Title: Optimal Hyperspectral Narrowbands for Discriminating Agricultural crops

Keywords: Hyperspectral data, Narrowbands, Agricultural crops, Stepwise discriminant analysis, Discriminant model, and Overall accuracy.

The authors is: Dr. Prasad S. Thenkabail, Center for Earth Observation (CEO), Yale University, P.O. Box 208109, New Haven, Connecticut 06511, USA.

Please address all correspondence to:

Dr. Prasad S. Thenkabail, Center for Earth Observation (CEO), Department of Geology and Geophysics, Kline Geology laboratory, P.O. Box 208109, 210 Whitney Avenue, Yale University, New Haven, Connecticut 06520-8109, USA. Fax: 203-432-3134, Tel.: 203-432-3440, e-mail: prasad.thenkabail@yale.edu

Optimal Hyperspectral Narrowbands for Discriminating Agricultural Crops

Prasad S. Thenkabail

Abstract

The main goal of this paper was to establish the best hyperspectral narrowbands for discriminating agricultural crops and to determine the accuracy with which such discrimination was possible. Six crops (wheat, barley, chickpea, lentil, vetch, and cumin) were studied. The best 12 narrowbands provided the most rapid increase in spectral discrimination. Further addition of narrowbands, only marginally increased discrimination capability reaching a plateau around 30 narrowbands. The overall accuracy (and K_{hat}) in separating the six crops increased rapidly from 73 percent (K_{hat}=71) when 6 best bands were used to 84 percent (K_{hat}=79) when 12 best bands were used. Peak overall accuracies of 94 percent (K_{hat}= 92) were achieved with about 30 narrowbands. Possibility of significant improvements by using mid infrared bands were indicated. Principal component derived hyperspectral narrowbands explained 61-92 percent variability in biomass and 70-87 percent variability in LAI. The best narrowbands to study agricultural crops have been recommended.

1.0 Introduction and Background

The new generation of sensors from the Earth Observation System (EOS), the New Millennium System (NMS), and the private industry systems (e.g., IKONOS, Orbview-4) provide data in many more spectral, spatial, and radiometric resolutions when compared with sensors of older generation such as the Landsat, the Le Systeme Probatoire d'Observstion de la Terre (SPOT), and the Indian Remote Sensing Satellite (IRS) series. This opens an exciting era of newer applications and improved information extractions. However, these advantages are often offset by multifold increases in demand on computing time, analyst time, storage volumes, and backup technologies. Newer sophisticated analytical techniques and software tools need to be evolved. Many projects are overwhelmed with significant amounts of “redundant” data leading to great strain on resources. Thereby, the need to determine optimal hyperspectral wavebands for every application is of great significance. This would help in avoiding handling “redundant” wavebands and/or help design of optimal waveband sensors.

Currently, we are in the very beginning of an evolution in spaceborne hyperspectral and hyperspatial sensors. The first spaceborne hyperspectral sensor, Hyperion, was recently launched onboard Earth Observing-1 (EO-1) by the National Aeronautics and Space Administration’s (NASA’s) New Millennium Program. The Hyperion gathers near- continuous data in 220 discrete narrowbands along the 400 to 2500 nanometer spectral range at 30 meter spatial resolution and in 12-bits. Each image is 7.5 kilometer in swath by 100 kilometer along track. The volume of data collected using Hyperion for an area equivalent to Landsat TM image area will increase by about 37 times. For the same area, another hyperspectral sensor, Warfighter-1, with 200 spectral bands and 8 meter spatial resolution onboard Orbview-4 will result in increase in data volume by 469 times. Such increases in data volume pose great challenges in data handling and manipulation requiring a more pragmatic approach of using selected optimal wavebands for given applications. However, currently there is little knowledge with regard to optimal (or redundant) wavebands for different applications.

The goal in this research will be to determine optimal hyperspectral narrow wavebands that best help model agricultural crop biophysical and yield characteristics in the visible and near-infrared portion of the spectrum. In the process, the redundant narrow wavebands will be determined. Recent research (Nolin and Dozier, 2000; Thenkabail et al. 2000; Blackburn 1998; Carter 1997; and Elvidge and Chen 1995) has shown that the narrow wavebands located in specific portions of the spectrum have the ability to provide required optimal information sought for a given application.

There is no single best approach to determine the optimal number of narrow wavebands required to provide best estimates of agricultural crop characteristics. In the past, researchers have used reflectance from individual narrow bands (Mariotti, 1996), various ratio indices (Aoki, 1981, Carter, 1994, Lichtenthaler 1996), derivatives of reflectance spectra (Elvide and Chen, 1995, Curran et al. 1991), or a combinations of these (Thenkabail et al., 2000), and linear mixture modeling approach (Mass, 2000; McGwire, et al. 2000). In this paper we will use 2 distinct statistical approaches to determine optimal wavebands. The two models are: (1) stepwise discriminant analysis (SDA) (Draper and Smith, 1981, Chapter 6), and (b) discriminant model (DM). Three crop separability indicators within stepwise discriminant analysis will be: (a) Wilks' Lambda, (b) Pillai trace, and (c) average squared canonical correlation. The discriminant model will classify each crop observation or parameter into one of the crop types leading to determination of classification accuracies for different numbers of input narrow or broad spectral wavebands. Both SDA and DM statistical tests lead to determining optimal wavebands.

The study will use data from 512 hyperspectral narrowbands in the ultra violet-visible-near infrared (UVNIR) portion of the spectrum for 6 agricultural crops, marginal lands, and fallow farms. Hyperspectral wavebands and indices will be tested for their ability to separate crop types and their ability to model biophysical characteristics. The results will be compared with multispectral Landsat broadbands. Hyperspectral wavebands providing best options for discriminating agricultural crop characteristics will be recommended.

2.0 Study area

The study area is located in the Syrian Arab Republic and is a representative benchmark area for the semi-arid desert margins of the Globe. The area was selected since it has both researcher and farmer managed farms (Figure 1a) with crops of high importance for the dry areas of the World. The area is characterized by Mediterranean climate: hot and dry summers and cool and wet winters. The long-term normal rainfall in the study area being 373 mm. This climate pattern occurs in several locations in the 30-40 degree latitude belts in both hemispheres. The largest single contiguous area experiencing this climate is in the Mediterranean basin covering parts of West Asia, North Africa, and Southern Europe. Many summer and winter crops in this climate are common with similar growing seasons and management practices. Hence a study conducted in such a representative area (Aleppo) has broader application across much of the other Mediterranean regions falling within similar climatic, vegetation, and agricultural patterns.

In the desert-margins of Southwest Asia (Figure 1a) agriculture faces more complex challenges than in areas with adequate rainfall. Over a billion people live in the desert margins with approximately 50 percent of the work force earning it's living directly from agriculture placing great stress on the sustainability of land and water resources. Worldwide an estimated one billion people currently live in countries and regions included in the desert-margins with the population growth rates of 2.1 percent in the Central Asian Republics and 3.6 percent in the Mediterranean regions.

Insert Figure 1a

3.0 Data characteristics

Spectral and ground truth data were gathered for 6 major crops: Barley (*Hordeum vulgare L.*; sample size 44), wheat (*Triticum aestivum L. or Triticum durum Desf.*; 64), lentil (*Lens esculenta Moench. Or Lens orientale (Boiss.) Schmalh. Or Lens culinaris Medikus*; 23), cumin (*Cuminum* *cyminum L.*; 17), chickpea (*Cicer arietinum L.*; 14), and vetch (*Vicia narbonensis L.*; 14). Measurements were also taken from marginal lands (20) and fallow farms or top soils (9).

3.1 Ground truth data

Ground truth data were collected from 205 locations the spectral characteristics (Figure 1b) that are influenced by crop phenology, background reflectance, tissue structure, health, vigor, and a host of other conditions (Asner et al. 2000). The data includes 176 locations of agricultural crops. The data included: GPS locations, crop samples to obtain quantitative characteristics such as LAI and biomass, observations of plant conditions and growth stages, canopy cover, digital photographs, and slide photographs. Sample locations were chosen randomly by driving around the study area and stopping for measurements at various locations. At each farm, an area of 30 m by 30 m that was considered representative portion of the farm was chosen for all measurements. Six main crops that occupy most of the cultivated area were identified for spectral and crop biophysical measurements. The spring (November-May) is the main cropping season in the study area. The ground truth data were collected during April-May 1998 when most crops were in critical, or tillering or late vegetative growth phases. In the laboratory, plant samples were analysed for leaf area (m^2), wet weight (kilograms), and in case of cotton for yields (kg lint/ha). Leaf area was obtained by running the leaves over a LI-COR 3100 leaf area meter. The leaf area obtained from one representative plant is multiplied by the number of plants in one m^2 area to obtain leaf area index (m^2/ m^2). Plants were cut and weighed on a simple weighing machine. This weight was multiplied by number of plants in one m^2 to obtain biomass (kg/ m^2). Leaf area index (LAI) and wet biomass (WBM), and plant height (PLNTHT) measurement were made for all 6 crops. There is a strong support for making measurements at the whole plant rather than leaf (Asner et al. 2000). Crop yield was obtained only for selected wheat farms through after harvest actual yield measurements (tonnes per hectare). The dried plants were crushed and assessed for plant crued protein (percentage) and Nirogen (percentage) for all crops and marginal lands. The mean nitrogen content (in percent) was: Vetch 3.24, lentil 2.7, wheat 1.66, barley 1.17, chickpea 3.01, cumin 3.13, and marginal lands 1.45.

3.2 Hyperspectral narrowband data

Hyperspectral data are acquired using a hand-held spectroradiometer of analytical spectral devices™ (Fieldspec, 1997) which gathers data in every 1.43 nanometers (nm) wide, 512 bands, discrete narrow wavelengths that includes a portion of ultraviolet, entire visible, and much of the near infrared portion (3318 to 1064 nm, exactly) of the spectrum. The spectroradiometer unit consisted of a main spectrometer, a personal computer, fiber optic cable, a pistol grip, and different field of view (FOV) cones. Inside the spectrometer instrument light is projected from the fiber optics onto a holographic diffraction grating where wavelength components are separated and reflected for independent collection by the detector(s) (FieldSpec, 1997). Each detector converts incident photons into electrons that are stored. At the read out time, the photoelectric current for each detector is converted to a voltage and is digitized by a 16-bit analog to digital (A/D) converter. This data is directly transferred to the computer main memory, which is in turn available for further processing, by the controlling software (FieldSpec, 1997). Gathering spectra at a given location involved optimizing the integration time (typically set at 17 milliseconds), providing foreoptic information, recording dark current, collecting white reference reflectance, and then obtaining target reflectance. The target reflectance is the ratio of energy reflected off the target (e.g., crops) to energy incident on the target (measured using a $BaSO_4$ white reference). Since the dark current varies with time and temperature it was gathered for each integration time (virtually for each new reading).

Data was gathered for the 6 desert-margin spring crops (barley, wheat, lentil, chickpea, cumin, vetch), marginal lands or grasslands of steppe, and other land cover types (e.g., fallows) of Southwest Asia. The mean spectral characteristics of these data are shown in Figure 1b. This spectrum is free of any atmospheric effects since data was measured at field level. At each site 10-20 reflectance measurements were consistently taken, along a transect, with a nadir view from a height of 1.2 meters for 6 crops, fallow farms, and grasslands using a 18 degree FOV. This resulted in viewing an area of 1134 cm^2.

Spectral data in 350-395 nm and 1010-1064 nm had significant noise problems. This was determined by observing hundreds of spectra individually. Also, from earlier studies (Broge and Leblancb, 2001, Thenkabail et al., 2000) it is clear that an optimal narrowband width is about 15 nm. Given these considerations, we decided to use data in the 395-1010 nm range and by aggregating data of 10 discrete intervals of 1.43 nm. This will give us a total of 43 narrow bands [(1010 nm - 395 nm) / (1.43*10)]. These 43 narrowbands are centered starting from 404 nm through 1004 nm (see Table I, for band centers of these narrowbands).

Insert Figure 1b

3.3 Landsat-5 TM broadband data

Broadband data can be derived by simulating narrowbands or from Landsat-5 TM image. Preliminary investigations showed the simulated broadband data provided significantly similar results as atmospherically corrected at-satellite exatmospheric reflectance based Landsat-5 TM broadband data in their relationships with agricultural crop variables. Thereby, it will suffice to report only one set of broadband data. In this paper broadband data derived from Landsat-5 TM sensor have been reported and will simply be referred to as "broadband" data.

The broadband data are corrected for atmospheric effects using simple blackbody subtraction technique and improved blackbody subtraction techniques or Chavez technique (Chavez, 1988, 1989; Milton, 1994), and the digital numbers are converted to radiance and at-satellite exotmospheric reflectance (Price, 1987) before being compared with narrowband data. A near-real-time Landsat-5 TM image of April 06, 1998 was acquired (image date corresponding with ground-measured spectra).

Digital values were extracted from 6 non-thermal bands of the Landsat-5 TM data from: (a) entire field; (b) a 3 by 3 pixel area within each field; and (c) 3 random pixels within each field. For the procedures to work the ground samples must adequately portray the conditions contributing to vegetation index response (Anderson et al. 1993). By averaging the vegetation index over a 9-pixel area, for example, a single ground sample had to capture the average condition of an 8100 m^2 region. The entire field area was between 3-6.5 hectares (30000 to 650000 m^2). A more desirable situation would be the comparison of the sample point to the vegetation index value obtained for a single pixel (900 m^2 area) (see Anderson et al. 1993). However, locating a precise pixel could be difficult. Hence 3 random pixels (2700 m^2) within each field were selected. These pixels were selected from areas within the field that were visually found to be representative of that field when the false color composites are displayed on a computer screen. Preliminary analysis showed that the 3-pixel procedure provided the better results. The digital numbers for each of its 6 non-thermal bands were converted to at-satellite exatmospheric reflectance's. A direct comparison is then made between the spectroradiometer measured narrow-band reflectance's (without atmospheric effects) with Landsat-5 TM measured broadband reflectance's (with atmospheric effects) as illustrated for wheat crop (Figure 1c) and for all the crops (Figure 1d). The satellite measured and ground-measured reflectances were then correlated with ecological variables.

Insert Figure 1c and Figure 1d

4.0 Methods

4.1 Stepwise discriminant analysis

The stepwise discriminant analysis (Draper and Smith, 1981, Chapter 6) was performed using PROC STEPDISC algorithm of SAS (SAS, 1999) through a stepwise selection (Klecka, 1980). The independent spectral waveband variables are chosen to enter or leave the model using: (a) significance level of F test analysis of covariance, where the variables already chosen act as covariates and the variable under consideration is the dependent variable, or (b) the squared partial correlation for predicting the variable under consideration from the CLASS variable, controlling for the effects of the variables selected for the model (SAS, 1999).

Stepwise selection begins with no variable in the model. At each step if a variable already in the model fails to meet the criterion to stay, the worst such variable is removed. Otherwise, the variable that contributes most to the discriminatory power of the model is entered. When all variables in the model meet the criterion to stay, and more of the other variables meet the criterion to enter, the stepwise selection process stops (SAS, 1999). The discriminatory power (or separability) between variables of different crops is measured by 3 indicators, which are multivariate measure of group differences over several variables:

(a) Wilks' Lambda,
(b) Pillai trace, and
(c) Average canonical correlation.

Determinants (variance) of the S matrices are found. Wilks' Lambda is the test statistic preferred for Multivariate analysis of variance (MANOVA), and is found through a ratio of the determininants.

$$\Lambda = \frac{|S_{error}|}{|S_{effect} + S_{error}|}$$

where, S is a matrix which is also known as: "sum-of-squares (SS) and cross-products," "cross-products," or "sum-of-products" matrices. Wilks' lambda and Pillai trace are based on the eigenvalues **Γ** of $\mathbf{A^*W^{-1}}$ where A is the among SS and cross-products matrix, and W the pooled SS and cross-products matrix:

Wilk's $\mathbf{\Lambda = \Pi\ 1/(1+\lambda_i)}$

Pillai's trace -= $\mathbf{\Sigma}\ \lambda_i / (1 + \lambda_i)$

Wilks' Lambda is the most commonly available and reported, however Pillai's criterion is more robust and therefore more appropriate when there are small or unequal sample sizes.

Canonical (latent) correlation is the correlation between a set of narrow waveband independent variables (IVs) and a set of dependent variables (DVs) of agricultural crops. Whereas multiple regressions are used for many-to-one relationships, canonical correlation is used for many-to-many relationships. Canonical correlation squared is the percent of variance in the dependent set explained by the independent set of variables along a given dimension. The general equations for performing a canonical correlation are relatively simple. First, a correlation matrix (R) is formed. This is composed of: correlations between DVs (R_{yy}), correlations between IVs (R_{xx}), and correlations between DVs and IVs (R_{xy}).

$$R = R^{-1}_{yy} R_{yx} R^{-1}_{xx} R_{xy}$$

For canonical analysis solve the above equation for eigenvalues and eigenvectors of the matrix R. Eigenvalues consolidate the variance of the matrix, redistributing the original variance into a few composite variants. Eigenvectors, transformed into coefficients, are used to combine the original variables into these composites. The eigenvalues are related to the canonical correlation by the following equation:

$$\lambda_i = r^2_{ci}$$

That is, each eigenvalue equals the squared canonical correlation for each pair of variants.

Normality is not required to perform Canonical Correlation. This is an advantage specially when the sample sizes are small. However, canonical Correlation is very sensitive to missing data in the analyzed matrix and to outliers. These issues are resolved before using canonical correlation.

4.2 Classification criterion or discriminant model

PROC DISCRIM (SAS, 1999) develops a discriminant model to **classify** each observation of a given crop variable into one of the groups. The discriminant model, also known as a classification criterion, is determined by a measure of generalized squared distance (Rao, 1973). The classification criterion can be based on either the individual within-group covariance matrices or the pooled covariance matrix; it also takes into account the prior probabilities of the groups. Each observation is placed in the class from which it has the smallest generalized squared distance. DISCRM can also compute the posterior probability of an observation belonging to each class.

Generalized Squared Distance (D) Function for crop types:

$$D^2_j(X) = (X-\bar{X}_j)' \, COV^{-1} \, (X-X_j)$$

Where, x is the observation, $\bar{X}$ is the mean of all observations for crop type j, and COV is for covariance.

Posterior Probability of Membership in each observation x of crop j is:

$$\overline{P}r(j|X) = \exp(-.5\, D_j^2(X)) / \mathrm{SUM}_k \exp(-.5\, D_k^2(X))$$

4.3 Principal Component Derived Vegetation Indices

Principal component analysis (PCA) was used to reduce the 43 wavebands hyperspectral data to a few bands that explain most of the variability. The original high dimensional data is thus transformed to few bands that contain most of the information in the original bands. PCA involves factor analysis using PROC PRINCOMP in SAS (SAS, 1999) in which a new factor (factor loading) is created for each variable in the data set.

In addition, normalized difference vegetation indices (NDVI) are computed from the first 2 principal components (PCA 1 and PCA 2) that explain about 90 percent variability in all the data. Each principal component involves factor loadings from all the wavebands.

NDVI of principal component 1 and 2 = NDVIPCA 1&2 = (PCA1BV-PCA2BV)/(PCA1BV+PCA2BV).

Where, PCA1BV and PCA1BV are the brightness values of principal component 1 and 2. New brightness values of PCA1BV and PCA2BV and so on can be calculated using the various wavebands and the coefficients for each waveband. For example: digital numbers of 6 TM bands for field 112 (barley crop) were: 57, 24, 24, 73, 52, and 18. The PCA1 coefficients were:-0.0564, 0.42323, 0.4455, 0.44708, 0.45723, and 0.45857. Thereby, the new brightness value, PCA1BV, for barley field number 112 will be: 82.3019. R^2 values are then calculated for relationships between NDVIPCA 1&2 with crop biophysical variables.

5.0 Results and discussions

5.1 Spectral and biophysical characteristics of crops

Agricultural crops were in the late vegetative or in critical growth phases. The mean canopy cover (in percentage) of crops was: vetch 88, lentil 90, wheat 97, barley 97, chickpea 69, cumin 48, and marginal lands 68. The sample sizes varied between 14 to 64. The mean spectral reflectance characteristics across the wavelength for these crops are shown in Figure 1b. At each sample location spectra were averaged from 10 to 20 measurements accounting for within and between field variability. For the same sample locations in Landsat-5 TM data, digital numbers were converted to radiance and at-satellite exo-atmospheric reflectance. The narrowband reflectance's were simulated and converted to corresponding broadband Landsat-5 TM reflectances. The two reflectances are plotted in Figure 1c (for wheat) and Figure 1d (for all crops). The plots show the affect of atmosphere in blue band (which is most sensitive to changes in atmospheric conditions). The blue broadband (TM1) has significantly higher reflectance compared to blue narrowband. Thereby, TM data were corrected for atmospheric effects by using improved blackbody subtraction techniques or Chavez technique (Chavez, 1988, 1989; Milton, 1994) before using them in this study.

The general characteristics of spectro-biophysical relationships of all 6 crops and the marginal lands are presented for atmospherically corrected broadband data with biophysical variables (Figure 2) and the corresponding narrowband data with biophysical variables (Figure 3). When the broadband index (NDVI23) is computed using TM green band (TM2) and TM red band (TM3) and related to LAI (Figure 2a) and wet biomass (Figure 2b) the R^2 values were 0.50 and 0.51, respectively. The corresponding narrowband index is computed using narrow bands with a width of 15 nanometers (nm) centered at 550 (nm) and 680 nm leading to an index of NDVI550680. The NDVI550680 has highly improved relationship with biophysical variables resulting in an R^2 value of 0.67 when related to LAI (Figure 3a) and WBM (Figure 3b). The commonly, used NDVI involving near-infrared and red bands provide much better relationships with LAI and biomass. The best narrowband NDVI involves a bands centered at NIR peak (920 nm) and absorption maxima (680 nm). The resulting index, NDVI920680, has significantly higher R^2 values with LAI ($R^2 = 0.71$; Figure 3c) and biomass ($R^2 = 0.74$; Figure 3d) compared to corresponding values of 0.59 (Figure 2c) and 0.60 (Figure 2d) with TM derived broadband NDVI43. These results indicate the potential of narrow bands when derived from specific wavelengths to provide greater sensitivity to changes in biophysical characteristics.

The discussions that follow focus on demonstrating the capabilities or limitations of various combinations of narrowbands or broadbands in separating (or discriminating) agricultural crop characteristics (or crop types). It needs to be noted that the narrowband data is acquired in the spectral region of the first 4 bands of Landsat TM broadbands.

Insert Figure 2a through 2d

Insert Figure 3a through 3d

5.2 Spectral separability of agricultural crops

Narrowband and broadband spectral data of wheat, barley, chickpea, lentil, vetch (hay), and cumin were used in stepwise discriminant analysis to determine crop separability (Table I). Three crop separability indicators were used: (1) Wilks' lambda, (b) Pillai trace, and (c) average squared canonical correlation.

Insert Table I here

The ability of spectral data to separate crop types was tested for various band combinations and is listed for best combinations of (Table I): (a) 4 or 6 broadbands, and (b) 4, 6, 12, 16, or 43 narrowbands. The best waveband combinations are determined by values of crop separability indicators such as, for example, Wilks’ lambda. The best 4 narrowbands (without mid-infrared bands) perform as well as the 6 broadbands (with mid-infrared bands) in separating 6 crop types (wheat, barley, chickpea, lentil, vetch, and cumin). By placing the 4 TM bands at specific narrowbands at 547 nm, 675 nm, 718 nm, and 904 nm highly significant increase in crop discriminatory power is seen compared to 4 TM broadbands centered at 485 nm, 560 nm, 665 nm, and 830 nm. For example, Pillai trace increases to 1.31 for the 4 narrowbands mentioned compared to 1.02 for the 4 TM broadbands (Table I). This indicates a clear advantage of using narrowbands from specific sensitive portions of the electromagnetic spectrum. For narrowbands, all three separability indicators change dramatically with increase in the number of wavebands till 12 or 16 bands (Figure 4a and 4b). For example, the wilks' lambda decreases dramatically with increase in number of wavebands from 4 to 6, and from 6 to 12 and to a much lesser extent when the number of bands increases from 12 to 16 (Figure 4a). The change becomes near asymptotic beyond 16 bands. The 6 crops will be perfectly separated if Wilks' lambda is zero. So the low decimals of Wilks' lambda indicate significant spectral separability between crops. The Pillai trace and the Cononical correlation provide similar results. Only a small incremental improvements (often statistically insignificant) in discriminatory power occurs with additional wavebands beyond 16 bands (Figure 2a and 2b; Table I). Wilks’ Lambda of 0.025 for 43 bands compared to 0.035 for 16 bands show such small incremental increase. However, addition of wavebands beyond 12 or 16, will require strong justification. Thereby, results indicate 12 to 16 narrowbands may be optimal in determining crop separability. These results confirm with earlier findings of Lelong et al. (1998) and Thenkabail et al. (2000).

Insert Figure 4a through 4d

The 4 narrowbands that best separate the 6 crops are (Table I): (a) 547 nm (green band center), (b) 675 nm (red absorption maxima), (c) 718 nm (red-edge center), and (d) 904 nm (near infrared peak). The 12 narrowbands that best separate the 6 crops are (Table I): (a) 489 nm (longer wavelength portion of blue band), (b) three green bands at 518 nm (point of maximum positive change in slope of spectra per unit change in wavelength), 547 nm (green band center), 575 nm (point of maximum negative change in slope of spectra per unit change in wavelength), (c) 3 red bands at 604 nm (beginning of red band), 661 nm (absorption pre-maxima), 675 nm (maximum absorption in red band), (d) 2 red-edge bands at 704 nm (beginning of red-edge when spectra dramatically begins to change from one of high absorption to one of high reflectance), 718 nm (center of the red-edge where the change in slope of spectra per unit change in wavelength in maximum in the visible and NIR), (e) 846 (center of the NIR shoulder), 904 nm (NIR peak), and (f) 975 nm (crop moisture absorption trough in NIR).

The ability of different combinations of narrowbands to better separate crop types spectrally when compared with broadbands is highlighted using bispectral plots (Figure 5 through 7). For example, wheat could not be separated from barley using two TM broadbands involving the red (TM3) and NIR (TM4) as shown in Figure 5c. When two narrowbands of red (λ_1 = 675 nm) and NIR (λ_1 = 904 nm) are used separability is highly significant (Figure 6c). Similarly, a red-edge band (λ_1 = 718 nm) and NIR peak band (λ_1 = 918 nm) provide highly significant separability between wheat and barley (Figure 7d). In each of these cases bandwidths ($\Delta\lambda_1$ and $\Delta\lambda_2$) were 14 or 15 nm wide. Similarly, cumin, lentil, and marginal lands had highly mixed spectral responses in TM blue band (TM1) and green band (TM2) (Figure 5a). When 2 narrow blue and green bands centered at 489 nm and 547 nm are used spectral separability becomes distinct (Figure 6a). Spectra of cumin, lentil, and wheat are highly mixed using green and red broadband combinations of TM2 and TM3 (Figure 5b). When narrower green and red band combinations of (λ_1 = 547 nm) and (λ_2 = 675 nm) (Figure 6b) and (λ_1 = 518 nm) and (λ_2 = 646 nm) (Figure 7b) are used spectral separability of cumin, lentil, and wheat becomes very clear. The advantage of the narrowbands is also in the numerous possibilities of searching for appropriate wavebands that provide best separability such as a green band centered at 447 nm and a NIR band centered at 761 nm (Figure 7a) that show clear discrimination between cumin, marginal lands and vetch. Such well-located wavebands are often missing in broadbands, often seriously limiting their use for specific application.

Literature indicates a high level of relevance of each of the wavebands found most sensitive in this study. The waveband centered at 489 nm is sensitive to carotenoid pigments (Blackburn, 1998; Tucker, 1977). Nichol et al. (2000) found 518 nm and 575 nm to be sensitive to pigment content and chloroplast. Strongly relationships with total chlorophyll were detected at 547 nm (Schepers et al. 1996). Waveband 661 nm and 675 nm varies significantly due to changes in factors such as biomass, LAI, soil background, cultivars types, canopy structure, nitrogen, moisture, and stress in plants (Elvidge and Chen, 1995; Carter, 1997; Blackburn, 1998, Thenkabail et al. 2000). Greatest crop-soil contrast is around 675 nm for most crops in most growing conditions (Thenkabail et al. 2000). Plant stress is best detected at red-edge bands (704 nm and 718 nm). These bands also provide additional information about chlorophyll and nitrogen status of plants (Carter, 1994, Clevers, 1999, Shaw et al. 1998, Elvidge and Chen, 1995). Strong correlation with total chlorophyll was detected at 846 nm (Schepers et al. 1996). Around 904 nm is the region of peak or maximum reflectance region of the NIR spectrum for certain types and/or growth stages of vegetation or crops (Thenkabail et al., 2000). Plant moisture sensitivity is best detected in the visible and NIR portion of the spectrum at 975 nm (Peñuelas et al. 1995, Thenkabail et. al. 2000). At this wavelength, direct measurements of water vapour in and over vegetation canopies is feasible (Richey et al. 1989).

For the 16 best bands, the 4 bands that add to the 12 mentioned above are: (a) 2 pigment sensitive blue bands centered at 418 nm and 461 nm, (Tucker, 1977) and (b) 2 NIR peak bands centered at 875 nm and 918 nm (Thenkabail et al. 2000).

Insert Figure 5a through 5d
Insert Figure 6a through 6d
Insert Figure 7a through 7d

5.3 Discriminant model for classification accuracies

The various best-broadband and best-narrowband combinations determined above are used to classify 6 crops into distinct groups using discriminant model. Accuracies of such a classification are established (Table II).

Insert Table II here

For broadbands, addition of 2 mid-infrared bands (TM5 and TM7) increase classification accuracies significantly compared to just using 4 VNIR bands (TM1, TM2, TM3, and TM4). The 6 narrowbands provide significantly better classification accuracies for 4 crops (wheat, lentil, vetch, and wheat) compared to 6 TM broad bands (Table II). However, for 2 other crops (chickpea, and cumin) 6 TM broad bands perform better than 6 narrow bands. This is as a result of the presence of MIR bands (TM5 and TM7) in TM, which are most useful when dealing with crops with complex conditions of soil background effects and mixed conditions of pigmentation of dry and green biomass (Thenkabail et al., 1995). Cumin and chickpea with 48 percent and 69 percent canopy covers are the crops with greatest soil background effects.

5.4 Classification accuracies and error matrices to determine optimal hyperspectral wavebands

The overall accuracy and K_{hat} increased rapidly for the best 12 bands after which the increase was in small increments (Figure 4c). An overall accuracy of 84 percent was achieved in discriminating 6 crop types when 12 narrowbands are used (Table III). The Cohen's Kappa or K_{hat} statistic allows for chance, and ranges from 0 in the case of the most confused classification to 1 in the case of the most accurate classification (Congalton 1988, 1991, Czaplewski, 1994, and Goodchild, 1994). A K_{hat} of 79 percent (Table III) for discriminating 6 crops is a very significant result. The increase in overall accuracy and K_{hat} from 12 to 16 bands is statistically insignificant (Figure 4c). This trend of small incremental increase in accuracies with addition of wavebands continues. Peak overall accuracies and K_{hat} of just over 90 percent can be achieved with about 30 narrowbands (Figure 4c). For all 43 narrowbands the overall accuracy is 94 percent with K_{hat} of 92 percent. Omission and commission errors decrease sharply for each crop as the number of wavebands increases, attaining less than 10 percent error in most cases when all 43 wavebands are used (Figure 4d).

Insert Table III here

The accuracy with which the individual crops are classified using various broad and narrowband combinations is summarized in Table II. Accuracies increase steeply up to 12 narrowbands after which the increase is by small increments (Table II and Figure 4c). For example, when 12 narrowbands are used accuracies go above 80 percent for 4 crops and above 60 for the other 2 crops (Table II): barley (84.1 percent), chickpea (85.7), cumin (100), lentil (65.2), vetch (64.3), and wheat (82.2). With addition of 4 bands, for 16 bands classification accuracies almost remain the same as that of 12 bands. Classification accuracies cross 90 percent for all crops except 1 (vetch with 85.7 percent) only when a large number of wavebands (43 bands) are added. The 43 bands take into account the entire visible and the near infrared region (395 to 1010 nm) with one band every 14 or 15 nonometers whereas the 12 bands are selected in the most sensitive portion of this wavelength, thus avoiding many redundant bands. Overall, for an increase of 10 percent accuracy, from 84 percent when 12 bands are used to 94 percent when all 43 bands are used, 31 additional wavebands are required.

The broadband results (Table II) indicate that addition of just two broadband TM mid-infrared (MIR) bands increase classification accuracies by 10 percent compared to the four TM VNIR bands. This is an encouraging sign that implies that addition of selected MIR bands in addition to VNIR bands increase crop discrimination and provide better classification accuracies. Based on these results, it is likely that some of the 12 best narrowbands may be added to and/or replaced by the MIR bands helping to increase overall classification accuracies.

Classification accuracy matrix is presented for classifying 6 crops using the best 12 bands through discriminant model (Table III). The error matrix includes errors of omission (producer's error) and errors of commission (user's error). Wheat fields are most accurately classified with low producer (2 percent) and user (11 percent) errors. This was followed by barley with omission and commission errors of 18 and 16 percent, respectively. Both wheat and barley had large sample size that helped account for within crop variability better and hence better spectral characterization. Commission errors occur when the classifier incorrectly commits pixels of the class being sought to other classes. A significant number of lentil fields are classified as chickpea, cumin, or vetch leading to large commission errors of 53 percent for lentil. Similarly, 21.43 percent of vetch fields are classified as barley leading to a commission error of 36 percent for vetch. Thereby, commission errors are indicative of spectral ability to discriminate within crop variability. The results indicate that within field variability of lentil and vetch crops is not fully accounted using 12 narrowbands. In contrast, omission errors occur when one class on the ground is misidentified as some other class by basic spectra and/or by classifier (Table III). Barley, chickpea, and lentil field with low biomass levels and/or with stress or senescing conditions also get classified as cumin leading to a high omission error of 41 percent for cumin. Thereby, an omission error is a measure of between crop (or class) discrimination. The results indicate that some of the barley, chickpea, and lentil fields get spectrally confused with cumin fields. The commission and omission errors can be reduced significantly (Figure 4d) using a large number of wavebands in the visible and the NIR. However, there is good likelihood that such errors may be reduced substantially by addition and/or substitution of few mid infrared bands helping us to avoid using a very large number of narrowbands.

5.5 Principal component analysis

The first 2 principal components explained: (a) 88 to 97 percent of all variability in broadband data, and (b) 86 to 95 percent of all variability in hyperspectral narrowband data (Table IV). Thereby, the procedure allows reducing many dimension of data into just 2 dimensions (bands) by retaining most of the variability seen in multidimensional data. Band reduction is done by computing the 2 new band values through Eigen value (or factor loadings). Information content, inherent in many dimensions (or bands) is retained in just 2 dimensions.

Insert Table IV here

For each crop, the first principal component (PCA1) of broadbands explained far greater variability in data when compared with PCA1 of the narrowbands (Table IV). This was because; the hyperspectral narrowbands contain far more complex and detailed information than the few broadbands of Landsat TM sensor. Each principal component represents unique nature of data that is highlighted by taking a normalized difference vegetation index of two principal components (NDVIPCA).

The NDVIPCA was related to various crop biophysical characteristics (Table IV). The broadband NDVIPCA explained 50-82 percent variability in wet biomass, and 56-74 percent variability in LAI. Using narrowband NDVIPCA, 61-92 percent variability in biomass and 70-87 percent variability in LAI. This is a highly significant improvement over broadband data. The results demonstrate the ability to model crop biophysical characteristics using only 2 principal components reduced from many narrowbands.

6.0 Conclusions

The study established the hyperspectral narrowbands that best discriminate agricultural crops. Six agricultural crops (wheat, barley, chickpea, lentil, vetch, and cumin) were studied using a 1.43 nanometer wide, 512-band spectroradiometer data. The stepwise discriminant analysis with 3 separability indicators (Wilks' lambda, Pillai trace, and average canonical correlation), Classification criterion or discriminant model, and Principal Component Derived Vegetation Indices were used to determine the crop discrimination power. The discrimination power increased rapidly with increase in number of wavebands up to best 12 narrowbands. Additional wavebands, beyond 12 bands, only increased crop discrimination by very small increments. For example, for six crops studied in this paper, values of Pillai's trace increased rapidly from 1.49 when 6 narrowbands were used to 1.97 when 12 narrowbands were used- an increase of 0.08 per addition of a waveband. Beyond 12 bands, Pillai trace increased only by very small increments reaching a value of 2.14 when all 43 wavebands are used- a increase of 0.028333 per addition of a waveband. Decrease in Wilks' lambda (Λ) values are indicative of increase in the ability of spectral data to discriminate crops (Λ = zero, for a perfect discrimination). For 6 crops studied in this paper, Λ values decreased rapidly from 0.1 when 6 narrowbands were used to 0.047 when 12 narrowbands were used- a decrease of 0.008833 per addition of a waveband. Beyond 12 bands, Λ decreases only by very small increments reaching a value of 0.025 when all 43 wavebands are used- a decrease of 0.00071 per addition of a waveband.

The overall accuracy (and K_{hat}) in separating 6 crops increased rapidly from 73 percent (K_{hat} = 71) for 6 best narrowbands to 84 percent (K_{hat} = 79) for 12 best narrowbands. The 12 narrowbands provide a high degree of accuracy consistently and can be considered as the best bands. The 12 waveband centers are: λ_1 = 489 nm, λ_2 = 518 nm, λ_3 = 547 nm, λ_4 = 575 nm, λ_5

= 604 nm, λ_6 = 661 nm, λ_7 = 675 nm, λ_8 = 704 nm, λ_9 = 718 nm, λ_{10} = 846 nm, λ_{11} = 904 nm, and λ_{12} = 975 nm. A nominal bandwidth ($\Delta\lambda$) of 15 nm can be used for all wavebands.

Further increase in wavebands increased accuracies in small increments reaching about 94 percent (K_{hat} = 92) when 30 narrowbands were used. Increase in accuracy when more than 30 bands were used was statistically insignificant. The spectral ability to discriminate within (commission) and between (omission) crop variability increased significantly with increase of number of narrowbands. For the 6 crops, omission errors decreased to a range of 0 to 27.3 percent when all 43 bands were used. In comparison, omission errors were 0 to 41.4 percent when 12 bands were used and 1.7 to 70 percent when 6 bands were used. Similarly, commission errors decreased to a range of 0 to 11.1 percent when all 43 bands were used when compared with 0 to 35.7 percent when 12 bands were used, and 7.8 to 57.1 percent when 6 bands were used. However, indications are that addition of few mid infrared bands as a substitute for and/or addition to the 12 narrowbands in the visible and NIR are likely to provide significantly improved results. The two mid-infrared (MIR) bands of Landsat TM (TM5 and TM7) provided increased sensitivity in handling crops with considerable soil background effects, thus helping to resolve these crops better when compared to using only the visible wavebands.

The study also highlighted the strength in reducing many dimensional narrowband data to few principal component analyzed (PCA) information rich bands. For example, 43 narrowbands were reduced to 2 PCA bands. The NDVI from the 2 of the so derived PCA bands take into account much of the variability in the 43 band hyperspectral data and hence explained 61-92 percent variability in agricultural crop biomass and 70-87 percent variability in LAI.

7.0 References

Anderson, G.L., Hanson, J.D., and Haas, R.H., (1993) Evaluating Landsat Thematic Mapper Derived Vegetation Indices for Estimating Above Ground Biomass on Semiarid Rangelands, *Remote Sensing of Environments*, 45:165-175.

Aoki, M., Yabuki, K. and Totsuka, T., (1981) An evaluation of chlorophyll content of leaves based on the spectral reflectivity in several plants. Res. Rep. Natl. Inst. Environ. Stud. Jpn. 66, pp. 125-130.

Asner, G.P., Wessman, C.A., Bateson, C.A., Privette, J.L. (2000) Impact of Tissue, Canopy, and Landscape Factors on the Hyperspectral Reflectance Variability of Arid Ecosystems. Remote Sensing of Environment. 74(1): 69-84.

Blackburn, G.A., (1998) Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves, INTERNATIONAL JOURNAL OF REMOTE SENSING, 19 (4): 657-675

Broge, N.H., Leblanch, E. (2001) Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sensing of Environment. 76(2): 156-172.

Congalton, R.G. (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37: 35-46.

Congalton, R.G. (1988) A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. Photogrammetric Engineering and Remote Sensing 54(5): 593-600.

Czaplewski, R.L. (1994) Variance Approximations for Assessments of Classification Accuracy. Research Paper RM-316, Fort Collins, Colorado: United States
Department of Agriculture, Forest Service.

Carter, G.A., (1994) Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. Int. J. Remote Sens. 15, pp. 697¯703.

Carter, G.A. (1997) Reflectance wavebands and indices for remote estimation of photosynthesis and stomatal conductance in pine canopies. *Remote Sensing of Environment*. 63:61-72.

Chavez, P.S., (1988) An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data, *Remote Sensing of Environment*, 24, 459-479.

Chavez, P.S., (1989) Radiometric calibration of Landsat thematic mapper multispectral images, *Photogrammetric Engineering and Remote Sensing*, 55, 1285-1294.

Clevers, J.G.P.W. (1999) The use of imaging spectrometry for agricultural applications. ISPRS Journal of Photogrammetry and Remote Sensing, 54(5-6):299-304.

Curran, P.J., Dungan, J.L., Macler, B.A. and Plummer, S.E., (1991) The effect of a red leaf pigment on the relationship between red-edge and chlorophyll concentration. Remote Sens. Environ. 35, pp. 69-75.

Draper, N.R., and Smith, H. (1981) Applied regression analysis, New York: Wiley.

Elvidge, C.D. and Chen, Z., (1995) Comparison of broad-band and narrow-band red and near-infrared vegetation indices. Remote Sens. Environ. 54, pp. 38-48.

FieldSPEC. (1997) User's Guide, manual release, Analytical Spectral Devices, Inc. Boulder, Colorado, USA.

Goodchild, M.F. (1994) Integrating GIS and remote sensing for vegetation analysis and modeling: Methodological issues. Journal of Vegetation Science 5:
615-626.

Klecka, W. R. (1980) Discriminant analysis. Quantitative applications in social sciences series, No. 19. Thousand oaks, CA: Sage Publications.

Lelong, C.C.D., Pinet, P.C., Poilve, H. (1998) A Case Study on Wheat in Beauce (France). Remote Sensing of Environment. 66(2): 179-191.

McGwire, K., Minor, T., and Fenstermaker, L. (2000) Hyperspectral Mixture Modeling for Quantifying Sparse Vegetation Cover in Arid Environments. Remote Sensing of Environment. 72(3): 360-374.

Lichtenthaler, H.K. (1987) Chlorophylls and carotenoids:pigments of photosynthetic membranes. *Methods in Enzymology*, 148:350-382.

Mariotti, M., Ercoli, L. and Masoni, A., (1996) Spectral properties of iron-deficient corn and sunflower leaves. Remote Sens. Environ. 58, pp. 282-288.

Milton, E.J., (1994) Teaching atmospheric correction using a spreadsheet. Photogrammetric Engineering and Remote Sensing, 60, 751-754.

Nichol, C.J., Huemmrich1, K.F., Black, T.A., Jarvis, P.G., Walthall, C.L., Grace, J., Hall, F.G., (2000) Remote sensing of photosynthetic-light-use efficiency of boreal forest, Agricultural and Forest meteorology, 101(2-3): 131-142.

Nolin, A.W., and Dozier, J., (2000) A Hyperspectral Method for Remotely Sensing the Grain Size of Snow. Remote Sensing of Environment, 74(2), pp. 207-216.

Price, J.C., (1987) Special Issue on Radiometric Calibration of Satellite Data, Remote Sensing of Environment, 22(1): 1-158.

Richery, J.E., Adams, J.B., and Victoria, R.L., (1989) Synoptic-Scale Hydrological and Biogeochemical Cycles in the Amazon River Basin: A Modelling and Remote Sensing Perspective, In Remote Sensing of Biosphere Functioning, (Editors: Hobbs R.J., and Mooney, H.A.), *Ecological Studies 79*, Springer-Verlag, New York. Pp.249-268.

SAS Institute, (1999) SAS/STAT User’s Guide and Software Release 6.12 Edition (Cary, North Carolina: SAS Institute Inc.).

Schepers, J.S., Blackmer, T.M., Wilhelm, W.W., and Resende, M., (1996) Transmittance and reflectance measurements of corn leaves from plants with different nitrogen and water supply, J. Plant Phys., 148:523-529.

Shaw, D.T., Malthus T.J., Kupiec, J.A. (1998) High-spectral resolution data for monitoring Scots pine (Pinus sylvestris L.) regeneration, International Journal of Remote Sensing, 19 (13): 2601 -- 2608.

Thenkabail, P.S., Smith, R.B., and De Pauw, E. (2000) Hyperspectral Vegetation Indices for determining agricultural crop characteristics, Remote Sensing of Environment, 71:158-182.

Thenkabail, P.S., Ward, A.D., and Lyon, J.G. (1995) Landsat-5 Thematic Mapper models of soybean and corn crop characteristics. *International Journal of Remote Sensing*. 15:49-61.

Tucker, C.J. (1977) Spectral estimation of grass canopy variables, *Remote Sensing of Environment*, 6:11-26.

Acknowledgements

The author would like to acknowledge financial support from the National Aeronautics and Space Administration (NASA) Earth Science Enterprise (grant number NAG5-3853). Thanks to Dr. Ronald B. Smith (Director), Dr. Roland Geerken (Scientist), and Mr. Laurent Bonneau (Lab Manager), of the Center for Earth Observation (CEO), Yale University for their support and for all the discussions. Excellent field work support was provided by the International Center for Agricultural Research in the Dry Areas (ICARDA). Many thanks are due to: Dr. Eddy de Pauw (Agroecologist), Prof. Dr. Adel El-Beltagy (Director General), Dr. John H. Dodds, (Assistant Director General, Research), Mr. Afif Dakermanji (research Manager), Mr. Pierre Hayak (crop scientist), Mr. A.F. Tarsha (farm Manager), and Mr. Samir Masri (soil scientist), and several others of ICARDA for all their support while collecting ground-truth data used in this paper.

Table I. Stepwise discriminant analysis to determine optimal wavebands. Hyperspectral and Landsat-5 TM data of 6 crops (barley, chickpea, cumin, lentil, vetch, and wheat) were used[*,¶].

Crops tested for Seperability	**Spectral Data type and number of wavebands used**	**Number of Wavebands**[#]	**Waveband Centers (nanometers)**	**Crop Separability Indicators**		
				Wilks' Lambda[*] (lesser the value greater the seperability between crops)	**Pillai Trace**[¶]. (greater the value greater the seperability between crops)	**Average Squ Cononical Correlation**[¶] (greater the va the seperabilit crops)
Barley, Chickpea, Cumin, Lentil, Vetch, and Wheat	**1. Landsat-5 TM broadbands**[5]	(a) 4 bands: 4 VNIR bands	485, 560, 665, 830, (TM1, TM2, TM3, TM4)	0.19	1.02	0.20
		(b) 6 bands: 4 VNIR and 2 MIR bands	485, 560, 665, 830, 1650, 2215 (TM1, TM2, TM3, TM4, TM5, TM7)	0.13	1.31	0.26
	2. Hyperspectral narrowbands[6]	(a) best 4 bands	hy547, hy675, hy718, hy904	0.13	1.31	0.26
		(b) best 6 bands	hy489, hy547, hy675, hy718, hy904, hy975	0.10	1.49	0.29
		(c) best 12 bands	hy489,hy518,hy547, hy575 hy604,hy661, hy675,hy704,hy718 hy846,hy904,hy975	0.047	1.97	0.39
		(d) best 16 bands	hy418,hy461, hy489, hy518,hy547,hy575 hy604,hy661,hy675, hy704,hy718 hy846, hy875, hy904,hy918, hy975	0.035	2.14	0.43
		(e) best 43 bands	hy404,hy418,hy432,hy447 hy461,hy475,hy489,hy504 hy518,hy532,hy547,hy561 hy575,hy589,hy604,hy618 hy632,hy646,hy661,hy675 hy689,hy704,hy718,hy732 hy746,hy761,hy775,hy789 hy804,hy818,hy832,hy846 hy861,hy875,hy889,hy904 hy918,hy932,hy946,hy961 hy975,hy989,hy1004	0.025	2.32	0.46

Note: [*] = Lower the value of Wilks' lambda better the seperability between different crops. Wilks' lambda is close to zero if groups are well seperated. Wilks' lambda here represents the discriminatory power of spectral data for 6 crops. Note that with addition of spectral wavebands Wilks' lambda is decreasing. and reaches near optimum for 16 bands (Wilks' lambda = 0.01495911). Further addition of spectral bands decreased Wilks' lambda only by insignificant margin.; [¶]**.** = Greater the value of Pillai's trace and greater the value of Average squared cononical correlation (ASCC) greater is the discriminantion between groups of data.; # = "hy" refers to hyperspectral (for example, hy404 refers to hyperspectral narrowband centered at 404 nonometers); +=significant at 99.9% probability; 5 = Landsat-5 TM of April 06, 1998; 6= hyperspectral data from hand-held spectroradiometer 43 narrowwavebands each of about 14.3 nanometer wide in 395 to 1010 nanometer spectrum range.

Table II. Classification accuracies using hyperspectral and Landsat-5 TM data. Discriminant model (of spring 1998 data) used to classify each observation into one of the groups[*,¶].

	Hyperspectral narrowwaveband classification accuracies (in percentage)				**Lan** **cl**
	Number of Hyperspectral Narrow bands =43	**Number of Hyperspectral Narrow bands =16**	**Number of Hyperspectral Narrow bands =12**	**Number of Hyperspectral Narrow bands =6**	**Nu** **Land** **Bro** **(4 VN** **MIR)**
Waveband center Or band name **Crop Type**	hy404 hy418 hy432 hy447 hy461 hy475 hy489 hy504 hy518 hy532 hy547 hy561 hy575 hy589 hy604 hy618 hy632 hy646 hy661 hy675 hy689 hy704 hy718 hy732 hy746 hy761 hy775 hy789 hy804 hy818 hy832 hy846 hy861 hy875 hy889 hy904 hy918 hy932 hy946 hy961 hy975 hy989 hy104	hy418 hy447 hy475 hy532 hy547 hy561 hy604 hy661 hy646 hy704 hy718 hy732 hy761 hy789 hy889 hy104	hy489 hy518 hy547 hy575 hy604 hy661, hy675 hy704, hy718 hy846 hy904 hy975	hy489 hy547 hy675 hy718 hy904 hy975	tm1 tr tm4 tr
1. Barley	89.0 (percent)	81.8	84.1	77.3	70.5
2. Chickpea	92.9	85.7	85.7	42.8	50.0
3. Cumin	94.1	94.4	100.0	64.7	94.1
4. Lentil	95.7	69.6	65.2	47.8	39.1
5. Vetch	92.3	64.3	64.3	50	35.7
6. Wheat	96.9	93.4	89.1	92.2	65.6

Note: ***= hyperspectral narrowwaveband data were obtained using a hand-held spectroradiometer. Band centers are indicated by, for example, "hy404" meaning hyperspectral narrowband centered at 404 nanometers. Band width for each narrowband was 14.3 nanometers. Hyperspectral data obtained in 395-1010 nanometer range was aggregated to 43 narrowbands each of 14.3 nanometers; ¶= broad band data was obtained from Landsat-5 TM overpass of April 06, 1998 (Path:174, Row: 035). Note that the mid-infrared (MIR) bands are available only for the broad bands.**

Table III. Classification matrix of spring 98 crop data using 12 Hyperspectral narrow bands. Discriminant model used to classify each observation into one of the groups.

```
Generalized Squared Distance Function:       Posterior Probability of Membership in each
CROPTY:
  2         _       -1   _                                    2                     2
 D (X) = (X-X )' COV  (X-X )                 Pr(j|X) = exp(-.5 D (X)) / SUM exp(-.5 D
(X))
  j         j            j                                     j        k           k

                        Number of Observations and Percent Classified into CROPTY:

 From CROPTY            ba          ch          cu          le          ve          wh
Total   errors of commission

      ba                37           0           5           0           1           1
44      16
                     84.09        0.00       11.36        0.00        2.27        2.27
100.00

      ch                 0          12           2           0           0           0
14      14
                      0.00       85.71       14.29        0.00        0.00        0.00
100.00

      cu                 0           0          17           0           0           0
17      0
                      0.00        0.00      100.00        0.00        0.00        0.00
100.00

      le                 0           3           3          15           2           0
23      53
                      0.00       13.04       13.04       65.22        8.70        0.00
100.00

      ve                 3           1           1           0           9           0
14      36
                     21.43        7.14        7.14        0.00       64.29        0.00
100.00

      wh                 5           1           1           0           0          57
64      11
                      7.81        1.56        1.56        0.00        0.00       89.06
100.00

      Total             45          17          29          15          12          58
176     84
      Percent        25.57        9.66       16.48        8.52        6.82       32.95
100.00

Errors of omission  18           29          41           0           25         2
```

$$K_{hat} = \left(N \sum_{i=1}^{R} X_{ii} - \sum_{i=1}^{r} X_{+i} * X_{i+}\right) / \left(N^2 - \sum_{i=1}^{r} X_{i+} * X_{+i}\right)$$

where, r is the number of rows in the matrix, X_{ii} is the number of observations in row i and column i, X_{i+} and X_{+i} are the marginal totals of i and column i respectively, and N is the total number of observations (Bishop et al. 1975).

Thereby,

K_{hat} = ((176)* (147) - (6936)) / ((176)^2 - (6936))

Where, (44*45) + (14*17) + (17*29) + (23*15) + (14*12) + (64*58) = 6936; **K_{hat} = 0.79**

Table VI. Principal component derived indices versus crop variables. Relationships of hyperspectral and multispectral PCA indices with crop variables using: (a) broad-band Landsat-5 TM data[*], and (b) spectroradiometer hyperspectral narrow-band data[¶].

Crop	Waveband	Percent variability in data explained by PCA1 and PCA2 and the wavebands bands highly influencing in explaining this variability					R^2 value when NDVI calculated using bands PCA1 and PCA2 are related with biophysical variables[@]:	
		PCA1[#] (percent)	Bands providing highest correlation (or factor loadings) for PCA1[+]	PCA2[#] (percent)	Bands providing highest correlation (or factor loadings) for PCA2[+]	PCA1 + PCA2[#]	WBM	LAI
1. Barley	Landsat-5 TM Broad-bands[*]	70	TM3, TM7, TM5, TM2, TM1	18	TM4	88	0.50	0.56
	Hyperspectral narrow-bands[¶]	68	HY532,HY718,HY518 HY547,HY561,HY575 HY432,HY447,HY732 HY589,HY704,HY461 HY418,HY504,HY932 HY489,HY475,HY918 HY604,HY904,HY889 HY875,HY975,HY989 HY861,HY746,HY846 HY618,HY832,HY946 HY818,HY804,HY789 HY761,HY961	24	HY675,HY661,HY646 HY689,HY632,HY618 HY604,HY489,HY475 HY504,HY404,HY461 HY704,HY589,HY447 HY432,HY418	92	0.78	0.77
2. Wheat	Landsat-5 TM Broad-bands[*]	77	TM3, TM2, TM1, TM7, TM5	16	TM4	93	0.66	0.59
	Hyperspectral narrow-bands[¶]	52	HY432,HY447,HY461 HY518,HY532,HY475 HY547,HY418,HY561 HY489,HY575,HY504 HY718,HY404,HY704 HY589,HY604,HY618 HY632,HY646,HY689 HY661,HY675,HY732	43	HY818,HY804,HY832 HY789,HY761,HY846 HY775,HY889,HY861 HY875,HY918,HY932 HY904,HY746,HY946 HY975,HY989,HY961 HY104,HY732	95	0.80	0.70
3. Lentil	Landsat-5 TM Broad-bands[*]	78	TM5, TM1, TM7, TM3, TM2	18	TM4	96	0.69	0.65
	Hyperspectral narrow-bands[¶]	50	HY775,HY789,HY761 HY804,HY818,HY832 HY846,HY746,HY861 HY875,HY889,HY918 HY904,HY732,HY932 HY989,HY975,HY104 HY718	40	HY575,HY518,HY561 HY447,HY432,HY461 HY532,HY504,HY547 HY489,HY475,HY704 HY589,HY418,HY718 HY604,HY618,HY632 HY689,HY646,HY404 HY989,HY661,HY675 HY975,HY104	90	0.75	0.79
4. Cumin	Landsat-5 TM Broad-bands[*]	79	TM2, TM3, TM5, TM1, TM7	15		94	0.49	0.62
	Hyperspectral narrow-bands[¶]	61	HY518,HY532,HY575 HY561,HY547,HY704 HY718,HY447,HY504 HY475,HY461,HY489 HY432,HY589,HY604 HY889,HY918,HY618 HY875,HY732,HY418 HY689,HY861,HY632 HY846,HY832,HY818 HY646,HY804,HY789	25	HY761,HY775,HY789 HY804,HY746,HY818 HY832,HY846,HY861 HY875,HY889,HY732 HY904,HY918,HY718	86	0.61	0.72

5. Chickpea	Landsat-5 TM Broad-bands[*]	77	TM1, TM5, TM2, TM3, TM7	20		97	0.82	0.74
	Hyperspectral narrow-bands[¶]	61	HY732,HY746,HY889 HY875,HY846,HY861 HY761,HY832,HY818 HY775,HY789,HY804 HY932,HY918,HY904 HY989,HY946,HY718 HY961,HY975,HY104 HY532,HY547,HY518 HY561,HY575,HY504 HY489	27	HY604,HY618,HY632 HY704,HY689,HY589 HY646,HY661,HY675 HY404,HY504,HY489 HY418,HY575,HY475 HY461,HY432,HY447 HY561	89	0.92	0.87
6. Vetch	Landsat-5 TM Broad-bands[*]	72	TM7, TM3, TM5, TM1, TM2	20		92	0.69	0.61
	Hyperspectral narrow-bands[¶]	56	HY718,HY432,HY547 HY532,HY518,HY447 HY461,HY418,HY561 HY475,HY575,HY489 HY504,HY704,HY732 HY589,HY604,HY618 HY404,HY632,HY689 HY646,HY746,HY818 HY832,HY804,HY846 HY761,HY789,HY775 HY861,HY889,HY875 HY661,HY904	26	HY932,HY918,HY946 HY975,HY904,HY104 HY989,HY961,HY889 HY875,HY861,HY846 HY761,HY775,HY818 HY789,HY832,HY804 HY746,HY732	92	0.80	0.70

Note:

* = based on digital values of Landsat-5 TM image dated April 06, 1998.

¶ = based on spectral reflectivity values from a hand-held spectroradiometer.

= variability in data explained by principal component 1. In case of broad-bands, Landsat-5 TM bands 1,2,3,4,5,and 7 of April 06, 1998 was used. In case of narrow-bands, hyperspectral narrow-bands obtained from a hand-held spectroradiometer were used. Hyperspectral data was obtained in 350-1050 nanometer waveband range in discrete intervals of 1.43 nanometers and aggregated into 43 narrow bands each of 14 or 15 nanometer s and centered between 404 and 1004. A high percent of variability in Landsat-5 TM data is explained in the first principal component itself . The complexity of hyperspectral data is greater due to large number of narrow-bands. This complexity in hyperspectral data results in relatively lower pecentage of data variability explained in the first component of hyperspectral data.

+ = factor loadings or bands providing highest correlation for a principal component are ranked and listed.

@ = NDVI of principal components 1 and 2 can be defined as: NDVIPCA 1&2 = (PCA1BV-PCA2BV)/(PCA1BV+PCA2BV). Where, PCA1BV stands for brightness value of principal component 1. Similarly, brightness value of principal component 2 is PCA2BV. New brightness values of PCA1BV and PCA2BV and so on can be calculated using the various wavebands and the coefficients for each waveband. For example: digital numbers of 6 TM bands for field 112 (barley crop) were: 57, 24, 24, 73, 52, and 18. The PCA1 coefficients were:-0.0564, 0.42323, 0.4455, 0.44708, 0.45723, and 0.45857. Thereby, the new brightness value, PCA1BV, for barley field number 112 will be: 82.3019. R^2 values were then calculated by relating NDVI's calculated from the principal component bands and biophysical variables.