SOIL CLASSIFICATION VIA MID-INFRARED SPECTROSCOPY

Raphael Linker
Faculty of Civil and Environmental Engineering, Technion-Israel Institute of Technology, Haifa, 32000, ISRAEL, linkerr@technion.ac.il*, http://www.technion.ac.il/~civil/linker/*

Abstract: The need for rapid and inexpensive techniques for soil characterization has led to the investigation of modern technologies, and in particular those based on reflectance spectroscopy. While near-infrared has been traditionally used, mid-infrared in the 400-4000 cm-1 range is becoming increasingly common due to the specificity of the absorbance bands in this spectral range. The present work discusses two methods based on mid-infrared spectroscopy for soil classification: attenuated total reflectance (ATR) and photoacoustic spectroscopy. The ATR method requires a soil sample close to water saturation, and as a result only the 800-1600 cm-1 interval of the spectrum yields a useful signal. Typical ATR soil spectra consist mostly of several broad bands in the 800-1200 cm-1 region and a calcium carbonate band around 1450 cm-1. By comparison, photoacoustic measurements are conducted with air-dried samples, and the photoacoustic spectra exhibit a larger number of clearly-defined bands. Both methods were tested on data sets containing over 100 samples of various soils commonly used in Israeli agriculture. Data analysis was conducted by wavelet decomposition and neural network classifiers. Very good classification performances were achieved, with correct classification rates of the validation samples typically above 95%.

Key-Words: Fourier transforms infrared (FTIR); attenuated total reflectance (ATR); photoacoustic spectroscopy (PAS); wavelets; neural networks

1. INTRODUCTION

Precision farming and similar modern approaches for efficient management of land resources require fast and accurate methods for soil characterization. Standard laboratory techniques for soil analysis are labour- and time-consuming, and extensive research has been devoted to the development of new methods for rapid screening of large number of soil samples (Viscarra et al., 1998; McBratney et al., 2006). Among the approaches investigated, spectroscopy, both in the near-infrared (NIR) (Ben-Dor et al., 1995; McCarty et al., 2002; Daniel et al., 2003) and mid-infrared ranges, has yield very promising results (Viscarra et al., 2006). While NIR spectra consist of non-specific overtones that are difficult to interpret, mid-infrared spectra consist of specific bands that can be directly associated with soil constituents. With respect to soil analysis, most mid-infrared studies were conducted in transmittance (Haberhauer et al., 1998; Haberhauer et al., 1999; Gerzabek et al., 2006), diffuse reflectance (DRIFT) (McCarty et al., 2002; Haberhaue et al., 1999; Janik et al., 1998; Nguyen et al., 1991) and attenuated total reflectance (ATR) (Linker et al., 2004; Linker et al., 2005; Linker et al., 2006) modes. Transmittance studies revealed numerous absorbance bands that could be associated with organic as well as inorganic soil components. However, this technique requires the time-consuming preparation of KBr pellets and is not suitable for routine analysis of large amounts of samples. In addition, such measurements involve very small quantities of soil (typically less than 1 mg per sample), and the representativeness of the sample may be questionable. Although some of these limitations are overcome with the DRIFT technique which does not strictly require the preparation of pellets, soil grinding and dilution with KBr usually improves the results significantly. By comparison, the ATR technique requires very minimal sample preparation. However, contrary to transmittance and DRIFT techniques, ATR requires very good contact between the sample and a crystal that serves as a waveguide for the IR beam (Fig. 1.). The IR beam is directed in such a way that it hits the crystal/sample interface several times. Each time, the evanescent wave penetrates a few microns into the sample, so that the signal that reaches the detector contains information about the absorbance of the sample. Since the penetration depth is limited to a few microns, very good contact between the sample and the crystal is required. For soil samples, this can be achieved by working with samples close to water saturation (Linker et al., 2004; Linker et al., 2005; Linker et al., 2006; Shaviv et al., 2003). Unfortunately, water exhibits very strong absorbance bands in the mid-IR range, which may distort or hide bands of interest.

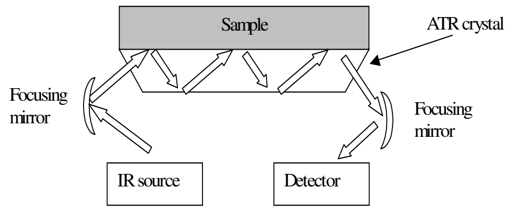

Figure 4. Schematic description of ATR spectroscopy

Changwen et al. (Changwen et al., Changwen et al.,) recently suggested the use of photoacoustic (PAS) mid-IR spectroscopy for soil identification. Photoacoustic spectroscopy is based on absorption-induced heating of the sample, which produces pressure fluctuations in a surrounding gas. These fluctuations are recorded by a microphone, and constitute the PAS signal (McClelland et al., 2001). The major advantage of photoacoustic spectroscopy is that it is suitable for highly absorbing samples, such as soils, without any special pre-treatment.

Regardless of the method used to obtain the spectra, mathematical processing is required for automated soil classification. Due to the high dimensionality of the spectra, this typically involves a data-reduction stage followed by some classification tool. The most commonly used method for data reduction is principal component analysis (PCA) (Jolliffe, 1986) that performs a linear decomposition of the data. This approach was used by Linker et al. (Linker et al., 2006) and Chanwen et al. (Changwen et al,) for ATR and PAS soil spectra, respectively. Wavelet transform is another data-reduction method that is becoming increasingly popular for spectrum analysis (Walczak et al., 1997; Trygg et al., 1998; Ehrentreich et al., 2002; Liu et al., 2004; Figueiredo et al., 2007). The main feature of wavelet decomposition is that the resulting coefficients contain information about both the location and shape (sharpness) of the spectral bands. With respect to spectroscopy, three types of wavelet-based methods have been investigated (continuous wavelet, discrete wavelet and wavelet packet transform), and details relative to each method can be found in the literature (Walczak et al., 1997; Figueiredo et al., 2007; Leung et al., 1998; Jahn et al.,). Wavelet transformation by itself does not produce a compressed representation of the original data, and data reduction is achieved by eliminating the wavelet coefficients that do not contain valuable information. This is a non-trivial task and various approaches have been reported in the literature, such as eliminating all "small" coefficients using either simple thresholding (Ehrentreich et al., 2002; Liu et al., 2004; Figueiredo et al., 2007; Leung et al., 1998), mutual information (Alsberg et al., 1998) or genetic algorithms (Depczynski et al.,1999; Zhang et al., 2003).

The present paper presents a method based on wavelet decomposition and a neural network classifier for classification of both ATR and PAS spectra.

In this study, the approach recommended by Trygg and Wold (Trygg et al., 1998), which consists in retaining the coefficients with the highest variance, was used. The selected coefficients were used as inputs to a neural network (NN) classifier.

2. MATERIALS AND METHODS

2.1 Sample preparation and spectroscopic measurements

Details concerning sample preparation and spectroscopic measurements can be found in (Linker et al., 2006) and (Changwen et al.,; Changwen g et al.,) for ATR and PAS, respectively, and only crucial information is recalled here. The ATR measurements were conducted with 202 samples close to water saturation, representative of five soil types commonly encountered in Israeli agriculture (Table 1). For the PAS measurements, 160 air-dried samples of the same types of soils were used. Although these samples were not strictly identical to the ones used for the ATR study, they had very similar characteristics and belonged to the same soil types (details not shown, see (Changwen et al.,).

Table 1. Properties of the soils used for ATR measurements

Soil type denomination in text	Clay content (%)	$CaCO_3$ content (%)	Organic matter content (%)
Grumosol	50-70	5 -25	1.1-1.3
Loess	15-30	10-30	0.8 -1.1
Rendzina	40-55	35-45	0.8 -1.1
Hamra	5-35	0 – 1	0.5 - 0.8
Terra Rosa	45-70	<1	1.0-1.4

2.2 Data analysis

2.2.1 Pre-processing of spectra

The ATR spectra were smoothed using a second-order Savitzky-Golay filter with a 15-point window, and the method developed by Linker et al. (Linker et al., 2004) was applied for water-subtraction and baseline correction. The corrected spectra were then normalized so that the integral of the spectra would be equal to one. For the photoacoustic spectra, only smoothing (using a first-order Savitzky-Golay filter with a 25-point window) and normalization was applied.

2.2.2 Data reduction

Data reduction was achieved using the discrete wavelet transform with a Coiflet mother-wavelet. The approach recommended by Trygg and Wold (Trygg et al., 1998) was used to determine which of the resulting wavelet coefficients contained most of information. According to this method, the coefficients were sorted according to their variances and only the N coefficients with the highest variances were retained. The procedure is depicted schematically in Fig. 2 and the main steps are (1) wavelet transformation of each spectrum, (2) concatenation of the wavelet coefficients, (3) calculation of the coefficient variances and (4) extraction of the coefficients with the largest variances.

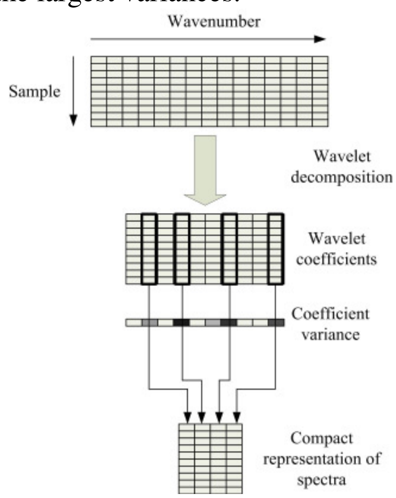

Figure 5: Schematic description of the wavelet-based procedure for data reduction.

2.2.3 Classification

Feedforward neural networks (NN) with sigmoid activation functions were used as non-linear classifiers (Haykin et al., 1999) The inputs of the classifiers consisted of the wavelet coefficients selected at the previous stage. The NN had five outputs, corresponding to the five soil classes included in the study. A "winner-takes-all" approach was used, according to which the identified type was that of the output node with the largest value. All the classifiers were calibrated using half of the data (chosen randomly) and validated with the remaining data.

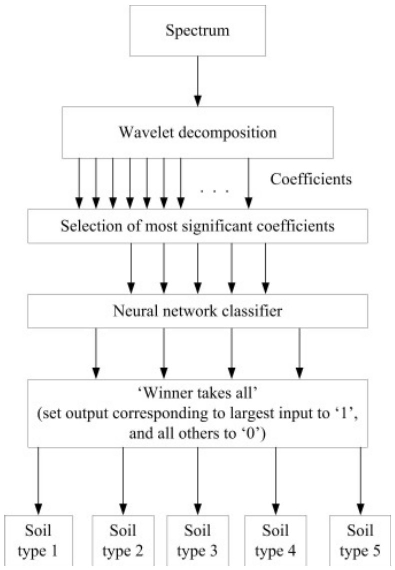

Figure 6: Schematic representation of the whole data-reduction & classification procedure

3. RESULTS

3.1 Attenuated total reflectance

Figure 4 shows typical water-subtracted and normalized spectra. For each soil type, only five spectra are shown for clarity. All the soils have various absorbance bands in the 800-1200 cm-1 interval, which are centered at 870, 915, 1025, 1110 cm-1. In addition, the calcareous soils have a strong absorbance band centered at 1440 cm-1 that corresponds to calcium carbonate (left frames in Fig. 4). Comparison of these results with Fig. 5 in (Changwen et al.,) (which was based on non-normalized spectra) shows that within-type variability was greatly reduced by normalization.

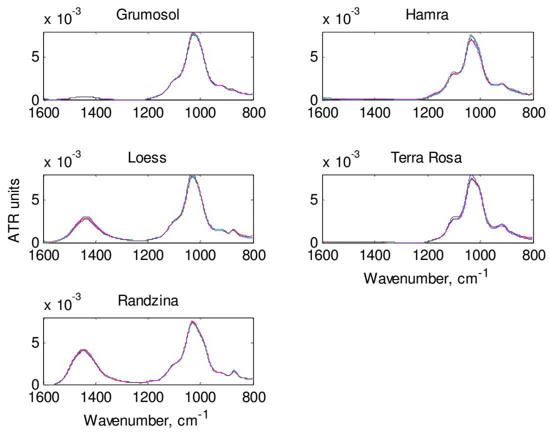

Figure 7. Typical ATR soil spectra after water subtraction, baseline correction and normalization

Three types of classifiers were investigated: (1) based on the 800-1200 cm-1 interval only, (2) based on the 1250-1550 cm-1 interval only, and (3) based on the whole 800-1550 cm-1 interval. In each case numerous classifiers based of various levels of wavelet decomposition, number of coefficients used as NN input and number of NN hidden nodes, were tested. Table 2 presents the results obtained for the NNs that had the highest classification rate for the validation samples. Regardless of the spectral interval used, very good results are obtained and all the validation samples are correctly classified except some Hamra samples that are classified as Terra Rosa. Such misclassification is not surprising due to the very high similarity of the spectra of both soils (Fig. 4, right frames). However, as pointed out by Linker et al. (Linker et al., 2006), this is not a serious shortcoming since in practice these two soils are not found in the same regions. Furthermore, the water contents of the saturated pastes of these soils are very different (due to difference in clay content), so that it is possible to ensure perfect classification by adding the paste water content to the NN inputs (not shown).

Table 2. Classification results based on ATR spectra (validation spectra only)

Spectral interval	1250-1550 cm-1	800-1200 cm-1	800-1600 cm-1
Decomposition level	5	5	5
Number of coefficients selected	4	5	5
Number of NN hidden nodes	4	4	4
	Percentage correct classification		
Grumosol	100	100	100
Loess	100	100	100
Randzina	100	100	100
Hamra	91	91	96
Terra Rosa	100	100	100

3.2 Photoacoustic spectroscopy

Figure 5 shows five typical PAS spectra for each soil type. Numerous bands can be observed in the 600-2000 cm-1 interval, and around 2550 cm-1, 2900 cm-1 and 3700 cm-1. Linker et al. (Changwen et al.,) showed that these bands can be associated with soil constituents such as clays, carbon carbonate and organic matter. It must be noted that while for ATR spectra only the 800-1600 cm-1 interval yielded useful information (due to the presence of water), the PAS spectra include useful bands throughout the whole spectral range.

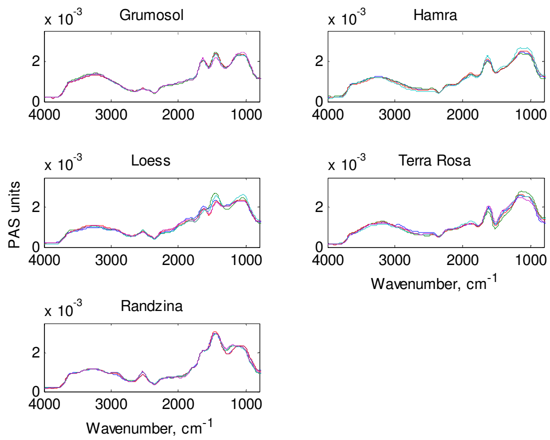

Figure 8. Typical normalized PAS soil spectra

The results of the classification procedure are summarized in Table 3. Again, various classifiers were tested, and only the best ones are reported in the Table. The best results were obtained when using the whole spectrum, in which case only a few Hamra samples were incorrectly classified.

Table 3. Classification results based on photoacoustic spectra (validation spectra only)

Spectral interval	800-4000 cm-1	800-2700 cm-1	800-2300 cm-1
Decomposition level	5	10	10
Number of coefficients selected	10	20	20
Number of NN hidden nodes	4	4	4
	Percentage correct classification		
Grumosol	100	100	100
Loess	100	100	93
Randzina	100	100	100
Hamra	96	95	96
Terra Rosa	100	100	85

4. ONCLUSION

Mid-infrared attenuated total reflectance and photoacoustic spectroscopy both appear to be very promising techniques for rapid analysis of soil samples. The main limitation of the ATR approach is that it requires samples close to water saturation. As a result, the spectral range that yields useful information is rather limited. By comparison, PAS measurements are conducted with air-dried samples, and useful absorbance bands are observed throughout the whole spectrum. The bands are also sharper and more clearly defined than the ATR ones.

Both types of spectra can be used for soil classification, which requires data reduction and classification. A new method based on wavelet decomposition and neural network classification has been developed, which results in correct classification of over 95% of the validation samples.

REFERENCES:

Alsberg B. K., Woodward A. M., Winson M. K., Rowland J. J. and Kell D. B., Variable selection in wavelet regression models, Analytica Chimica Acta, Vol. 368, 1998, pp. 29-44.

Ben-Dor E. and Banin A., Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties, Soil Science Society of America Journal, Vol. 59, 1995, 364-372.

Changwen D., Linker R. and Shaviv A. Characterization of soils using photoacoustic spectroscopy, Applied Spectroscopy (Accepted)

Changwen D., Linker R. and Shaviv A. Soil identification using mid-infrared photoacoustic spectroscopy. Submitted to Geoderma

Daniel K. W., Tripathi N. K. and Honda K., Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand), Australian Journal of Soil Research, Vol. 41, 2003, pp. 47-59.

Depczynski U., Jetter K., Molt K. and Niemoller, A., Quantitative analysis of near infrared spectra by wavelet coefficient regression using a genetic algorithm, Chemometrics and Intelligent Laboratory Systems, Vol. 47, 1999, pp. 179-187.

Ehrentreich F., Wavelet transform applications in analytical chemistry, Analytical and Bioanalytical chemistry, Vol. 372, 2002, pp. 115-121.

Figueiredo dos Santos R. N., Galvao R. K. H., Araujo M. C. U. and Cirino da Silva E., Improvement of prediction ability of PLS models employing the wavelet packet transform: A case study concerning FT-IR determination of gasoline parameters, Talanta, Vol. 71, 2007, pp. 1136-1143.

Gerzabek M. H., Antil R. S., Kogel-Knabner I., Knicker H., Kirchmann H. and Haberhauer G., How are soil use and management reflected by soil organic matter characteristics: A spectroscopic approach, European Journal of Soil Science, Vol. 57, 2006, pp. 485-494.

Haberhauer G. and Gerzabek M. H., DRIFT and transmission FT-IR spectroscopy of forest soils: an approach to determine decomposition process of forest litter, Vibrational Spectroscopy, Vol. 19, 1999, pp. 413-417.

Haberhauer G., Rafferty B., Strebl F. and Gerzabek M. H., Comparison of the composition of forest soil litter derived from three different sites at various decompositional stages using FTIR spectroscopy, Geoderma, Vol. 83, 1998, pp. 331-342.

Haykin S., Neural networks: A comprehensive foundation, Prentice-Hall, 1999.

Jahn B. R., Linker R., Upadhyaya S. K., Shaviv A., Slaughter D. C. and Shmulevich I. Mid infrared spectroscopic determination of soil nitrate content. Biosystems Engineering, Vol. 94, 2006, pp. 505-515.

Janik L. J., Merry R. H. and Skjemstand J. O., Can mid infrared diffuse reflectance analysis replace soil extractions, Australian Journal of Soil Research, Vol. 38, 1998, pp. 681-696.

Jolliffe I. T., Principal Component Analysis, Springer-Verlag, New York, 1986.

Leung A. K., Chau F. T., Gao J. B. and Shih T. M. Application of wavelet transform in infrared spectrometry: spectral compression and library search, Chemometrics and Intelligent Laboratory Systems, Vol. 43, 1998, pp. 69-88

Linker R., Kenny A., Shaviv A., Singher L. and Shmulevich I. FTIR/ATR nitrate determination of soil pastes using PCR, PLS and cross-correlation, Applied Spectroscopy, Vol. 58, 2004, pp. 516-520.

Linker R., Shmulevich I., Kenny A. and Shaviv A., Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy, Chemosphere, Vol. 61, 2005, pp. 652-658.

Linker R., Weiner M., Shmulevich I. and Shaviv A., Nitrate determination in soil pastes using FTIR-ATR mid-infrared spectroscopy: Improved accuracy via soil identification, Biosystems Engineering, Vol. 94, 2006, pp. 111-118.

Liu Y. and Brown S. D., Wavelet multiscale regression from the perspective of data fusion: new conceptual approaches. Analytical and Bioanalytical chemistry, Vol. 380, 2004, pp. 445-452.

McBratney A. B., Minasny B., Viscarra Rossel R. A., Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis, Geoderma, Vol. 136, 2006, pp. 272-278.

McCarty G. W., Reeves III J. B., Follett R. F. and Kimble J. M., Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement, Soil Science Society of America Journal, Vol. 66, 2002, pp. 640-646.

McClelland J. F., Jones R. W. and Bajic S. J., Photoacoustic Spectroscopy, In Handbook of Vibrational Spectroscopy (Volume II), J. M. Chalmers and P. R. Griffiths, Eds., Wiley & Sons, 2001.

Nguyen T. T., Janik L. J.and Raupach M., Diffuse reflectance infrared Fourier trasnform (DRIFT) spectroscopy in soils studies, Australian Journal of Soil Research, Vol. 29, 1991, pp. 49-67.

Shaviv A., Kenny A., Shmulevich I., Singher L. Reichlin Y. and Katzir A. IR fiberoptic systems in situ and real time monitoring of nitrate in water and environmental systems, Environmental Science & Technology, Vol. 37, 2003, pp. 2807-2812.

Trygg J. and Wold S., PLS regression on wavelet compressed NIR spectra, Chemometrics and Intelligent Laboratory Systems, Vol. 42, 1998, pp. 209-220.

Viscarra Rossel R. A. and McBratney A. B., Soil chemical analytical accuracy and costs: Implications from precision agriculture, Australian Journal of Experimental Agriculture, Vol. 38, 1998, pp. 765-775.

Viscarra Rossel R. A., Walvoort D. J. J., McBratney A. B., Janick L. J. and Skjemstad J. O., Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties, Geoderma, Vol. 131, 2006, pp. 59-75.

Walczak B. and Massart D. L., Noise suppression and signal compression using the wavelet packet transform, Chemometrics and Intelligent Laboratory Systems, Vol. 36, 1997, pp. 81-94.

Zhang X., Jin J., Zheng J. and Gao H., Genetic algorithms based on wavelet transform for resolving simulated overlapped spectra, Analytical and Bioanalytical Chemistry, Vol. 377, 2003, pp. 1153-1158.