SOIL CLASSIFICATION IN ARID LANDS WITH THEMATIC MAPPER DATA

Clasificación de Suelos en Zonas Aridas con Datos Tipo Thematic Mapper

Juan José Martínez-Ríos[1] and H. Curtis Monger[2]

SUMMARY

Soil is an essential part of any terrestrial ecosystem. Scientists, technicians, and farmers have studied its physical and chemical properties for many years for agriculture and soil conservation. These studies usually require field sampling and laboratory analysis that are time-consuming and destructive to the samples being analyzed. Remotely sensed data are an alternative that provide reliable information at low cost based on a non-destructive technique. The objective of this study was to evaluate the usefulness of Landsat Thematic Mapper data to classify soils in arid lands. To this end, a Thematic Mapper (TM) scene from the Chihuahuan Desert at Doña Ana County, New Mexico, mapped with the Soil Taxonomy System, was used. Furthermore, four remote sensing approaches were created to determine the best method to identify soil-mapping units. They were named **simple**, **technical**, **scaled**, and **complex**. The agreement of TM data and soils maps was tested using the error matrix approach in a supervised classification. Spectral signatures were selected by separability analysis applying the transformed divergency technique. The results revealed that the simple approach, based on thermal band discrimination, obtained classification accuracies of 70.67%, suggesting bands 2, 4, and 7 as the best for identifying soil mapping units. The technical approach, based on the principal components analysis technique, obtained accuracies of 66.86%, suggesting that data reduction is possible through this technique. The scaled approach, based on band ratios, achieved accuracies of 61.43%, suggesting ratios 1/5, 3/4, and 5/4 as the best transformations. The complex approach, based on indices, obtained accuracies of 28.50%, distinguishing SAVI, SVI, and albedo as the best data transformations. Based on its data reduction and its statistical accuracy, the technical approach was selected as the best method to classify soils at the study area. Because of its agreement with the soil taxonomy system, remotely sensed data are a meaningful alternative for detecting different soil types in arid environments.

Index words: *Remote sensing, supervised classification, spectral signature, Landsat.*

RESUMEN

El suelo es una parte esencial de cualquier ecosistema terrestre. Los edafólogos, técnicos y agricultores han estudiado sus propiedades por mucho tiempo. Estos estudios requieren muestreos de campo y análisis de laboratorio los cuales son lentos y destructivos. Los datos obtenidos mediante sensores remotos son una alternativa que provee de información fiable a bajo costo utilizando una técnica no destructiva. El objetivo de este trabajo fue evaluar la utilidad de los datos tipo Thematic Mapper (TM) para clasificar suelos en zonas áridas. Para este fin, una escena Landsat TM del Desierto Chihuahuense ubicada en el Condado de Doña Ana, Nuevo México en los EUA y un mapa de suelos del área, realizado con el sistema Soil Taxonomy, fueron utilizados. Se crearon cuatro métodos para determinar el mejor en la identificación de las unidades de suelo en la escena. Los métodos se llamaron: simple, técnico, escalar y complejo. El ajuste de los datos tipo TM y el mapa de suelos se evaluó utilizando la técnica de la matriz de error para una clasificación supervisada. Las firmas espectrales de los diferentes suelos fueron seleccionadas mediante un análisis de separabilidad, aplicando la técnica llamada **transformación divergente**. Los resultados mostraron que el método simple, basado en la discriminación de la banda térmica, obtuvo una precisión de 70.67%, mostrando a las bandas TM 2, 4 y 7 como las mejores para identificar las unidades suelo en el mapa. El método técnico, basado en el análisis de componentes principales, obtuvo una precisión de 66.86%, lo que

[1] Facultad de Agricultura y Zootecnia, Universidad Juárez del Estado de Durango, México. Apartado Postal 1-142, 35000 Gómez Palacio, Durango, México. (martinez@taipan.nmsu.edu)
[2] New Mexico State University, Dept. of Agronomy and Horticulture, BOX 3Q, Las Cruces, NM 88003, USA. (cmonger@nmsu.edu)

Recibido: Marzo de 2000. Aceptado: Febrero de 2002.
Publicado en Terra 20: 89-100.

sugiere que la reducción de datos es posible mediante esta técnica. El método escalar, basado en operaciones matemáticas entre bandas TM, logró una precisión de 61.43%, y señaló a los cocientes banda 1/banda 5, banda 3/banda 4 y banda 5/banda 4 como los mejores. El método complejo, basado en índices espectrales, logró una precisión de 28.50%, distinguiéndose SAVI (Soil Adjusted Vegetation Index), SVI (Simple Vegetation Index) y albedo como las mejores transformaciones para la identificación de los suelos. Basado en los datos obtenidos y por su capacidad para la reducción de datos, el método técnico se seleccionó como el mejor para la identificación de diferentes unidades de suelo bajo el sistema USDA Soil Taxonomy, siendo una alternativa para clasificar los suelos de las zonas áridas.

Palabras clave: *Sensores remotos, clasificación supervisada, firma espectral, Landsat.*

INTRODUCTION

The soil is an essential part of any terrestrial ecosystem defined as the product of interactions between parent material, biota, topography, and climate through time. As a result of human activities, the soil is also one of the most affected parts of the global ecosystem (Flechsig, *et al*., 1995; Rapaport, *et al*., 1995; Schlesinger, 1991).

Soil scientists, technicians and farmers have studied physical and chemical properties of soils for many years, using this knowledge in the construction of soil maps that exhibit areas with high agricultural potential as well as unstable areas caused by environmental contamination and poor land use planning.

The first attempt to classify soils using a systematic approach was conducted in Russia in the 1880s. This rudimentary system was based on the identification of soil properties like texture and color that would lead to the creation of new and improved methods like the FAO system and the Soil Taxonomy during the present century.

These classification systems have traditionally involved field sampling and laboratory analysis that are time consuming, labor intensive, and destructive to the samples being analyzed. Under this scenario, the use of remote sensing data, defined as the collection and interpretation of information about an object or feature from a distant point, is an alternative to classify soils.

Remote sensing data began in the 1840s as balloonists took photographs of the ground employing the newly invented photo-camera. Aerial photography was an important reconnaissance tool during the two World Wars and materialized as a valuable tool for earth monitoring with the inclusion of sensors on board of spacecraft like the Sputnik in 1957 and the Landsat, formerly ERTS (Earth Resources Technology Satellite), in the 1970s.

Although there will remain the need for field verifications to classify soils, remote sensing provides reliable information at low cost using a non-destructive technique. Furthermore, the arrival of new and improved space-borne sensors appears promising for earth monitoring studies. The goal of this study was to test the hypothesis that Landsat Thematic Mapper (TM) data can be useful in classifying soil types in arid and semi-arid regions. To accomplish this goal, a mapped region from Doña Ana County, New Mexico, USA, within the Chihuahuan Desert was selected and used as ground truth data. Four techniques of interpreting the remote sensing data on the TM scene were tested to determine the best method to classify soils in arid lands.

Remote Sensing

Remote sensing can be defined as the science of capturing information about an object using an instrument from a distance without physical contact with the object (Lillesand and Kiefer, 1994). The process for collecting such information includes many elements like (a) sources of energy; (b) transmission through the atmosphere; (c) energy interaction with the earth; (d) re-propagation through atmosphere; (e) sensing of energy by remote sensors; (f) data products in digital or pictorial format; (g) analysis and interpretation of data; (h) development or creation of products for users. When energy (electromagnetic radiation-EMR) is incident on a feature located on the earth surface, three energy processes (interactions) take place: (a) energy reflected; (b) energy absorbed; and (c) energy transmitted. According to the physical principle of conservation of energy, the mathematical relationship between those energy processes can be written as

$$E_I (\lambda) = E_R (\lambda) + E_A (\lambda) + E_T (\lambda)$$

Where: E_I represents the incident energy, E_R represents the reflected energy, E_A represents the absorbed energy, and E_T represents the transmitted energy, with all energy components as a function of wavelength λ. From the previous equation, the energy reflected (E_R) is given by

$$E_R (\lambda) = E_I (\lambda) – [E_A (\lambda) + E_T (\lambda)]$$

Which means that E_R is in equilibrium with the sum of E_A and E_T , subtracted from E_I.

According to Campbell (1996), remote sensing relies on observed spectral differences in the emitted or reflected energy from objects, features, or entities of interest. Based upon this knowledge, the spectral reflectance of an object (Pλ) can be calculated taking the energy of wavelength (λ) reflected from the object, divided by the energy of wavelength (λ) incident upon the object expressed as percentage.

$$P\lambda = E_R (\lambda) / E_I (\lambda)$$

Many researchers have constructed graphs relating the spectral reflectance values (Pλ) against the wavelength (λ), generating the plots called spectral reflectance curves (Robinson, *et al.*, 1995). Some regions of the electromagnetic radiation spectrum are more appropriate for distinguishing land features than other regions (Jensen, 1996). Figure 1 shows the typical reflectance curves for soils, water, and vegetation. It has been determined that the curve for healthy green vegetation has a “peak-and-valley” configuration, caused by the energy absorption by chlorophyll in the segment from 0.45 to 0.67 µm. It also shows a high reflectance in the range from about 0.7 to 1.3 µm generated by the internal structure of plant leaves. On the other hand, the soil curve in the same figure shows less peak and valley configuration due to many factors that affect its reflectance, like soil moisture, texture (clay content), and organic matter among others. The graph shows that the best region in the spectrum to study soils may be the region from 1.4 to 1.9 µm. Finally, regarding the water curve, the chart shows that the only segment useful for studying water is the region from 0.4 to 0.8 µm. Beyond that point, no information is available.

Some researchers (Stoner, *et al.*, 1980; Thompson, *et al.*, 1983; Huete, *et al.*, 1985; Coleman and Montgomery, 1987; Escadafal, *et al.*, 1989; Henderson, *et al.*, 1989; Ben-Dor and Banin, 1994) have conducted research on the usefulness of Thematic Mapper data to identify the best spectral region to analyze soil properties. Although most of these studies have been performed under lab conditions using hand held radiometers, few studies have been done using Landsat-5 Thematic Mapper data (Weismiller and Kaminsky, 1978; Satterwhite and Henley, 1987; Bhatti, *et al.*, 1991; Thenkabail, 1992; Van Deventer, 1992; Van Deventer, *et al.*, 1997).

To identify spectral characteristics of soils, there are several factors that must be considered, like soil roughness and soil texture, because it has been observed (Swain and Davis, 1978) that soil roughness decreases its reflectance, and soil texture, due to its relationship to water holding capacities, affects in several ways the relative reflectance values. Satterwhite and Henley (1987) pointed out that in arid and semi-arid ecosystems, the spectral response curves for soils, vegetation, and the association soil-vegetation are usually located in the 0.4 µm - 1.1 µm region. As a result, to discriminate between these surfaces is a difficult task.

METHODS AND MATERIALS

Regional Setting

The study area comprises 397.59 km^2 (39 759 ha) and is located in southern New Mexico, USA, in the northern part of Doña Ana County, near Las Cruces, NM (Figure 2). Its coordinates are 32.5338° N, 106.92° W (upper left); and 32.4334° N, 106.56° W (lower right). This area is located within the USDA Desert Soil-Geomorphology Project established in 1957 and finished in 1972 by the Soil Survey Investigations of the Soil Conservation Service (Hawley, 1975). The main objective of that project was to study the relationship between geomorphology and soils in arid and semiarid conditions to extrapolate this knowledge to other regions with similar geology and climatic conditions. This area is one of the most studied regions regarding geomorphology and soils in the United States (Hawley, 1975; Gile and Grossman, 1979; Monger, *et al*., 1998). Its extensive reports are used continuously in studies related to ecological research, range management, and desert stability. Gile and Grossman (1979) give a complete report on this area.

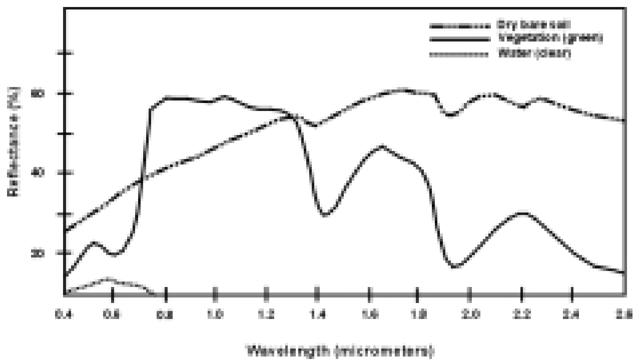

Figure 1. Typical spectral reflectance curves for soil, vegetation, and water (Lillesand and Kiefer, 1994).

Satellite Data

According to the Landsat World Reference System (WRS), the satellite image for the study area is located at Path 33, Row 37. Its ID is LT5033037009515610 and was obtained from the sensors on board of Landsat 5 on June 07, 1995. This scene was acquired from the Jornada LTER (Long Term Ecological Research) computer laboratory, located in the Biology Department at New Mexico State University. The image was previously rectified and georeferenced using the UTM grid system for zone 13 and the North American Datum of 1927 (NAD27). The USGS (United States Geological Survey) EROS Data Center previously resampled the image to 30 m per pixel using the cubic convolution technique. The software used to perform this task was the NLAPS (National Landsat Archive Production System).

Characteristics of the Landsat TM Data

The Thematic Mapper data has a swath width of approximately 185 km from an altitude of 705 km (Jensen, 1996). Its detectors record the electromagnetic radiation (EMR) in seven bands that are: Band 1 (Blue, 0.45-0.52 µm wavelength), Band 2 (Green, 0.52-0.60 µm), Band 3 (Red, 0.63-0.69 µm), Band 4 (Reflective-infrared, 0.76-0.90 µm), Band 5 (Mid-infrared, 1.55-1.74 µm), Band 6 (Thermal-infrared, 10.40-12.50 µm), and Band 7 (Mid-infrared, 2.08-2.35 µm). The spatial resolution of a Thematic Mapper image is 30 m x 30 m for Bands 1 to 5 as well as Band 7; the Band 6 has a spatial resolution of 120 m x 120 m (Lillesand and Kiefer, 1994). The Landsat platforms operate following a sun-synchronous, near polar orbit. Its radiometric resolution is 8 bit, and its temporal resolution is 16 days, with an image overlap that varies from 7 percent at the Equator to nearly 84 percent at 81° North or South latitude (Lauer, *et al*., 1997).

Data Input

A soil map of the study area made by Gile, *et al*. (1981), from their study on soils and geomorphology in the basin and range area of southern New Mexico, was digitized using ARC/INFO®. After this process, a vector layer was obtained with a final root mean square error (RMSE) of 0.02, ensuring an accurate registration between the reference soils map and the Thematic Mapper scene.

After editing the attribute table of the vector layer, using the Soil Taxonomy system, the layer was converted to raster format (grid) assigning a cell size of 30 m, corresponding to that of the Thematic Mapper data. This task was performed using ESRI's Spatial Analyst® software. This final map was the ground truth map in digital form, which would be compared to the computer-classified map for the accuracy assessment analysis.

Data Processing

Supervised Classification

Digital image classification can be defined as the procedure assigning individual pixels in an image to different classes or categories (Campbell, 1996). The result is a thematic scene that exhibits the original data as categorical information. The supervised classification is a user-controlled technique in which pixels are evaluated and assigned to a class previously recognized from collateral sources, such as ground truth, maps, and aerial photographs (Mausel, *et al.*, 1990; Smith, *et al.*, 1990; Jensen, 1996). Supervised classification requires that the user select training areas for use as a basis for the classification process.

Because a soils map for the calibration area was available (ground truth), the supervised classification technique was used in this study. Studies from Congalton (1991), Knick, *et al.* (1997), and Lo and Watson (1998) report that this technique allows separation between several spectral classes with reliable precision for scene classification.

Because the supervised classification method requires the selection of training sites, and because the

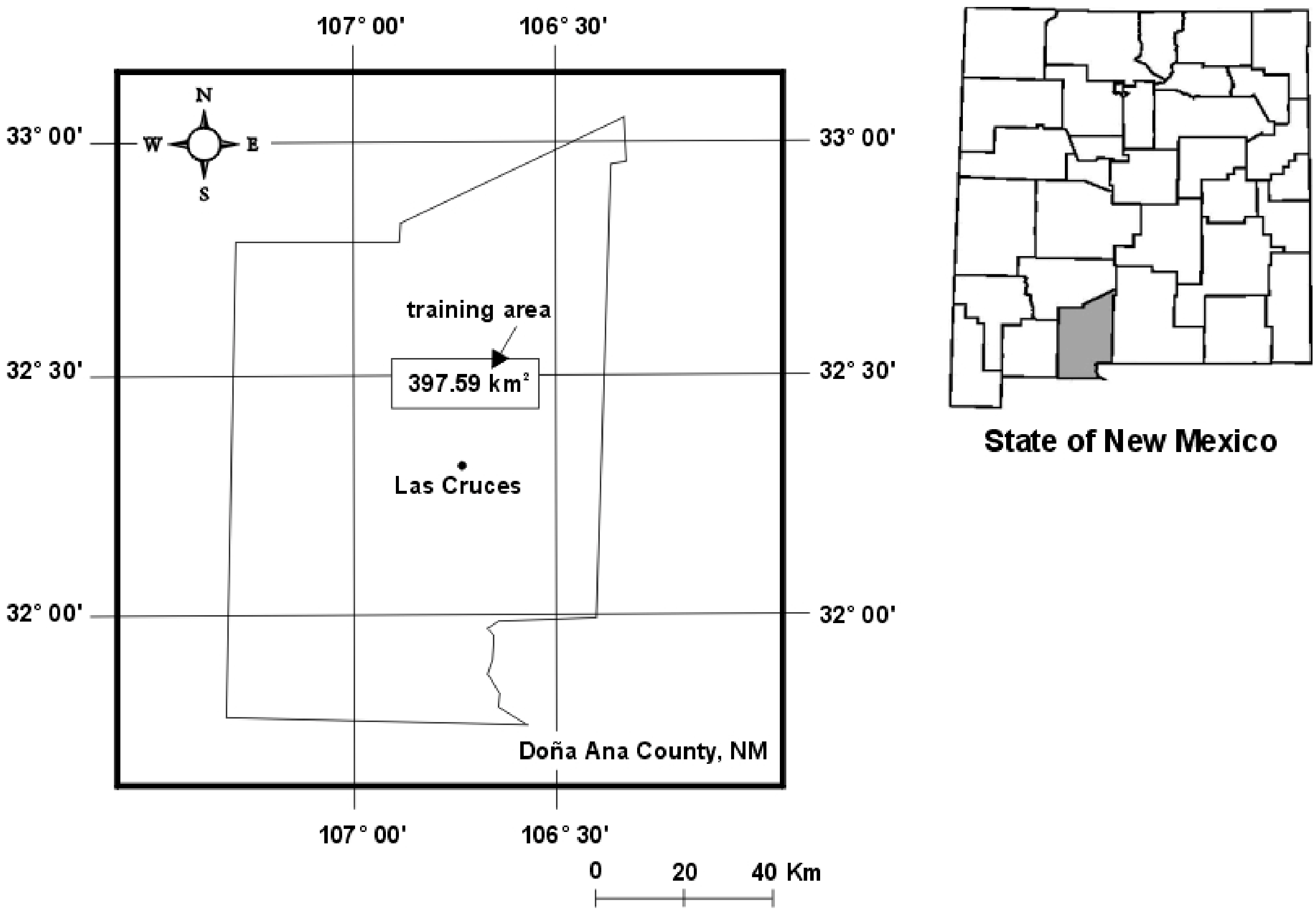

Figure 2. Localization of the study area at Doña Ana County, New Mexico, USA.

success of a good classification process is based on how correctly the spectral signatures were identified (Hepner, 1990), four approaches were created and used in this study. They were called simple, technical, scaled, and complex approach (note: these names were chosen for the specific purposes of this study, meaning nothing in remote sensing literature).

Simple approach. In this process, the original seven bands Thematic Mapper scene for the calibration area was reduced to six bands by eliminating the thermal band (Band 6) because it was considered as 'irrelevant' in this particular study. After this process, the selection of training sites was performed on the remaining six bands.

Technical approach. This approach was based on the concept of Principal Components Analysis, also called PCA, or Karhunen-Loeve analysis (Jensen, 1996). This statistical procedure has demonstrated to be an excellent tool in the analysis of multispectral remotely sensed data (Morse, *et al.*, 1990; Rees, 1990; Lillesand and Kiefer, 1994; Wilcox, *et al.*, 1994). The transformations to the original Thematic Mapper data using PCA reduce data redundancy and can improve interpretation because the PCA bands are independent and non-correlated (ERDAS, 1995). Additionally, this technique reduces contributions of noise and error (Campbell, 1996). Similarly, PCA can be used to reduce the information included in the raw data (seven bands) into two or three bands without losing significant information.

It has been reported that with Landsat images of land surfaces the three first components retain over 96 percent of the total sample variance (Gong and Howarth, 1992; Wilcox, *et al.*, 1994; Jensen, 1996). In this approach, a three-band composite image of principal component (PC) bands was generated from the original Thematic Mapper (TM) data as follows: Band 1 was the first PC of the three TM visible bands (1-3); Band 2 was the raw TM near-IR band (4); and Band 3 was the first PC of the two TM mid-IR bands (5, 7). Using this technique, the classification process can benefit because: (1) it reduces the total number of bands from 7 to 3 being more affordable to process; and (2) it retains more than 96 percent of the total variance in the original data set, and preserves the ability to interpret the inherent differences between the visible, near-IR, and mid-IR bands.

Scaled approach. This approach was based on band ratios, which is a procedure accomplished by the simple arithmetical division of one band by another in the Thematic Mapper data. Some studies (Satterwhite and Henley, 1987; Frazier, 1989; Frazier and Cheng, 1989; Bauer, *et al.*, 1994; Stella and Hoffer, 1998) indicate that to determine the spectral differences between soil attributes, the use of band ratios and indices is advisable, since they help to normalize differences in solar radiation due to differences in sun angle, light intensity, and atmospheric disturbances.

Researchers like Van Deventer, *et al.* (1997) state that spectral differences between soil attributes in Thematic Mapper data can be acquired by the development of indices, and declare "...indices include original brightness values as well transformations such as band differences, band ratios, and normalized differences of two or more bands..."

It has also been found (Wilcox, *et al*., 1994) that some of Thematic Mapper band ratios may help to discriminate between bare soil, eroded soil and soil organic carbon levels.

With the objective to determine the behavior of soil spectral curves, six band ratios were created using the Thematic Mapper data. They were 1/4, 1/5, 3/4, 5/4, 5/3, and 5/7. These ratios were scaled from 0 to 255 to take advantage of the radiometric resolution of the data, improving the contrast in the image by making full use of the 256 gray levels available in the display system. These ratios were assigned to the image created from Band 1 (ratio 1/4) through Band 6 (ratio 5/7).

Complex approach. In this particular approach, and in order to quantify the spectral differences in the study area, some known transformations were used in this study. These transformations included: NDVI (Normalized Difference Vegetation Index), SAVI (Soil Adjusted Vegetation Index), NDTI (Normal Difference Tillage Index), SVI (Simple Vegetation Index), and Albedo. These transformations were calculated as follows:

NDVI = (TM band 4 – TM band 3) / TM band 4 + TM band 3)
SAVI = [(TM band 4 – TM band 3) / TM band 4 + TM band 3 + 0.7)] * (1 + 0.7)
NDTI = (TM band 5 – TM band 7) / (TM band 5 + TM band 7)
SVI = TM band 4 / TM band 3
Albedo = (0.322 * TM band 3) + (0.725 * TM band 4)

Selection of the Training Sites

To obtain the training sites in every approach, a digitization on the image was performed to determine the AOIs (areas of interest), this process was performed using the image processing software ERDAS Imagine® ver 8.2.

The AOIs are defined as small polygons or areas defined by the analyst on a multispectral image as representative regions. To correctly identify the training sites, the soil ARC/INFO® polygon coverage was overlaid on the image. After this process, several training sites were selected based on their spectral response on the image. They were chosen to be representative and homogeneous of the spectral response of known soil types. Table 1 shows the soil spectral signatures selected for every approach and their respective soil classification according to the USDA Soil Taxonomy.

Separability Analysis

According to Swain and Davis (1978), there are several statistical methods to determine the signature separability for classification purposes, they are: Divergency, Transformed divergency, and Jeffreys-Matusita distance. Additionally, Jensen (1996) also includes the Bhattacharyya distance as another separability estimator. However, when working with several signatures (classes), the Jeffreys-Matusita and Transformed divergency have demonstrated to be better techniques (Mausel, *et al.*, 1990).

Because of its availability in the image processing software, the Transformed divergency (Td) technique was used to find the separation between the soil spectral signatures. This technique was based on the criteria proposed by Jensen (1996) who states that a value of 2000 may be considered as excellent 'between-class' separation, while values above 1900 can be considered good separability. On the other hand, Td values below 1700 suggest poor separability. According to ERDAS (1995), Transformed divergency (Td) is calculated using the following equations:

$$Td_{ij} = 2000\left(1 - e^{-D_{ij}/8}\right)$$

Where: i and j = the two signatures (classes) being compared

Table 1. Spectral signatures from soil units selected for the separability analysis in the study.

Spectral signature	Soil classification according to Soil Survey Staff, Soil Taxonomy (1994)
13MC †, ‡, §, ¶	Mixed, Typic Torripsamments
13Y †, §, ¶	Sandy, mixed, Typic Torripsamments
10L †, ‡	Loamy-carbonatic, shallow, Typic Paleorthid
11R †, §	Loamy-skeletal, mixed, Typic Calciorthid
16MB †	Coarse-loamy, mixed, Typic Calciorthid
13MB †, §, ¶	Coarse-loamy, Typic Torriorthent
14V ‡, §, ¶	Fine-loamy, mixed, Typic Haplargid
16VG †, §	Coarse-loamy, Typic Haplargid
12V †, §	Loamy-skeletal, mixed, shallow, Ustollic Paleargid
13V †	Coarse-loamy, Typic Torriorthent
15M †, ‡, §, ¶	Fine-loamy, mixed, Typic Haplargid
13L †	Fine-silty, mixed Typic Torrifluvent
13MM †, ¶	Coarse-loamy, Typic Haplargid
13MO ‡, §	Coarse-loamy, mixed, Thermic Aridic Haplustoll
13LG ‡	Coarse-loamy, Typic Torriorthent
10LL †, ‡, §, ¶	Loamy, shallow, Typic Paleorthid
16LS †	Fine-loamy, Typic Haplargid
10V ‡, ¶	Loamy, mixed, shallow, Typic Paleorthid
51S ‡, ¶	Fine-silty, mixed, Ustollic Calciorthid
16L ‡, §, ¶	Fine-loamy, mixed, Ustollic Haplargid
11L ‡	Coarse-loamy, carbonatic, Typic Calciorthid
16V §	Coarse-loamy, carbonatic, Typic Calciorthid
55S †	Fine, mixed, Ustollic Haplargid (overflow)

† Selected in the simple approach. ‡ Selected in the technical approach.
§ Selected in the scaled approach. ¶ Selected in the complex approach.

D_{ij} = divergence between two signatures.

Divergency (D) can be calculated by:

$$D_{ij} = \frac{1}{2}\,\mathrm{tr}\left[\left(C_i - C_j\right)\left(C_j^{-1} - C_j^{-1}\right)\right] +$$
$$\frac{1}{2}\,\mathrm{tr}\left[\left(C_i^{-1} - C_j^{-1}\right)\left(\mu_i - \mu_j\right)\left(\mu_i - \mu_j\right)^T\right]$$

Where: i and j = the two signatures (classes) being compared
C_1 = the covariance matrix of signature i
μ_i = the mean vector of signature i
tr = the trace function (matrix algebra)
T = the transposition function

Accuracy Assessment

To test the accuracy of the computer-classified maps in every approach, the error matrix technique (confusion matrix) was used. The confusion matrix is the simplest descriptive statistic used to compare a classification result with ground truth information. "...This accuracy measure indicates the probability of a reference pixel being correctly classified and is really a measure of omission error..." Congalton (1991). In addition to the error matrix approach to test the classification accuracy, the computation of Kappa coefficient of agreement was performed. The Kappa coefficient is a discrete multivariate measure developed by J. Cohen in 1960 (Lo and Watson, 1998) and is used widely by remote sensing scientists (Congalton and Mead, 1983; Congalton, *et al*., 1983; Congalton and Green, 1999). According to Congalton (1991), to perform the Kappa coefficient analysis, the KHAT statistic (an estimate of Kappa) must be calculated. Congalton and Green (1999) express that the KHAT value measures how adequately the computer-aided classification coincides with the data of reference (ground truth), and state that this estimator can be calculated with the following equation:

$$KHAT = \frac{n\sum_{i=1}^{k} n_{ii} - \sum_{i=1}^{k}\left(n_{i+}\cdot n_{+i}\right)}{n^2 - \sum_{i=1}^{k}\left(n_{i+}\cdot n_{+i}\right)}$$

Where: k = the number of rows in the matrix
n_{ii} = observation in row *i* and column *i*
n_{i+} and n_{+i} = the marginal totals of row *i* and column *i*, respectively
n = the total number of observations

RESULTS

Simple Approach

After performing the statistical comparison of reflectance values for the training sites in the six-band TM image, the transformed Divergency technique (Td) established, in general, an excellent statistical separation between most of the signatures collected. The separability analysis also established that the best separation was found when using Bands 2, 4, and 7, having an average value of 1974 and a minimum of 1303. Table 2 shows the results of the accuracy assessment analysis performed with the 15 classes collected in the supervised classification. The highest accuracy was obtained with signature 12V (Ustollic

Table 2. Errors of inclusion and exclusion in the simple approach after the supervised classification.

Class	Commission	Omission	Accuracy
	- - - - - - - - - - - % - - - - - - - - - - -		
15M	13.286	18.960	81.040
16VG	29.435	39.258	60.742
13MB	63.727	49.963	50.037
13MC	22.357	16.882	83.118
13Y	33.553	55.981	44.019
11R	12.433	21.821	78.179
10L	38.108	30.457	69.543
12V	9.257	10.344	89.656
16MB	29.802	29.051	70.949
13V	22.856	46.053	53.947
13L	104.040	13.484	86.516
13MM	11.943	41.683	58.317
10LL	31.075	20.557	79.443
16LS	40.559	15.526	84.474
55S	9.862	29.554	70.446

Paleargid) whereas the lowest accuracy was obtained with 13Y (Typic Torripsamment). The error matrix showed an overall accuracy of 70.67% and a KHAT (Kappa estimator) value of 0.682, which according to Congalton and Green (1999), represents a moderate agreement between the ground truth and the classified classes.

Technical Approach

Principal Components Analysis

Visible bands. After performing the principal components analysis (PCA) on the three visible bands of the Thematic Mapper scene, the analysis showed that the first principal component (PC1) accounted for 93.4% of the total variance in the three bands. Regarding the second and third components, they accounted for 6.37% and 0.21% of the total variance, respectively. This means that it is possible to compress the information provided in the three visible bands onto the first component (PC1) without losing any substantial information.

Bands 5 and 7. Once the principal component analysis was performed on Bands 5 and 7 of the original TM scene, a strong correlation between these bands was established. The analysis showed that the first principal component (PC1) accounted for 99.24% of the total variance for those bands, whereas the second component (PC2) accounted for just 0.75% of the total variance, meaning that PC1 can be used in place of Bands 5 and 7 without losing any substantial information.

Separability analysis. After collecting 11 classes for separability analysis in the technical approach, the Td technique, applied taking three bands at a time, established a best minimum separability average of 1844, which was considered a good separation between signatures, having just four combinations out of 56 (7%) values lower than 1700.

Accuracy assessment. Table 3 shows the results of the accuracy assessment analysis performed using 11 classes in the supervised classification. The error matrix shows an overall accuracy of 66.86% and a KHAT (Kappa estimator) coefficient value of 0.6239, which, according to the criteria proposed by Congalton and Green (1999), expresses a moderate agreement between the ground truth and the classified classes. Moreover, Table 4 shows that the highest accuracy using this technique was obtained with signature 10V (Typic Paleorthid) with a value of 100%. The second best accuracy was obtained by

Table 3. Error matrix for the supervised classification in the technical approach (the horizontal axis is the ground truth, whereas the classified data is shown in the vertical axis).

	13MC	14V	10V	10L	51S	16L	15M	11L	13LG	10LL	13MO	Total
13MC	2952	0	0	0	0	8	105	34	0	0	13	3112
14V	0	4449	0	0	0	0	0	0	14	43	32	4538
10V	0	0	6732	0	0	0	0	0	0	0	0	6732
10L	0	0	0	4835	6	1	0	0	621	622	4	6089
51S	0	0	0	8	6239	8552	487	0	214	0	0	15500
16L	0	0	0	709	4331	10871	763	840	5178	1	61	22754
15M	546	0	0	0	114	186	3648	231	0	0	329	5054
11L	19	0	0	57	511	925	188	9775	22	7	506	12010
13LG	0	0	0	59	713	82	2	58	402	66	2351	3733
10LL	0	0	0	958	75	7	0	3	649	10342	131	12165
13MO	28	28	0	35	50	2	1	3	690	3376	11704	15917
Total	3545	4477	6732	6661	12039	20634	5194	10944	7790	14457	15131	107604

Overall classification accuracy: 66.86%. KHAT (Kappa) coefficient: 0.6239.

Table 4. Errors of inclusion and exclusion for the technical approach after the supervised classification.

Class	Commission	Omission	Accuracy
	- - - - - - - - - -	%	- - - - - - - - - -
13MC	4.513	16.728	83.272
14V	1.988	0.625	99.375
10V	0.000	0.000	100.000
10L	18.826	27.413	72.587
51S	76.925	48.177	51.823
16L	57.589	47.315	52.685
15M	27.070	29.765	70.235
11L	20.422	10.682	89.318
13LG	42.760	98.840	5.160
10LL	12.610	28.464	71.536
13MO	27.844	22.649	77.351

signature 14V (Typic Haplargid) with 99.4%, presenting an omission value of 0.625%, and a commission value of 1.988%. In contrast, the lowest accuracy was obtained by signature 13LG (Typic Torriorthent) showing an accuracy of 5.16% (94.84% omission and 42.76% commission). Likewise, low accuracies were also obtained by signatures 51S (Ustollic Calciorthid) and 16L (Ustollic Haplargid) with 51.8% and 52.6%, respectively, suggesting that a good separation for those signatures may not be clearly achieved through this procedure.

Scaled Approach

Separability analysis. Once the separability analysis was performed for the 12 classes collected, the Td technique established an excellent statistical separation between most of the 12 signatures collected. This process, performed utilizing three bands at a time, just found a poor separability between the signatures 13Y (Typic Torripsamment) and 16V (Typic Haplargid) with a value of 1656, no other poor separability was found in the analysis. Additionally, the separability analysis established that the best minimum separation between signatures was obtained when using Bands 2, 3, and 4, that is, ratios 1/5, 3/4, and 5/4, respectively, having an average value of 1985 and a minimum of 1656.

Accuracy assessment. Table 5 shows the results of the accuracy assessment analysis performed utilizing 12 signatures in the supervised classification process. The error matrix presents an overall accuracy of 61.43% and a KHAT value of 0.5706, these results indicate, based on the criteria previously used, that there is a moderate agreement between the ground truth classes and the classified data. Moreover, Table 6 reveals that the highest accuracy in this approach was obtained with signature 15M (Typic Haplargid), displaying an accuracy of 95.784%, presenting a commission value of 25.587%, and an omission value of 4.216%. On the other hand, the lowest accuracy was obtained by signature 16V (Ustollic Haplargid) with an accuracy of 41.105% (63.883% commission and 58.895% omission).

Complex Approach

Separability analysis. Due to the fact that data collected in the training areas was not normally distributed, the separability analysis was performed using the Euclidean Distance method (Swain and Davis, 1978).

Table 5. Error matrix for the supervised classification in the scaled approach (the horizontal axis is the ground truth, whereas the classified data is shown in the vertical axis).

	13Y	13MC	13MB	14V	11R	12V	16VG	16V	16L	15M	10LL	13MO	Total
13Y	4068	0	865	57	1	0	77	39	0	0	30	1726	6863
13MC	6	3270	0	0	0	0	129	53	3	42	0	163	3666
13MB	914	38	3051	461	18	17	393	245	91	0	9	182	5419
14V	56	0	875	3701	0	3033	0	1	0	0	0	1	7667
11R	0	0	0	0	4497	23	0	0	52	0	1219	0	5791
12V	14	0	163	149	81	3773	1	17	425	0	0	0	4623
16VG	864	52	511	109	0	2	2354	2111	209	6	44	2737	8999
16V	0	1	6	0	0	8	317	4903	6826	4	166	292	12523
16L	0	0	0	0	0	0	41	2950	11075	133	5	455	14659
15M	0	23	0	0	0	0	10	1	1293	4975	0	2	6304
10LL	544	0	327	0	715	1	12	819	212	0	4515	239	7384
13MO	1289	161	2	0	0	0	246	789	448	34	673	9334	12976
Total	7755	3545	5800	4477	5312	6857	3580	11928	20634	5194	6661	15131	96874

Overall classification accuracy: 61.43%. KHAT (Kappa) coefficient: 0.5706.

Table 6. Errors of inclusion and exclusion for the scaled approach after the supervised classification.

Class	Commission	Omission	Accuracy
	- - - - - - - - - -	% - - - - - - - - - -	
13Y	36.041	47.544	52.456
13MC	11.171	7.757	92.243
13MB	40.828	47.397	52.603
14V	88.586	17.333	82.667
11R	24.360	15.343	84.657
12V	12.396	44.976	55.024
16VG	185.615	34.246	65.754
16V	63.883	58.895	41.105
16L	17.369	46.326	53.674
15M	25.587	4.216	95.784
10LL	43.072	32.217	67.783
13MO	24.070	38.312	61.688

The criteria used in this case were that the bigger the distance, the more distinct the signatures are. The analysis established a good separability between most of the ten classes collected, showing an average distance of 10 (spectral distance from a pixel to the mean of a particular class).

Accuracy assessment. Table 7 shows the results of the accuracy assessment analysis performed using 10 signatures for the supervised classification. The error matrix presents an overall accuracy of 28.50% and a KHAT value of 0.2037; these results indicate, based on the criteria previously used, that a poor agreement exists between the ground truth classes and the classified data. Table 8 shows that signature 10V (Typic Paleorthid) achieved the higher accuracy, and indicates that this signature accomplished an accuracy of 78.298% with an omission value of 21.702%. Besides signature 10V, no other signature was observed portraying total accuracy values higher than 50%, suggesting a poor level of agreement between ground truth and classified data through this method. Table 9 shows a summary of the results obtained in the four approaches.

CONCLUSIONS

Based upon the results obtained through the four approaches utilized, the technical approach was selected as the best procedure to classify soils in arid lands. The selection of this procedure was based on the following considerations:

1. The KHAT differences between the simple and technical approaches were the closest considering the four methods used.
2. The technical approach based on the PCA scene compresses all the data in components without losing any major information. In contrast, the simple approach suggests the use of only three bands (2, 4, and 7) compromising therefore, in some degree, the final classification results.

Table 8. Errors of inclusion and exclusion for the complex approach after the supervised classification.

Class	Commission	Omission	Accuracy
	- - - - - - - - - - -	% - - - - - - - - - - -	
13MC	179.097	52.609	47.391
13MM	139.190	99.444	0.556
14V	125.977	64.440	35.560
16L	10.885	68.886	31.114
10V	120.410	21.702	78.298
51S	37.453	79.949	20.051
10LL	47.936	91.668	8.332
13MB	139.431	75.034	24.966
15M	101.040	82.364	17.636
13Y	72.921	77.808	22.192

Table 7. Error matrix for the supervised classification in the complex approach (the horizontal axis is the ground truth, whereas the classified data is shown in the vertical axis).

	13MC	13MM	14V	16L	10V	51S	10LL	13MB	15M	13Y	Total
13MC	1680	224	0	1563	3	1248	301	684	1033	1293	8029
13MM	585	25	0	3404	0	1875	135	60	139	60	6283
14V	4	782	1592	677	515	690	834	792	458	888	7232
16L	1	0	0	6420	0	2222	21	0	2	0	8666
10V	0	1592	820	1181	5271	1003	2785	262	191	272	13377
51S	0	0	0	4509	0	2414	0	0	0	0	6923
10LL	0	513	618	373	703	288	555	272	196	230	3748
13MB	138	728	1116	1066	214	968	1031	1448	1096	1730	9535
15M	224	321	280	632	22	605	451	1152	916	1561	6164
13Y	913	311	51	809	4	726	548	1130	1163	1721	7376
Total	3545	4496	4477	20634	6732	12039	6661	5800	5194	7755	77333

Overall classification accuracy: 28.50%. KHAT (Kappa) coefficient: 0.2037.

Table 9. Summary of the results obtained using the four approaches in the study area.

Approach	Accuracy	Best Thematic Mapper bands to identify soil unit	Best and worst map units identified
Simple	KHAT = 0.6822 accuracy = 70.67%	2, 4, 7	Best: 12V, 13L, 16LS Worst: 13Y, 13MB
Technical	KHAT = 0.6239 accuracy = 66.86%	1, 2, 3 [†]	Best: 10V, 14V, 11L Worst: 13LG, 51S
Scaled	KHAT = 0.5706 accuracy = 61.43%	2, 3, 4 [‡]	Best: 15M, 13MC Worst: 16V, 13MB, 13Y
Complex	KHAT = 0.2037 accuracy = 28.50%	2, 4, 5 [§]	Best: 10V Worst: 13MM

[†] Band 1 = PC1 from PCA analysis of three visible bands. Band 2 = Raw Thematic Mapper Band 4 (Near Infrared). Band 3 = PC1 from PCA analysis of TM Bands 5 and 7.
[‡] Band2 = TM ratio 1/5; Band 3 = TM ratio 3 / 4; Band 4 = TM ratio 5/4.
[§] Band 2 = SAVI; Band 4 = SVI; Band 5 = albedo.

3. The technical approach achieved the highest accuracies in the study area for signatures 10V and 14V with 100% and 99%, respectively. On the other hand, the highest accuracy obtained with the simple approach was 89% with signature 12V. Therefore, to obtain better accuracies in arid regions, the technical approach is recommended.

In Mexico, principally in the arid regions, soil classification by means of remotely sensing data, may be a good alternative to compliment the information already provided by INEGI (Instituto Nacional de Estadística Geografía e Informática). Furthermore, this technique can provide information to those sites where non-digital soil maps at large scale (1:50 000 and larger) have not been produced yet. Map generation by means of remotely sensed data may take advantage of the USDA Soil Taxonomy system, which provides more information than the FAO system currently used in Mexico.

In general, the simple and the technical approaches showed the highest accuracies, but the technical approach is suggested for classifying soils in arid lands because of its data reduction capability that facilitates its digital processing.

Despite the fact that soil classification systems are based on subsurface horizons, Thematic Mapper scenes detect a high percentage (~70%) of mappable soil variability.

REFERENCES

Bauer, M.E., T.E. Burk, A.R. Ek, P.R. Coppin, S.D. Lime, T.A. Walsh, D.K. Walters, W. Befort, and D.F. Heinzen. 1994. Satellite inventory of Minnesota forest resources. Photogrammetric Engineering and Remote Sensing 60: 287-298.

Ben-Dor, E. and A. Banin. 1994. Visible and near-infrared (0.4-1.1 μm) analysis of arid and semiarid soils. Remote Sensing:of Environment 46: 246-267.

Bhatti, A.U., D.J. Mulla, and B.E. Frazier. 1991. Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and Thematic Mapper Images. Remote Sensing of Environment 37: 181-191.

Campbell, J.B. 1996. Introduction to remote sensing. 2nd ed. The Guilford Press, New York.

Coleman, T.L. and O.L. Montgomery. 1987. Soil moisture, organic matter, and iron content effect on the spectral characteristics of selected vertisols and alfisols in Alabama. Photogrammetric Engineering and Remote Sensing 53(12): 1659-1663.

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. and K. Green. 1999. Assessing the accuracy of remotely sensed data: Principles and practices. Lewis Publishers. Boca Raton, FL.

Congalton, R.G. and R.A. Mead. 1983. A quantitative method to test for consistency and correctness in photointerpretation. Photogrammetric Engineering and Remote Sensing 49: 69-74.

Congalton, R.G., R.G. Oderwald, and R.A. Mead. 1983. Assessing Landsat classification accuracy using discrete multivariate statistical techniques. Photogrammetric Engineering and Remote Sensing 54(5): 593-600.

ERDAS. 1995. ERDAS Imagine field guide. 3rd ed. ERDAS. Atlanta, GA.

Escadafal, R., M.C. Girard, and D. Courault. 1989. Munsell soil color and soil reflectance in the visible spectral bands of Landsat MSS and TM data. Remote Sensing of Environment 27: 37-46.

Flechsig, M., M. Erhard, and R. Grote. 1995. Landscape models for the evaluation of ecosystem stability under environmental change: The Duebener Heide case study. pp. 493-500. *In:* M. Heit, H.D. Parker, and A. Shortreid (eds.). GIS applications in natural resources 2. GIS World. Vancouver, British Columbia, Canada.

Frazier, B.E. 1989. Use of Landsat thematic mapper band ratios for soil investigations. Adv. Space Res. 9(1): 155-158.

Frazier, B.E. and Y. Cheng. 1989. Remote sensing of soils in the eastern Palouse region with Landsat Thematic Mapper. Remote Sensing of Environment 28: 317-325.

Gile, L.H. and R.B. Grossman. 1979. The Desert Project soil monograph. National Resources Conservation Service. US Department of Agriculture and National Soil Survey Center. Lincoln, NE.

Gile, L.H., J. W. Hawley, and R.B. Grossman. 1981. Soils and geomorphology of the Basin and Range area of southern New Mexico-Guidebook to the Desert Project. Memoir 39. New Mexico Bureau of Mines and Mineral Resources. Socorro, NM.

Gong, P. and P.J. Howarth. 1992. Frequency-based contextual classification and gray-level vector reduction for land-use identification. Photogrammetric Engineering and Remote Sensing 58: 423-437.

Hawley, J.W. 1975. The Desert Soil-geomorphology project. pp. 183-185. *In:* W.R. Seager, R.E. Clemons, and J.F. Callender (eds.). Guidebook of the Las Cruces County, New Mexico Geological Society, 26th field conference.

Henderson, T.L., A. Szilagyi, M.F. Baumgardner, C.T. Chen, and D.A. Landgrebe. 1989. Spectral band selection for classification of soil organic content. Soil Sci. Soc. Am. J. 53: 1778-1784.

Hepner, G.F. 1990. Artificial neural network classification using a minimal training set:comparison to conventional supervised classification. Photogrammetric Engineering and Remote Sensing 56: 469-473.

Huete, A.R., R.D. Jackson, and D.F. Post. 1985. Spectral response of a plant canopy with different soil backgrounds. Remote Sensing of Environment 17: 37-53.

Jensen, J.R. 1996. Introductory Digital Image Processing: A remote sensing perspective. 2nd ed. Prentice-Hall, New Jersey.

Knick, T.S., J.T. Rotenberry, and T.J. Zarriello. 1997. Supervised classification of Landsat thematic mapper imagery in a semi-arid rangeland by nonparametric discriminant analysis. Photogrammetric Engineering and Remote Sensing 63: 79-86.

Lauer, D.T., S.A. Morain, and V.V. Salomonson. 1997. The Landsat program: Its origins, evolution, and impacts. Photogrammetric Engineering and Remote Sensing 63: 831-838.

Lillesand, T.M. and R.W. Kiefer. 1994. Remote sensing and image interpretation. 3rd ed. John Wiley. New York.

Lo, C.P. and L.J. Watson. 1998. The influence of geographic sampling methods on vegetation map accuracy evaluation in a swampy environment. Photogrammetric Engineering and Remote Sensing 64: 1189-1200.

Mausel, P.W., W.J. Kamber, and J.K. Lee. 1990. Optimum band selection for supervised classification of multispectral data. Photogrammetric Engineering and Remote Sensing 56: 55-60.

Monger, H.C., D.R. Cole, J.W. Gish, and T.H. Giordano. 1998. Stable carbon and oxygen isotopes in Quaternary soil carbonates as indicators of ecogeomorphic changes in the northern Chihuahuan Desert. Geoderma 82: 137-172.

Morse, A., T.J. Zarriello, and W.J. Kramber. 1990. Using remote sensing and GIS technology to help adjudicate Idaho water rights. Photogrammetry Engineering and Remote Sensing 56: 365-370.

Rapaport, D.J., W.G. Withford, and K. Korporal. 1995. Evaluating ecosystem health:opportunities for GIS. pp. 408-413. *In:* M. Heit, H.D. Parker, and A. Shortreid (eds.). GIS applications in natural resources 2. GIS World. Vancouver, British Columbia, Canada.

Rees, W.G. 1990. Physical principles of remote sensing. Cambridge University Press. New York.

Robinson, A.H., J.L. Morrison, P.C. Muehrcke, A.J. Kimerling, and S.C. Guptill. 1995. Elements of Cartography. 6th ed. John Wiley, New York.

Satterwhite, M.B. and J.P. Henley. 1987. Spectral characteristics of selected soils and vegetation in northern Nevada and their discrimination using band ratio techniques. Remote Sensing of Environment 23: 155-175.

Schlesinger, W.H. 1991. Biogeochemistry: An analysis of global change. Academic Press. New York.

Smith, M.O., S.L. Ustin, J.B. Adams, and A.R. Gillespie. 1990. Vegetation in deserts: I. A regional measure of abundance from multispectral images. Remote Sensing of Environment 31: 1-26.

Soil Survey Staff. 1994. Keys to soil taxonomy. 6th ed. US Dept. of Agriculture. Soil Conservation Service. Lincoln, NE.

Stella, W.T. and R.M. Hoffer. 1998. Responses of spectral indices to variations in vegetation cover and soil background. Photogrammetric Engineering and Remote Sensing 64: 915-921.

Stoner, E.R., M.F. Baumgardner, R.A., Weismiller, L.L. Biehl, and B.F. Robinson. 1980. Extension of laboratory-measured soil spectra to field conditions. Soil Sci. Soc. Am. J. 44: 572-574.

Swain, P.H. and S.M. Davis. 1978. Remote sensing: The quantitative approach. McGraw-Hill. New York.

Thenkabail, P.S. 1992. Capabilities of Landsat-5 Thematic Mapper (TM) data in studying soybean and corn crop variables. PhD. Dissertation. The Ohio State University. Columbus, OH.

Thompson, D.R., D.E. Pitts, and K.E. Henderson. 1983. Simulation of Landsat Multispectral scanner response of soils using laboratory reflectance measurements. Soil Sci. Soc. Am. J. 47: 542-546.

Van Deventer, A.P. 1992. Evaluating the usefulness of Landsat Thematic Mapper data to determine soil properties, management practices and soil water content. PhD. dissertation. The Ohio State University. Columbus, OH.

Van Deventer, A.P., A.D. Ward, P.H. Gowda, and J.G. Lyon. 1997. Using Thematic Mapper data to identify contrasting soil plains and tillage practices. Photogrammetric Engineering and Remote Sensing 63: 87-93.

Weismiller, R.A. and S.A. Kaminsky. 1978. Application of remote sensing technology to soil survey research. J. Soil Water Conserv. 33: 287-289.

Wilcox, C.H., B.E. Frazier, and S.T. Ball. 1994. Relationship between soil organic carbon and Landsat TM data in Eastern Washington. Photogrammetric Engineering and Remote Sensing 60: 777-781.