+PC2, and average of the first ten PCs.**

Neural Net Training				Principal Component Analysis			
Field 2		Field 3		Field 2		Field 3	
Waveband	Sensitivity	Waveband	Sensitivity	Waveband	Eigenvector	Waveband PC1+PC2 / PC1 / avg(10)	Eigenvector PC1+PC2
741	0.084	774	0.034	741	0.562	669 / 669 / 828	0.195
522	0.064	783	0.026	747	0.302	663 / 663 / 741	0.193
546	0.061	765	0.022	744	0.268	666 / 666 / 744	0.193
744	0.060	741	0.017	501	0.234	660 / 660 / 747	0.192
642	0.059	759	0.017	498	0.229	657 / 657 / 810	0.191
747	0.059	603	0.015	639	0.170	654 / 654 / 510	0.189
516	0.052	696	0.014	750	0.168	744 / 672 / 507	0.188
543	0.046	777	0.014	669	0.165	651 / 651 / 822	0.186
483	0.036	798	0.013	642	0.131	786 / 648 / 699	0.186
540	0.034	543	0.013	495	0.131	789 / 675 / 513	0.185
570	0.032	633	0.012	672	0.130	756 / 645 / 804	0.185
648	0.032	600	0.011	606	0.121	783 / 639 / 474	0.185
675	0.030	705	0.010	675	0.117	672 / 678 / 522	0.184
477	0.029	693	0.009	699	0.113	765 / 642 / 798	0.184
672	0.027	828	0.008	516	0.098	768 / 636 / 471	0.184
615	0.025	471	0.008	630	0.097	759 / 633 / 669	0.183
660	0.024	534	0.008	486	0.095	771 / 681 / 516	0.183
723	0.023	714	0.006	489	0.094	780 / 630 / 816	0.183
765	0.020	666	0.006	735	0.090	774 / 627 / 693	0.182
501	0.017	804	0.006	492	0.088	648 / 684 / 696	0.182



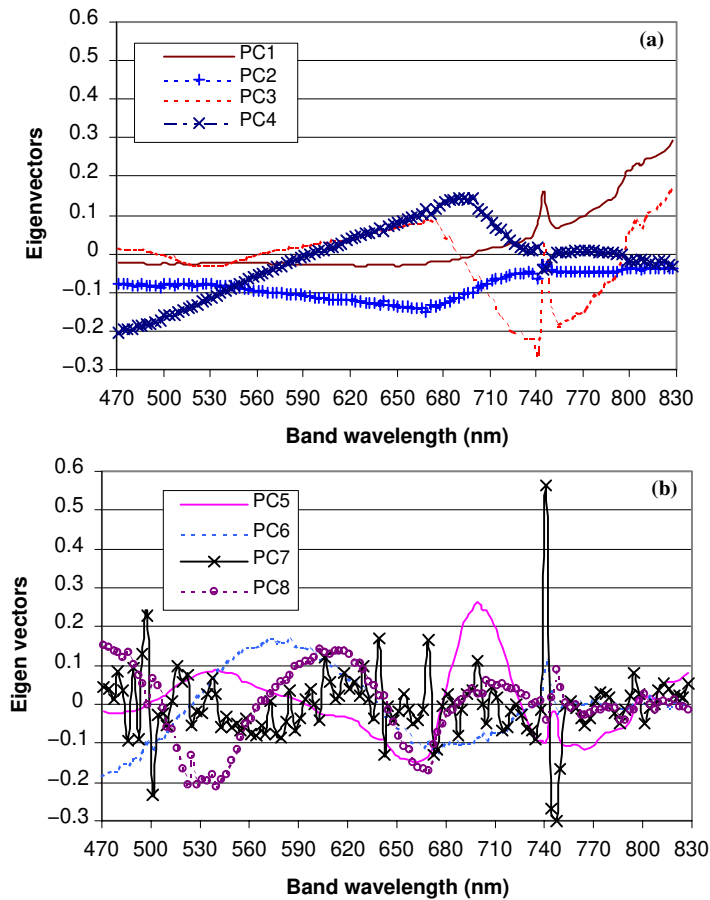


Figure 8. Individual band contributions to each principal component (band eigenvectors) of field 2 image acquired on April 26, 2000: (a) principal components 1 through 3, and (b) principal components 5 through 8.

range of 741–747 nm and 490–510 nm for PC7, 705–717 nm and 669–672 nm for PC5, and 693–700 nm and 470–499 nm for PC4 (fig. 7). The high correlation of  $EC_a$  with lower-order PCs such as 4, 5, and 7 shows that  $EC_a$  was not the dominating factor that determined the spatial variability in soil reflectance. Although some of the surface factors such as soil moisture, clay content, and mineral concentration that contribute to  $EC_a$  affect soil reflectance,  $EC_a$  characteristics of deeper layers may not affect soil reflectance. Similarly, other surface

factors such as crop residue may dominate soil reflectance but may not affect  $EC_a$  significantly.

Principal component analysis of the partially vegetated scene in field 3 showed a very different trend from that of the soil images (fig. 10). The first two PCs showed the highest correlations (0.38 and 0.35) with canopy density. This can be explained by the fact that the major variance factor in this image is crop, rather than soil. In the soil image, the  $EC_a$  only partially contributed to the overall variability in soil reflectance captured in the image. Subtle differences in scene re-

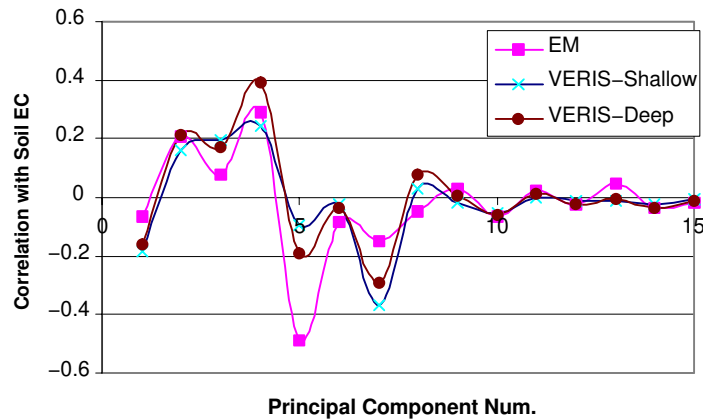


Figure 9. Correlation between principal components and soil characteristics plotted against principal component bands derived from hyperspectral images of bare soil in field 1.

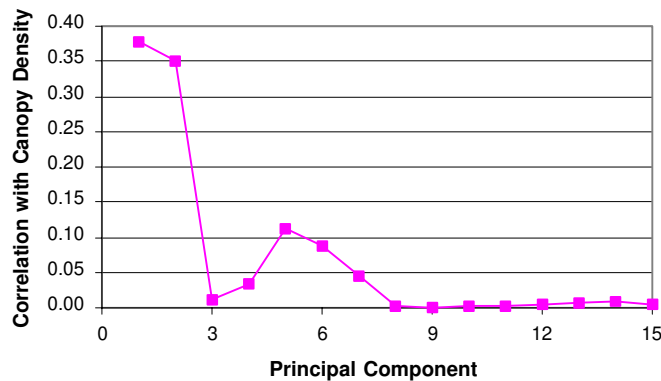


Figure 10. Correlation between principal components derived from hyperspectral image of partially vegetated field (field 3) and canopy density.

flectance are represented by higher-order PCs as opposed to major differences in scene reflectance. Major contributing wavebands to PC1 were from the red region of the spectrum and to PC2 were from the near-infrared region of the spectrum (fig. 11). Since both PC1 and PC2 were highly correlated to canopy density (fig. 10), and they represented two equally significant regions of the spectrum, the sum of the absolute values of PC1 and PC2 was considered as the best measure for selecting the top 20 bands.

#### COMPARISON OF BAND SELECTION METHODS

When the results from the five band selection methods described in this article were compared to each other, the wavebands in the 741–747 nm range became the most significant bands for soil characterization, identified by four out of the five methods (tables 3 to 5). Comparison of the three unsupervised techniques on soil images of two fields resulted in only two common bands (669 and 741 nm) in the red and NIR regions. However, there were several bands that were within  $\pm 3$  nm between the three unsupervised methods and the two fields. There were more common bands between entropy and FD, and between FD and SD. The top 20 bands corresponded very well in both fields under each measure. Out of the 20 selected bands, 17 under entropy measure, 11 bands under first derivative, and 19 bands under second derivative were common between the two fields.

The top 20 bands selected with the three unsupervised methods on the field 3 image that had partial vegetative cover

showed no common band (table 4). There were 8 common bands among the top 20 bands selected by the first and second derivative methods. The lack of conformity between the three measures lies in the way they work. Entropy measure is similar to variance and follows band variance very closely (fig. 12). Although entropy measure accounts for the information content in bands, it fails to consider information redundancy between different bands. Therefore, entropy may not be adequate as a standalone method for band selection. The first derivative measure looks at information redundancy between the two adjacent bands, and it does not select the second band if that band does not offer any new information. Similarly, second derivative compares three adjacent bands to select a band without redundancy. In the case of partially vegetated field image, most of the image variance or information content was in the red region (fig. 12). However, the first and second derivatives reduce redundancy by eliminating an adjacent band that contains redundant information.

Principal component 5 was significantly correlated to EM data in both fields 1 and 2. Therefore, the eigenvector of PC5 was considered as the principal component measure. Principal components 1 and 2 were most significant for characterizing crop canopy density; therefore, the average of eigenvectors of PC1 and PC2 was considered as the principal component measure for canopy density classification. The bare soil image of field 2 resulted in eight common bands among the top 20 bands selected based on principal component measure and neural net measure. Top 20 bands selected

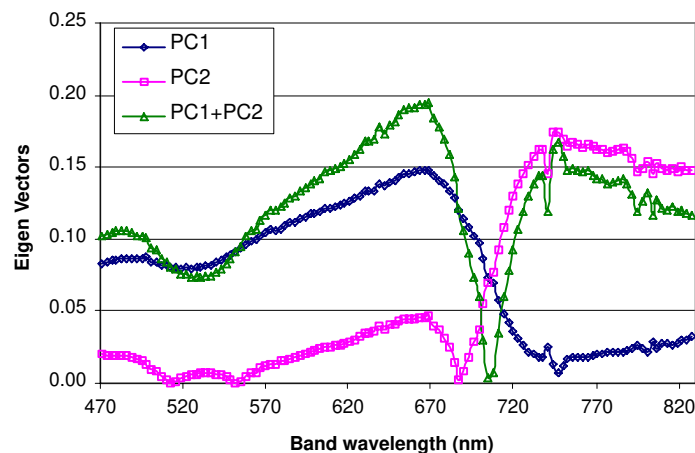


Figure 11. Absolute values of eigenvectors of first two principal components (PC1 and PC2), and their sum (PC1+PC2) of field 3 image plotted against wavebands.

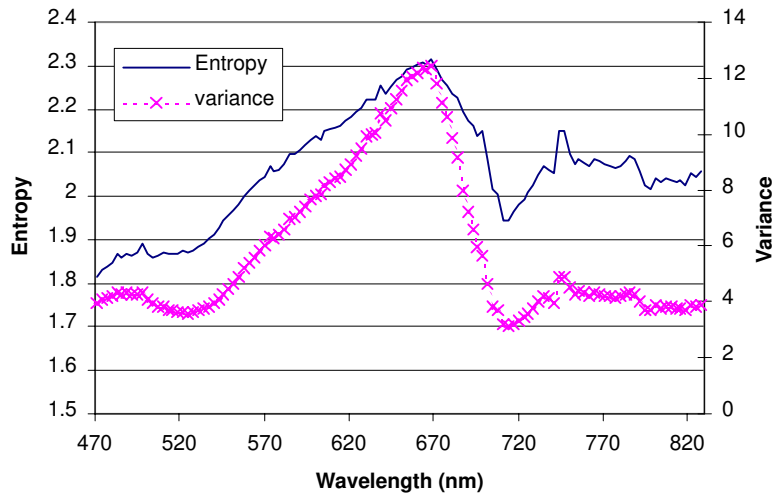


Figure 12. Entropy and variance of the 120 bands of field 3 image from June 8, 2000, plotted against the respective band wavelengths.

based on the PC1 measure showed only two bands in common with that of neural net selection. The lack of conformity is due to the difference in how PCA and ANN select bands. PCA is based on the variance in image bands, whereas ANN selects bands specifically on the sensitivity of a predictive model to each band. The PC measure of PC1+PC2 resulted in five bands in common with the neural net measure. This is expected, as the PC1+PC2 measure represents more information by accounting for the variance represented in two principal components. Band selection based on the PC1 measure resulted in all red bands, whereas the PC1+PC2 measure resulted in both red and NIR bands. A selection criterion based on an average of the first ten PCs resulted in a more varied selection from across all the spectral regions (table 3) since this measure contained more information than the PC1 and PC1+PC2 measures.

The neural net selection had seven bands in common with the PC measure based on average of first ten PCs. The principal components are uncorrelated bands computed by a linear combination of hyperspectral bands such that the maximum variance is captured in the higher-order bands. Theoretically, it is possible to capture the subtle variance caused by a non-dominating target characteristic by this method. However, band selection based on the eigenvectors of one or more principal component does not have a mechanism to eliminate redundant bands. The neural net measure selected bands from all spectral regions. It is safe to assume that neural net method of band selection leads to a reduction in information redundancy since the sensitivity analysis computes the additional information provided by a given band by adding that band to the neural net model.

Comparison of the supervised to unsupervised methods of band selection showed that there were no common bands selected by all the methods, mainly because of the unique technique by which each method selected bands. For field 3, principal component measure based on the PC most correlated to canopy density (PC1) resulted in exactly the same results as entropy measure since both these measures are based on band variance. Both entropy and PCA rely on the variance in digital numbers within each band to calculate the respective measure for information content. The principal component measure based on average of the top ten PCs had 7 to 8 common bands with the first and second derivative

based selections. As standalone methods, both entropy and PCA-based methods are unsuitable since they do not consider data redundancy in bands. Additionally, PCA is unreliable for temporal or multi-image comparisons since the principal components change with scene illumination. Two images of the same scene taken on the same day can result in different principal components. If principal components of multiple images need to be compared, then they should first be calibrated with respect to one image. Similar to PCA, entropy is only a measure of variance. If the target characteristic of interest is the dominant factor causing variations in scene reflectance, then entropy will be an adequate measure to estimate the information content specific to the scene characteristic of interest. In many precision agriculture applications, the scene characteristic of interest causes only subtle difference in the pixel reflectance. For example, soil fertility factors, specific crop stress in early crop growth stage, etc., only contribute minimally to the total scene reflectance. In such applications, both entropy and principal component analysis are inadequate without specific modifications.

The first and second derivatives and ANN methods try to remove information redundancy to some extent while selecting the most responsive bands. The ANN measure resulted in only one common band with entropy and six each with the first and second derivative for the soil image. For the canopy image, the ANN shared two common bands with entropy, four with the first derivative, and three with the second derivative for the canopy image. In spite of the lack of common specific bands, more than half of the bands selected by the ANN measure were within a  $\pm 3$  nm range of bands selected by the first and second derivative methods. Although the derivative methods were comparing the redundancy in information between adjacent bands, it is interesting to note that several of the adjacent bands had enough pixel-to-pixel variability to be included in the top 20 bands. Spectral bands 714–717 nm, 741–753 nm, and 816–822 nm in the second derivative and bands 690–699 nm, 765–768 nm, and 795–798 nm in the first derivative are examples of adjacent bands in the top 20 bands selected. The neural net measure resulted in the least number of adjacent bands in the top 20 bands. The three adjacent bands in the 741–747 nm range were selected by the first and second

derivatives, neural net, and PCA methods when applied to the soil image. The presence of these adjacent bands in the top 20 bands implies that these three bands provide distinctly different information useful for characterizing soils. Compared to the neural net selection, the derivative methods are flawed in their ability to compare information content between bands that are distant from each other on the spectral scale. However, derivative methods are useful in identifying absorption spectra associated with water or plant pigments in the scene. These absorption bands result in sharp changes in the reflectance of the absorption bands with respect to adjacent bands, resulting in peak values for the derivative measure.

Theoretically, an ANN is capable of learning nonlinear relationships between one or more independent variables (target characteristics) and a corresponding dependent variable (target reflectance). For a partially vegetated field image, the ANN method identified several near-infrared bands indicative of the plant cellular structure, chlorophyll absorption spectrum (around 696 nm), the green peak (543 nm), and a few wavelengths in the blue-green transitional area representative of light absorption by carotenoids and other plant pigments (table 5). As expected, the ANN measure identified primarily red and green bands from the soil image to map soil electrical conductivity. These bands may indicate soil clay particles and minerals that significantly influence soil electrical conductivity. For band selection applications specific to target characteristics, the ANN method may be the best. The ANN method is capable of performing exhaustive searches in the entire application domain and is the best among the five methods considered here to avoid data redundancy issues. The drawback of ANN analysis is that an exhaustive search procedure would require supercomputers. Therefore, the time and cost associated with exhaustive an ANN search could be a limiting factor in its application. The findings of our research will aid in plant stress discrimination for precision agriculture, future hyperspectral sensor design, and crop management. As a next step, we plan to continue our efforts to model specific target characteristics on bands selected with the five different methods, and validate the performance of these models to predict soil or plant characteristics that are important in agriculture.

## CONCLUSION

Five methods of band selection (three unsupervised and two supervised) were compared for selecting signature bands from hyperspectral imagery for characterizing soil EC<sub>a</sub> and canopy density in agricultural fields. The best 20 bands selected by the five methods varied between fields and methods. The entropy measure selected bands primarily from 780–828 nm for bare soil and 627–684 nm for the canopy image. Derivative measures selected bands from 500–510 nm, 690–705 nm, 738–750 nm, 740–756 nm, and 800–805 nm for bare soil images. From the partially vegetated images, the derivative measures selected bands from 690–705 nm, 740–756 nm, and 810–825 nm. Top 20 bands selected by the ANN from the canopy image included green bands (530–550 nm), chlorophyll absorption spectra (690–710 nm), several red bands, and NIR bands from 740–750 nm. For the soil images, the ANN selected a

different set of red bands than for the canopy image, green (530–550 nm), and NIR (740–750 nm) bands. The PCA selected 740–750 nm for both soil and canopy images. The remaining bands selected by PCA included 690–710 nm and several green bands for canopy characterization, and several red bands for soil characterization.

There was some agreement between the derivative measures and the ANN since both methods reduced redundancy to some extent. The entropy measure and principal component measure based on PC1 resulted in similar bands since both of these methods were based on band variance. The entropy measure can be adopted to select bands from hyperspectral data that have the most information content for characterizing global variability within a scene. Similarly, the principal component measure may be adopted to identify bands with the most information content or information specific to a field characteristic, as done in this study. Both principal component and entropy measures were very similar to each other in that they selected bands based on information content within each band without considering information redundancy. Therefore, these two methods, although ideal for identifying bands with the most information content, are not suitable for selecting a set of bands that will provide the best characterization of a given target characteristic without redundancy.

The first and second derivatives were reasonable measures for identifying redundancy or variability between adjacent bands. The derivative measures could provide reasonable accuracy in identifying absorption bands and minimize redundancy between adjacent bands. However, the derivative measures are unable to handle information redundancy in wavebands that are more than one or two steps away from a given band. Another caveat with derivative measures is that they magnify noise in the data, if the data is not filtered for noise.

Overall, the ANN has the capability to learn the subtle changes in scene reflectance caused by a specific scene characteristic in spite of large variations in the total scene reflectance. The ANN measure based on band sensitivity compared the information added by a single band in a model to characterize a specific target variable. Therefore, the ANN measure is capable reducing information redundancy between the selected bands. In addition to identifying the most responsive bands, the ANN measure can also be used for application-specific data compression.

With the exception of the entropy measure, all methods selected several bands from the 740–750 nm range among the top 20 from all three field images, and 690–710 from the partially vegetated image of field 3. Therefore, we recommend that future sensor designs for agriculture and environmental applications may include narrow wavebands (10 nm or less) in these two regions. This study also showed that it would be possible to capture the target-specific information provided by the top 20 bands with fewer bands, such as nine bands in the case of the principal component measure in field 2. This is possible by combining some of the adjacent bands that were among the top 20 into single broader bands. Therefore, these band selection methods are quite useful for considerably reducing data volume and dimensionality.

## ACKNOWLEDGEMENTS

The authors extend their appreciation to Dr. Don Bullock, Dr. Ken Sudduth, Dr. Newell Kitchen, and Mr. Palm Harlen

for providing access to some of the data used in this study. The authors also thank the United Soybean Board and the Illinois Council of Food and Agricultural Research for their financial support.

## REFERENCES

- Ahn, C.-W., M. F. Baumgardner, and L. L. Biehl. 1999. Delineation of soil variability using geostatistics and fuzzy clustering analysis of hyperspectral data. *SSSA J.* 63(1): 142–150.
- Bach, H., and W. Mauser. 1995. Multisensor approach for the determination of plant parameters of corn. Paper No. 0–7803–2567–2/95. Piscataway, N.J.: IEEE.
- Bajwa, S. G., and L. Tian. 2001. Modeling and mapping of spatial weed density within a soybean field from aerial CIR images. *Trans. ASAE* 44(6): 1965–1974.
- Bajwa, S. G., P. Groves, and L. Tian. 2001. An application-oriented hyperspectral data compression method for precision agriculture applications. In *Proc. 2001 ASPRS Annual Meeting and Conference*, (CD-ROM). Bethesda, Md.: ASPRS.
- Bakhsh, A., B. D. Jaynes, T. S. Colvin, and R. S. Kanwar. 1998. Spatio-temporal yield variability analysis for a cotton-soybean field in Iowa. ASAE Paper No. 981049. St. Joseph, Mich.: ASAE.
- Birk, R. J., and T. B. McCord. 1994. Airborne hyperspectral sensor systems. *IEEE AES System Magazine* 9(10): 26–33. Piscataway, N.J.: IEEE.
- Carter, G. A. 1998. Reflectance wavebands and indices for remote estimation of photosynthesis and stomatal conductance in pine canopies. *Remote Sensing of Environ.* 63(1): 61–72.
- Elliot, D. 1993. A better activation function for artificial neural networks. ISR Technical Report TR 93–8. College Park, Md.: University of Maryland, Institute of Systems Research.
- Fritz, R. M., D. D. Malo, T. E. Schumacher, D. E. Clay, C. G. Carlson, M. M. Ellsbury, and K. J. Dalsted. 1999. Field comparison of two soil electrical conductivity measurement systems. In *Proc. 4th International Conference of Precision Agriculture*, 1211–1217. P. C. Robert, R. H. Rust, and W. E. Larson, eds. Madison, Wisc.: ASA–CSSA–SSSA.
- Ghinelli, B. M. G., and J. C. Bennett. 1998. Feasibility of employing artificial neural networks for emergent crop monitoring in SAR systems. *IEEE Proc. Radar, Sonar, and Navigation* 145(5): 291–296.
- Goh, A. T. C. 1995. Modeling soil correlations using neural networks. *J. Computing in Civil Eng.* 9(4): 275–278.
- Goldberg, D. E. 1989. *Genetic Algorithms in Search Optimization and Machine Learning*. Reading, Mass.: Addison-Wesley.
- Goldberg, D. E. 1999. The race, the hurdle, and the sweet spot: Lessons from genetic algorithms for the automation of design innovation and creativity. In *Evolutionary Design by Computers*, 105–118. P. Bentley, ed. San Mateo, Cal.: Morgan Kaufmann.
- Gopalapillai, S., and L. Tian. 1999. In-field variability detection and yield prediction in corn using digital aerial imaging. *Trans. ASAE* 42(6): 1911–1920.
- Green, A. A., M. Berman, P. Switzer, and M. D. Craig. 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geoscience and Remote Sensing* 26(1): 65–74.
- Harsanyi, J. C., and C. Chang. 1994. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. *IEEE Trans. Geoscience and Remote Sensing* 32(4): 779–785.
- Hsieh, P.-F., and D. Landgrebe. 1998a. Low-pass filtering for increasing class separability. In *Proc. 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5: 2691–2693. Piscataway, N.J.: IEEE.
- Hsieh, P.-F., and D. Landgrebe. 1998b. Statistics enhancement in hyperspectral data analysis using spectral-spatial labeling, the EM algorithm, and the leave-one-out covariance estimator. In *Proc. SPIE: Imaging Spectroscopy IV*, 83–190. San Diego, Cal.: SPIE.
- Karimi, H. 1998. Data acquisition through emerging high-resolution satellite imagery. *J. Computing in Civil Eng.* 12(3): 126–128.
- Kitchen, N. R., K. A. Sudduth, and S. T. Drummond. 2000. Characterizing soil chemical and physical properties influencing crop yield using soil electrical conductivity. In *Proc. 2nd International Geospatial Information in Agriculture and Forestry Conference*, vol. 1: 1122–131. Ann Arbor, Mich.: Veridian ERIM International.
- Lund, E. D., P. E. Colin, D. Christy, and P. E. Drummond. 1999. Applying soil electrical conductivity technology to precision agriculture. In *Proc. 4th International Conference of Precision Agriculture*, 1089–1100. P. C. Robert, R. H. Rust, and W. E. Larson, eds. Madison, Wisc.: ASA–CSSA–SSSA.
- Mukherjee, A. 1997. Self-organizing neural network for identification of natural models. *J. Computing in Civil Eng.* 11(1): 74–77.
- NCSA. 2000. Image to Knowledge (I2K)– Software Overview and Documentation. Available at: [www.ncsa.uiuc.edu/Divisions/DMV/ALG/activities/projects/i2k/documentation/index.html](http://www.ncsa.uiuc.edu/Divisions/DMV/ALG/activities/projects/i2k/documentation/index.html).
- Petrie, G. M., and P. G. Heasler. 1998. Optimal band selection strategies for hyperspectral data sets. In *Proc. 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 3: 1582–1584. Piscataway, N.J.: IEEE.
- Russel, S., and P. Norvig. 1995. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, N.J.: Prentice-Hall.
- Schowengerdt, R. A. 1997. *Remote Sensing: Models and Methods for Image Processing*. San Diego, Cal.: Academic Press.
- Smith, G. M., and E. J. Milton. 1999. The use of empirical line method to calibrate remotely sensed data to reflectance. *International J. Remote Sensing* 20(13): 2653–2662.
- Taubman, D. S., and M. W. Marcellin. 2000. *JPEG 2000: Image compression fundamentals, standards, and practice*. Boston, Mass.: Kluwer Academic Publishers, Kluwer International Series. Available at: [www.jpeg.org/JPEG2000.htm](http://www.jpeg.org/JPEG2000.htm).
- Thenkabail, P. S., R. B. Smith, and E. De Pauw. 2002. Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric Eng. and Remote Sensing* 68(6): 607–621.
- Tian, L., J. F. Reid, and J. Hummel. 1999. Development of a precision sprayer for site-specific weed management. *Trans. ASAE* 42(4): 893–900.
- Tu, T.-M. 2000. Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach. *Optical Eng.* 39(4): 897–906.

