REMOTE SENSING OF SOILS IN THE SANTA MONICA MOUNTAINS: HIERARCHICAL FOREGROUND AND BACKGROUND ANALYSIS

Alicia Palacios-Orueta[1], Jorge E. Pinzón[2], Susan L. Ustin[1], and Dar A. Roberts[3]
[1]Department of Land, Air, and Water Resources, [2]Department of Applied Mathematics, University of California, Davis, CA 95616, USA; [3]Department of Geography, University of California, Santa Barbara, CA

1. INTRODUCTION

The extreme spatial and temporal variability of surface processes makes soil properties extremely variable and therefore, difficult to measure. Since soil and ecosystem processes occur at different scales, it is necessary to work at sufficiently large spatial resolution and coverage for generalizations to be made. Organic matter is a soil property closely related to soil quality, not only as an indicator of soil erosion and degradation, but also as a regulating factor of processes such as nutrient availability, water holding capacity, and permeability. Because values of organic content are highly variable and react very quickly to external changes (Gerrard, 1992), decomposition rates show high spatial variability. The spatial distribution of organic matter content can be an indicator of the rate of decomposition and other processes happening on the soil surface, such as differences in deposition and erosion rates or microclimate factors.

Imaging spectrometry offers a potential way to map certain soil properties that are relevant to surficial processes at the landscape scale. In the last few years the analysis of hyperspectral data and image processing techniques have improved to the point that they offer the potential for direct analysis of soil properties. Several multispectral sensors have already been used for discrimination between soils (Lewis et al, 1975; Agbu et al, 1990; Coleman et al, 1993). Specifically, there are several studies where organic matter has been analyzed in terms of its reflectance properties (Stoner and Baumgardner, 1981; Henderson et al, 1989). Hyperspectral data, specifically Advanced Visible Infrared Imaging Spectrometer (AVIRIS) data has been shown to be useful for improved discrimination of minerals (Clark et al, 1990; Kruse et al, 1990). Also, several other studies have dealt with soil identification and discrimination directly or indirectly using AVIRIS data (Smith et al, 1990; Roberts et al, 1993; Palacios-Orueta and Ustin, 1996, 1998b).

A significant problem for soil analysis is the presence of vegetation in most pixels. Because the signatures of soils and vegetation are so different, the lesser variability contained within the soil component is not significant enough for soil discrimination when vegetation is also present in the pixel. HFBA (Hierarchical Foreground and Background) (Pinzón et al, 1988) is a new steerable analytic technique where the Foreground and Background Analysis (FBA) equation (Smith et al, 1994) is applied at several levels in a hierarchical way, thus, the variability contained in the data set is confined at each step, making it possible to extract subtle absorption features. For this model FBA was modified to project the spectra into a property-specific axis of continuous variation. To examine the application of this method to extract and improve detection of soil properties using an imaging spectrometer, we applied it to map the spatial distribution of organic matter content from samples in two watersheds in the Santa Monica Mountains Recreation Area. In earlier work, Palacios-Orueta and Ustin (1988a) found that soils from each valley could be discriminated based on organic matter and iron content and that these could be spectrally estimated with reasonable accuracy in soil samples. The purpose of this work was to test the performance of HFBA (Hierarchical Foreground and Background Analysis) applied to AVIRIS data for the discrimination of these soils and soil properties.

2. METHODS AND MATERIALS

2.1 Study Sites, Soils, and Geologic Materials

The soils are from two south-to-north trending watersheds, La Jolla Valley and Serrano Valley, within the Point Mugu State Park in the western (coastal) region of the Santa Monica Mountains National Recreation Area, between Ventura and Los Angeles Counties, California. The climate is typically Mediterranean with dry summers and mild winters. In late 1993 a wildfire removed most of the vegetation in both valleys making it possible to observe soils with minimal vegetation cover. Further description of this area is found in Palacios-Orueta (1997) or Palacios-Orueta and Ustin (1988a). La Jolla Valley soils are formed from weathered sandstone and shale while Serrano Valley soils are derived from basic igneous rock. The soil moisture regime in both valleys is xeric and the soil temperature regime is considered thermic. The steep terrain and the distance to the ocean create different microsite environments which consequently result in high soil variability (Edwards et al, 1970).

2.2 Field Soil Data Collection

Seventy-four soil sample composites were collected from the valleys. Samples were selected to represent the range of aspect, slope, elevation and parent materials within the area, although this goal could not be completely achieved due to the roughness of the terrain. The locations of the soil samples were identified using a Global Positioning System unit (Trimble Navigation PROXL) with +/- 1 meter accuracy after differential correction.

2.3 Physico-chemical laboratory analyses

The soil samples were analyzed by the DANR (Division of Agriculture and Natural Resources) Analytical Laboratory at U.C. Davis for organic matter, iron content, and texture. Organic matter content was significantly higher in La Jolla Valley although variances were similar in both valleys (Palacios-Orueta and Ustin, 1988a). The soil sample preparation for spectrometry followed the standardized procedure from Henderson et al (1992). Further information about this procedure, spectroscopic technique and soil characteristics can be found in Palacios-Orueta (1997).

2.4 Spectroscopic Analysis

The spectral data set includes laboratory reflectance spectra (400-2500 nm) measured in a Varian Cary 5E spectrophotometer, and two AVIRIS scenes acquired April ll, 1994.

2.5 Geographic Information Systems Database

The geographic information was organized in a GIS (Arc/Info) database. The AVIRIS scenes were georeferenced using control points and combined in the database with ancillary information composed of a Digital Elevation Model, vegetation map, and the digitized geologic map. The organic matter content was also included in the database. The AVIRIS imagery used for this study was acquired approximately six months after the wildfire and at the end of the winter precipitation period. Two adjacent image scenes were used for this analysis. Apparent surface reflectance retrieval used a radiative-transfer based atmospheric model (MODTRAN 2) that accounts for spatial variation of the atmospheric conditions (Green et al, 1993).

3. METHODOLOGY

HFBA was developed by Pinzón et al (1995) as an improvement of FBA (Smith et al, 1994). The approach taken (Fig. 1) was to narrow the variance by stratification of the soil population into small but reliable ranges of soil variability that can be consistently detected. This process is done by first discriminating the soils between the valleys, and second, by investigating the variability related to organic matter content within each group. A smaller range of spectral variability is found in the second level. These vectors are calculated using a training set derived from laboratory data and Singular Value Decomposition (SVD) is used to solve the HFBA equation at each level in the analysis. The performance was tested with the whole laboratory spectral data set and then applied to the AVIRIS image. Each pixel was classified as being of the soil type of one of the two valleys or into a class level of organic matter content. The analyses were done in Matlab (1994).

4. RESULTS AND DISCUSSION

4.1 First Level of Soil Classification: Vector Training

Although other sources of variability between the soils at these two locations are likely, the spectral variability due to the combination of organic matter is summarized in this step. HFBA uses a supervised classification scheme where each valley was represented by a scale of values in which Serrano type soils ranged from 0 to 7 and La Jolla from 7 to14. Then, spectra in the training set were projected by the HFBA vector to the center of each class. Figure 2 shows the mean spectra for each valley and the HFBA vector that yielded the best discrimination between valleys. It can be observed that the two spectral areas most important for discrimination between the valleys were near 1000 nm and 2200 nm. Although the greatest weights were given to the band at 2200 nm, an area between 700 and 1400 nm and centered on 1000 nm was consistently negatively weighted. This means that a wide area around 1000 nm is important in the discrimination while only a few bands around 2200 nm are significant. Palacios-Orueta and Ustin (1988a) found that reflectance around 1000 nm was not only related to organic matter content but also iron, thus low reflectance in this band by itself is not sufficient to determine the organic matter content. From the mean spectra, it is observed that the reflectance at 1000 nm is significantly different between valleys. The absorption bands centered at 2200 and 2300 nm are most likely due to the presence of OH-Al and Mg-OH in dioctahedral and trioctahedral clays respectively (Hunt and Salisbury, 1970). The differences in geologic parent materials could produce this effect. These results combined with the analysis of error (Table 1) support the idea that although there must be other sources of variability, organic matter and iron contents play a critical role in the spectral discrimination between valleys.

4.2 Second Level: Organic Matter Content

At the second HFBA level, the analysis focused on extracting information related the biogeochemistry. In order to do this, two analysis tools were used: the quantization of the chemical data into ranges and the selection of the soil samples for the training set. In each group defined at the first level, two new vectors were trained to classify spectral samples for organic matter. Vectors **a** and **b** (Fig. 3 a,b) were trained with the soils classified either as Serrano or La Jolla types. Both HFBA vectors show a concave shape around the 700 nm region although in La Jolla the minimum value is slightly shifted towards 800 nm. In these soils the weights increase until reaching the highest value at 1400 nm. The band at 2200 nm is highly weighted in Serrano, while in La Jolla the band centered at 2300 nm has highly positive weights. In La Jolla, the vector is smoother over a wider range of wavelengths, possibly due to the higher organic matter and lower iron contents in this valley. Organic matter characteristics are stronger and their features are more clearly observed. Table 2 shows the quantization levels for organic matter content, and the number of samples in the whole data set and in the training set. Figure 4 shows the distributions of the measured and the predicted values for both valleys. The continuous line represents the predicted values and the dashed line represents the measured data. The r^2 from the regression analysis is 0.72, and only five samples were outside of one standard deviation. The distributions of the predicted and the measured data follow similar patterns assigning more samples to the centrally placed values.

4.3 Classification between valleys

The first vector was trained to assign each pixel a classification value that will locate it in one of the two soil types. Since many pixels are not pure soils they are classified outside the range of the original classes (0-14). Although the area under study was recently burned in a major wildfire there was a considerable amount of revegetation in some areas, mainly in the moister valley bottoms. Because the image was acquired following the wildfire and winter storms we expected the amount of dry vegetation to be low, thus, decreasing the possibility of confusion with soil.

There are also some terrestrial areas in the image that were not affected by the wildfire and remained vegetation covered. Masking the vegetation using an NDVI threshold is an arbitrary decision and pixels with small but undetermined amounts of vegetation still remain. Our interest lies in discriminating soil properties in pixels over a range of partial vegetation cover. Since the vectors are trained with pure soils, we expect that pixels having some vegetation will still show soil characteristics while pixels with higher levels of vegetation cover will be out of the range of the predicted soil property values. This allows an *a posteriori* decision about vegetation cover that is derived from the soil information rather than an *a priori* vegetation based decision. The NDVI (Fig. 9a) is shown as a reference and used to compare the spatial distribution of the vegetation derived from the HFBA but it was not used directly to mask vegetation in the analysis. Our results showed that the negative values projected by the classification vector were pixels with high NDVI (>0.5), providing some confirmation of the methodology.

A histogram of the results (Fig. 5) shows that the AVIRIS distribution forms a long tail with only a few pixels having values higher than 21. Nearly all pixels with values higher than 14 were located in the ocean, therefore we used this criteria to remove them from further consideration in the soil analysis. All pixels with values less than 0 were classified as vegetation. The remaining "potential soil" pixels in the image were classified at several levels. Pixels with values between 0 and 7 were assigned to Serrano type soil class, i.e., have the physicochemical properties of Serrano Valley soils, and pixels with values between 7 and 11 have the physicochemical characteristics of soils from La Jolla Valley soil type. Pixels with values between 11 and 14 are located in the beach areas, and although they have soil properties, due to the high albedo of the sand they are projected in the high extreme of the soil range. The Serrano soil type is assigned a light gray and soils classified as La Jolla are assigned dark gray in Figure 9b. Comparing these results with the NDVI shows that areas with NDVI > 0.5 (black) follow the same spatial pattern as the pixels that were not classified (white) in Figure 9b. The image the La Jolla soil type pixels are clustered in patches, and the pixels classified as Serrano soil type are distributed more continuously over most of the image.

The first level of classification allowed us to select pixels classified as Serrano or La Jolla soil types. Only pixels with enough spectrally expressed soil to fall within the laboratory data range were analyzed in the second step. Thus the variability due to soils alone is identified and this variability is divided into that produced by La Jolla and Serrano soil types. This hierarchy optimizes the application of the organic matter vector.

3.5 Organic Matter Content Determination

Organic matter was estimated applying the vectors trained with laboratory data. Vector **a** was used to predict organic matter content in pixels classified as Serrano type soils in the first level classification and Vector **b** was used for pixels classified as La Jolla type soils (Fig. 6). Although the predicted organic matter values range from –15 to 10%, the range most pixels are found within the 1 - 6%, the same range as laboratory data. Pixels where soil is not the primary component show projected values outside this range. The ocean pixels have extreme high projected values of organic matter content and pixels having high NDVI (> 0.5) are projected to low or negative values of organic matter content, thus making it straight forward to remove them from consideration. The distribution of the AVIRIS soil organic matter for the two soil types (Fig. 7) followed the same trend as the laboratory data; the range is the same although the distribution is different. Figure 9c shows the results of the analysis for AVIRIS pixels having a high soil component (i.e. organic matter between 0 – 6%). The light gray indicates low and dark gray indicates high organic matter, white indicates pixels out of range in the first classification level or had negative organic matter content values. Soils with high organic matter content are not uniquely associated with La Jolla Valley and it is observed that the pixels mapped as La Jolla type soils (Figure 9b) show high organic matter content in Figure 9c (e.g., northeast area). This pattern agrees with our laboratory soils data. The spatial distribution of organic matter was related to the aspect (Fig. 8). North and east facing aspects are generally cooler and more humid and characteristically accumulate higher levels of organic matter. It can be seen that high values are predominant on north facing aspects while lower values are predominant on south facing slopes. AVIRIS is not simply mapping soils in separate valleys but the distribution of organic matter as a continuous variable. The spatial variation in organic matter is probably related to the steep terrain and aspect. This property is not unique to the respective valley and because variability within the valleys is high it is representative of the larger region. Palacios-Orueta and Ustin (1998b) found spatial variability in iron content was independent of the organic matter content but the variation in the valleys was sufficient to map the larger region.

4. CONCLUSIONS

HFBA was found to be a suitable method for soil analysis to determine relative changes in organic matter content because it is sequentially structured so that soil properties can be quantized into different ranges of variance. There are a combination of features that makes this spectral model work more efficiently than standard classification methods:

1. It is a mixture model therefore it can use continuous data over the whole spectrum.
2. It is steerable model, maximizing variability between classes and minimizing variability within classes, optimizing the amount of information extracted.
3. As a supervised classification algorithm, it can be focused on specific soil properties.
4. The Singular Value Decomposition equation efficiently discriminates between foreground soil properties and background environmental conditions.

5. The hierarchy reduces variability at each step allowing subtle absorption features to be extracted.

The results obtained when training the vectors with the laboratory data showed that the organization of the system and the singular value decomposition transformation work effectively in predicting organic matter from spectral data. Although the classified soils were not uniquely associated with either valley, the predictions of organic matter content from the image agreed with the soil characteristics from the field sampling locations. This methodology is based on a hierarchical analysis, which implies that variability is reduced at several steps, each time becoming more specific. The use of HFBA provides a mechanism to efficiently reduce the number of field measurements, or to use a vector developed from an area having similar soil variability. HFBA would be very useful for identifying changes in soil properties in a temporal framework. To understand landscape soil processes more completely, the data could be further analyzed in a geographic context, e.g., using a relational GIS database.

5. REFERENCES

Agbu, P.A., D.J. Fehrenbacher, and I.J. Jansen, 1990, "Statistical Comparison of SPOT Spectral Maps with Field Soil Maps," Soil Sci. Soc. Am. J., vol. 54, no. 3, pp. 812-818.

Clark, R.N., A.J. Gallagher, and G.A. Swayze, 1990, "Material Absorption Band Depth Mapping of Imaging Spectrometer Data using a Complete Band Shape Least-Squares Fit with Library Reference Spectra," Proceedings of the Second Airborne Visible/Infrared Image Spectrometer (AVIRIS) Workshop, Pasadena, California, pp. 176-186.

Coleman, T.L., P.A. Agbu, and O.L. Montgomery, 1993, "Spectral Differentiation of Surface Soils and Soil Properties - Is It Possible from Space Platforms?" Soil Sci., vol. 155, no. 4, pp. 283-293.

Edwards, R.D., D.F. Rabey, and R.W. Kover, 1970, "Soil Survey, Ventura Area, California," United States Department of Agriculture, United States Soil Conservation Service, 148 pp.

Green, R.O., J.E. Conel, and D.A. Roberts, 1993a, "Estimation of Aerosol Optical Depth and Calculation of Apparent Surface Reflectance from Radiance Measured by the Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) using MODTRAN 2," SPIE Conf. 1937, Imaging Spectrometry of the Terrestrial Environment, 12 pp.

Henderson, T.L., M.F. Baumgardner, D.P. Franzmeier, D.E. Stott, and D.C. Coster, 1992, "High Dimensional Reflectance Analysis of Soil Organic Matter," Soil Sci. Soc. Am. J., vol. 56, no. 3, pp. 865-872.

Henderson, T.L., A. Szilagyi, M.F. Baumgardner, C.T. Chen, and D.A. Landgrebe, 1989, "Spectral Band Selection for Classification of Soil Organic Matter Content," Soil Sci. Soc. Am. J., vol. 53, no. 6, pp. 1778-1784.

Kruse, F.A., K.S. Kiereinyoung, and J.W. Boardman, 1990, "Mineral Mapping at Cuprite, Nevada with a 63-Channel Imaging Spectrometer," Photogramm. Eng. and Remote Sens., vol. 56, no. 1, pp. 83-92.

Lewis, D. T., P.M. Seevers, and J.V. Drew, 1975, "Use of satellite imagery to delineate soil associations in the sands hills region of Nebraska," Soil Sci. Soc. Am. Proc., vol. 39, no. 2, pp. 330-335.

Matlab (1994). The MathWorks, Inc.

Palacios-Orueta, A. 1997, "Soil Discrimination with Laboratory Spectra and Airborne Imaging Spectrometer Data (AVIRIS)," Department of Land, Air and Water Resources. Davis, University of California, 120 pp.

Palacios-Orueta, A. and S.L. Ustin, 1996, "Multivariate Classification of Soil Spectra. Remote Sens. Environ., vol. 57, no. 2, pp. 108-118.

Palacios-Orueta, A. and S.L. Ustin, 1998, "Remote Sensing of Soil Properties in the Santa Monica Mountains: I. Spectral Analysis," Remote Sens. Environ., (in press).

Pinzón, J.E., S.L. Ustin, C.M. Castaneda, and M.O. Smith, 1998, "Investigation of Leaf Biochemistry by

Hierarchical Foreground/Background Analysis," IEEE Trans. Geosci. and Remote Sens., (in press).

Pinzón, J.E., S.L. Ustin, Q.L. Hart, S. Jacquemoud, and M.O. Smith, 1995, "Using Foreground/Background Analysis to Determine Leaf and Canopy Chemistry," Summaries of the Fifth Annual JPL Airborne Earth Science Workshop: AVIRIS Workshop, vol. 1, pp. 129-132.

Roberts, D.A., M.O. Smith, and J.B. Adams, 1993, "Green Vegetation, Nonphotosynthetic Vegetation, and Soils in AVIRIS Data," Remote Sens. Environ., vol. 44, nos. 2/3, pp. 255-269.

Smith, M.O., D.A. Roberts, J. Hill, W. Mehl, B. Hosgood, J. Venderbout, G. Schmuck, C. Koechler, and J. Adams, 1994, "A New Approach to Quantifying Abundances of Materials in Multispectral Images," Institute of Electrical Electronics Engineering, Int. Geosci. Remote Sens. Trans., IGARSS '94, California Institute of Technology, Pasadena, CA (CD), vol. 4, pp. 2372-2374.

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 Sens. Environ., vol. 31, no. 1, pp. 1-26.

Stoner, E.R. and M.F. Baumgardner, 1981, "Characteristic variations in reflectances of surfaces soils." Soil Sci. Soc., Am. J., vol. 45, no. 6, pp. 1161-1165.

Table 1. Table of classification errors after first level classification step from spectral analysis of soil samples.

		Number of soil samples that belong to:		
		La Jolla	**Serrano**	**Total**
Number of soil samples classified in:	**La Jolla**	34	1	35
	Intermediate	5	2	7
	Serrano	3	29	32
	Total	42	32	74

Table 2. Quantization levels (R1-R4) for ranges of soil organic matter content and number of samples in the training set and in the whole laboratory soil sample data set.

	Organic Matter (%)							
	La Jolla Valley				**Serrano Valley**			
	R1	**R2**	**R3**	**R4**	**R1**	**R2**	**R3**	**R4**
Range Center	0.98	2.33	3.68	5.03	1.1	2.04	2.99	3.93
Training Set	3	6	3	3	2	8	1	2
Complete Data Set	4	15	13	10	7	13	7	5

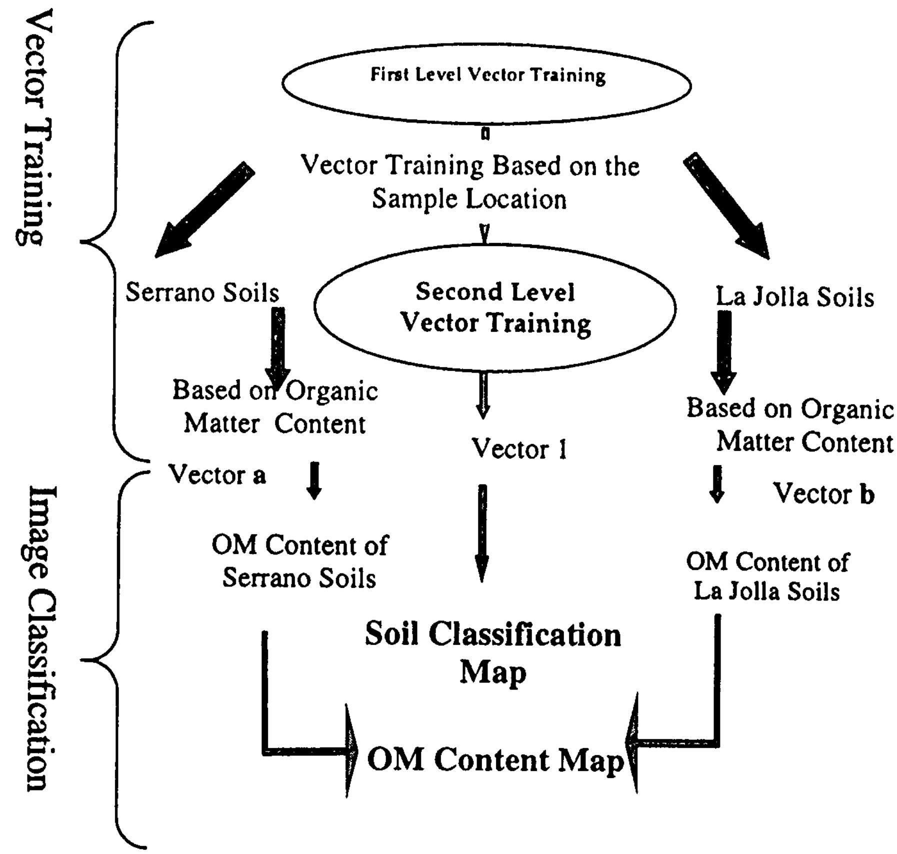

Figure 1. Schematic organization for HFBA Analysis.

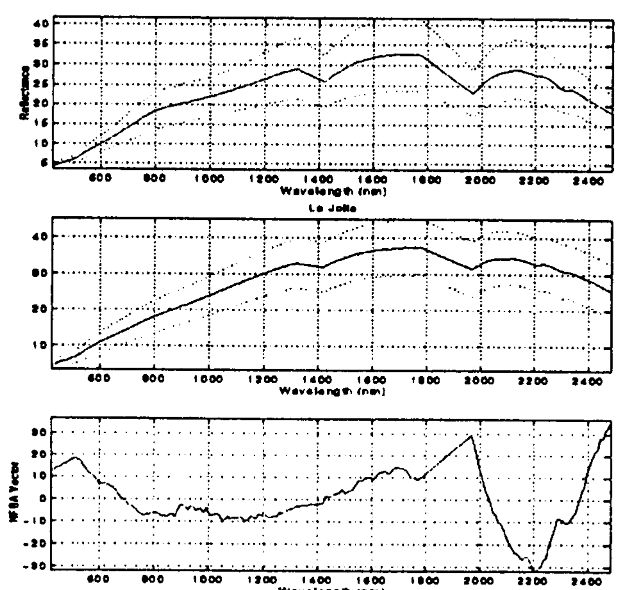

Figure 2. Mean reflectance spectra for La Jolla (a), Serrano Valleys (b) and HFBA vector (c) for first level of classification.

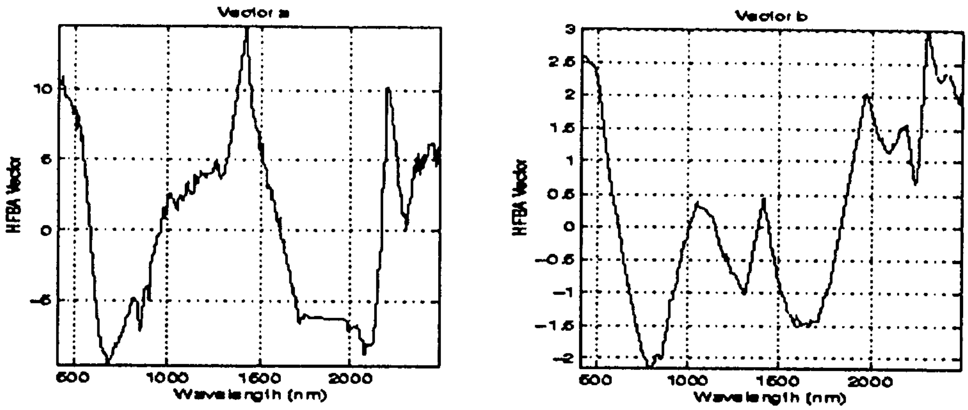

Figure 3. HFBA training vector weightings for predicting organic matter content. Vector *a* corresponds to Serrano and vector *b* to La Jolla type soils.

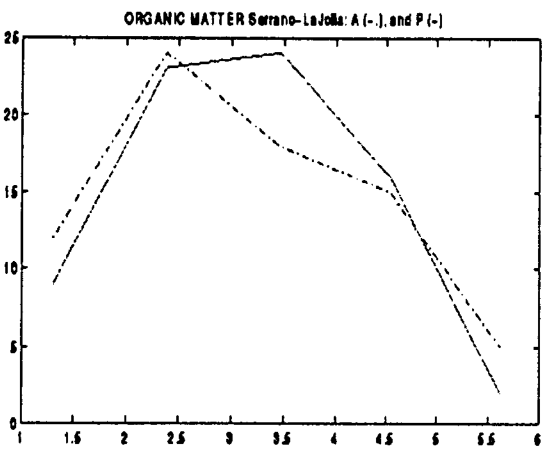

Figure 4. Distributions of the predicted and the measured data for organic matter content for laboratory soil samples. The continuous line represents the predicted values and the dashed line represent the measured values.

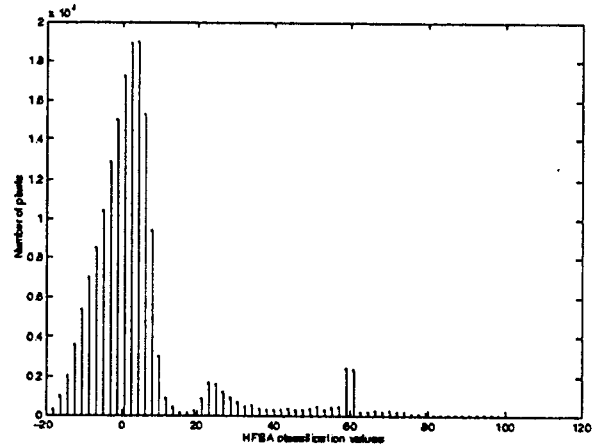

Figure 5. Histogram of the distribution of AVIRIS soils after applying the classification vector at the first level.

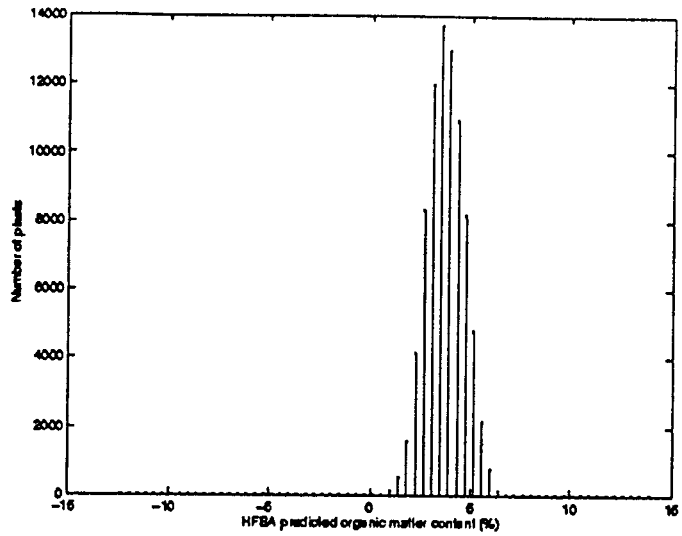

Figure 6. Predicted organic matter content distribution from AVIRIS data.

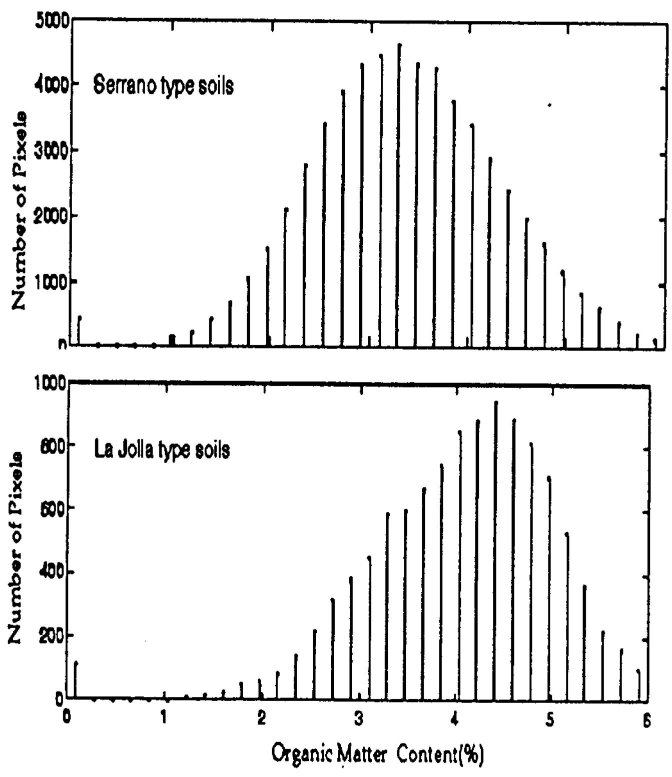

Figure 7. Predicted organic matter content distributions for soils having characteristics of Serrano and La Jolla type soils.

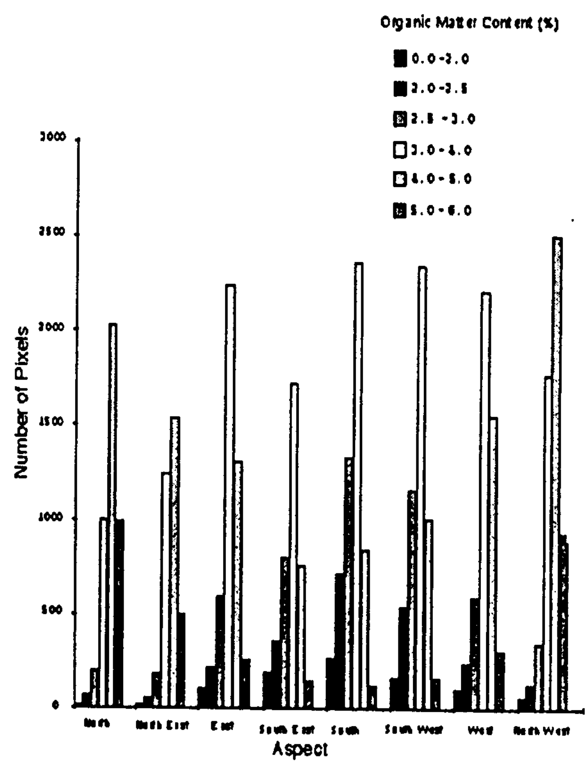

Figure 8. Predicted distribution of organic matter content for different aspects from AVIRIS data.

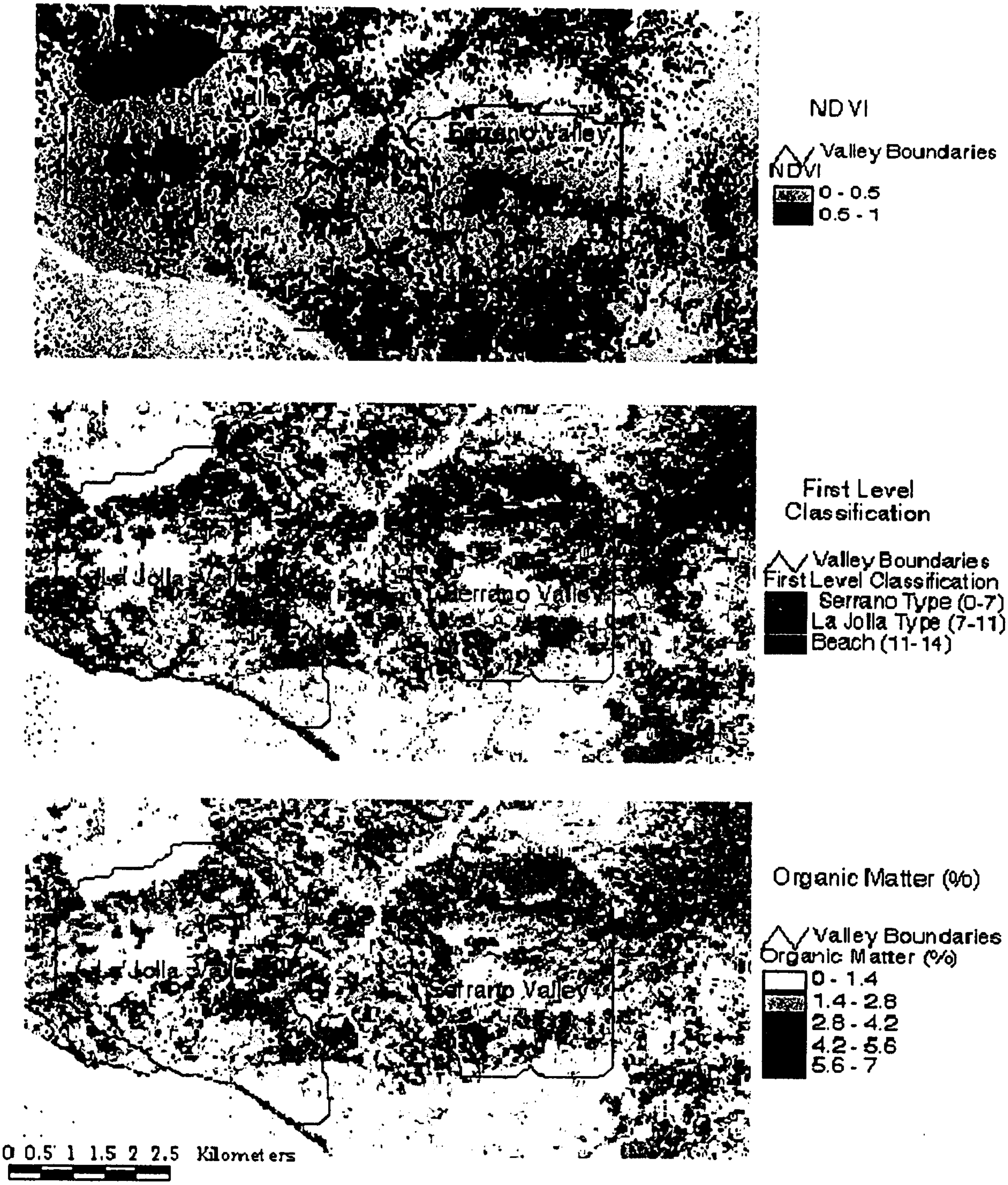

Figure 9a. AVIRIS NDVI results showing location of vegetated pixels. Black indicates high vegetation (NDVI>0.5), light gray indicates the low vegetation cover. Figure 9b. Classification results at the first level. White corresponds to ocean and vegetation dominated pixels. Light gray corresponds to Serrano type soils, dark gray to La Jolla type soils and black to the beach areas. Figure 9c. Predicted organic matter ranges for pixels having low vegetation cover. Black and dark gray correspond to high levels and light gray corresponds to low levels, while to ocean and vegetation dominated pixels.