Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties

Cheng-Wen Chang, David A. Laird,* Maurice J. Mausbach, and Charles R. Hurburgh, Jr.

ABSTRACT

A fast and convenient soil analytical technique is needed for soil quality assessment and precision soil management. The main objective of this study was to evaluate the ability of near-infrared reflectance spectroscopy (NIRS) to predict diverse soil properties. Near-infrared reflectance spectra, obtained from a Perstrop NIR Systems 6500 scanning monochromator (Foss NIRSystems, Silver Spring, MD), and 33 chemical, physical, and biochemical properties were studied for 802 soil samples collected from four Major Land Resource Areas (MLRAs). Calibrations were based on principal component regression (PCR) using the first derivatives of optical density [log(1/R)] for the 1300- to 2500-nm spectral range. Total C, total N, moisture, cation-exchange capacity (CEC), 1.5 MPa water, basal respiration rate, sand, silt, and Mehlich III extractable Ca were successfully predicted by NIRS ($r^2 > 0.80$). Some Mehlich III extractable metals (Fe, K, Mg, and Mn) and exchangeable cations (Ca, Mg, and K), sum of exchangeable bases, exchangeable acidity, clay, potentially mineralizable N, total respiration rate, biomass C, and pH were also estimated by NIRS but with less accuracy ($r^2 = 0.80~0.50$). The predicted results for aggregation (wt% > 2, 1, 0.5, 0.25 mm, and macroaggregation) were not reliable ($r^2 = 0.46~0.60$). Mehlich III extractable Cu, P, and Zn, and exchangeable Na could not be predicted using the NIRS–PCR technique ($r^2 < 0.50$). The results indicate that NIRS can be used as a rapid analytical technique to simultaneously estimate several soil properties with acceptable accuracy in a very short time.

S TANDARD PROCEDURES for measuring soil properties are complex, time-consuming, and expensive. A rapid and economical soil analytical technique is needed before farmers and land managers will be able to fully utilize soil testing as an aid in precision farming and both assessment and management of soil quality.

Near-infrared reflectance spectroscopy is a nondestructive analytical technique for studying interactions between incident light and a material's surface. Near-infrared reflectance spectroscopy is widely used in industry due to its simplicity, rapidness, and the need for little or no sample preparation. The technique was first developed more than three decades ago for rapid moisture analysis of grain (Ben-Gera and Norris, 1968). Now NIRS is the dominant analytical technique used for grain and forage quality assessments. Near-infrared reflectance spectroscopy is also used as an identification and/or characterization technique in the polymer, pharmaceutical, petrochemical, and other industries (Wetzel, 1983; Creaser and Davies, 1988a; Murray and Cowe, 1992; Workman, 1996).

Near-infrared spectra are dominated by weak overtones and combinations of fundamental vibrational bands for H-C, H-N, and H-O bonds from the mid-infrared region. In addition to the chemistry of a material, near-infrared spectra are also influenced by the physical structure of a material. The size and shape of the particles, the voids between particles, and the arrangement of particles affect the length of the light transmission passing through a sample and thereby influence reflectance (Wetzel, 1983). Near-infrared spectra are difficult to interpret directly because of the overlap of weak overtones and combinations of fundamental vibrational bands. As a result, multivariate calibration is required for quantitative analysis of sample constituents by NIRS.

Various calibration methods have been used to relate near-infrared spectra with measured properties of materials. Principal components regression, partial least squares regression (PLSR), stepwise multiple linear regression (SMLR), Fourier regression, locally weighted regression (LWR) and artificial neural networks are the most used multivariate calibration techniques for NIRS (Creaser and Davies, 1988b; Naes and Isaksson, 1990; Holst, 1992). None of these proposed calibration techniques have achieved universal acceptance because a calibration model that works well for one application may be unacceptable for another.

Several studies have focused on measuring soil properties using NIRS. Bowers and Hanks (1965) concluded that reflectance spectra are strongly affected by soil moisture content, organic matter content, and particle size. Al-Abbas et al. (1972) found that light reflectance is negatively correlated with organic matter and clay content. Krishnan et al. (1980) compared analyses based on reflectance at two wavelengths in the visible light region (623.3 and 564.4 nm) and two wavelengths in the near-infrared region (1136.0 and 1398.4 nm) for predicting soil organic matter content of 12 soils from Illinois ($r^2 = 0.92$ for visible light, and $r^2 = 0.87$ for infrared light). One hundred forty-four samples collected from three soil series in Queensland, Australia (six depth-intervals for eight profiles of each soil series) were analyzed in the study of Dalal and Henry (1986). They used reflectance at three wavelengths, in the form of log(1/R), to predict total organic C, total N, and

moisture content by multiple linear regression (r^2 = 0.86, 0.86, and 0.92, respectively). Using similar approaches, Morra et al. (1991) concluded that NIRS could be used to predict soil C and N content.

Henderson et al. (1992) found that reflectance of organic matter extracted from four Indiana agricultural soils strongly correlated with organic C content and significantly responded to concentrations of Fe and Mn oxides in the visible range for soils having the same parent material. Fritze et al. (1994) studied reflectance spectra of fire- or wood ash-treated humus samples, and found that the first two regression components (based on PLSR) obtained from spectroscopic characterization could explain 60 to 80% of the biological variation (soil respiration and microbial biomass). The prediction ability of NIRS for soil N availability was compared with other tests by Fox et al. (1993). A portable near-infrared spectrophotometer, designed by Sudduth and Hummel (1993a, 1993b), was used to predict soil organic matter ($r^2 > 0.85$), moisture ($r^2 > 0.94$), and CEC ($r^2 > 0.85$) in soils from Illinois. They also concluded that the predictions of these soil properties became less accurate as the geographic range of samples increased (Sudduth and Hummel, 1996).

Most NIRS studies related to prediction of soil properties were limited in one way or another due to small sample size, less diversity in soil types, and/or because only a few soil properties were tested. In this study, a diverse group of 802 soil samples collected from four MLRAs were analyzed by NIRS. The objectives of this study were (i) to evaluate the ability of NIRS to predict various soil properties and (ii) to understand why NIRS can be used to predict some soil properties that theoretically do not respond to light in the near-infrared region.

MATERIALS AND METHODS

Soil Quality-Natural Resource Inventory (SQ-NRI) Pilot Project

The USDA-NRCS Soil Quality Institute and the USDA-ARS National Soil Tilth Laboratory initiated a collaborative project for "Measuring and Assessing Soil Quality via the National Resource Inventory Sampling Frame" (Brejda et al., 2000a,b). The goal of this project was to test the feasibility of sampling soils and measuring selected soil properties within the framework of the Natural Resource Inventory. Surface (0~3 or 0~10 cm) and subsurface (3~10 or 10~30 cm) soils were collected from 448 sites in the four MLRAs, including MLRA 9 (the Palouse and Nez Perce Prairies located in eastern Washington and western Idaho), MLRA 67 (the Central High Plains located in eastern Colorado, southeastern Wyoming, and western Nebraska), MLRA 77 (the Southern High Plains located in New Mexico and the panhandle of Texas), and MLRA 105 (the Northern Mississippi Valley Loess Hills, located in northeastern Iowa, southeastern Minnesota, and southwestern Wisconsin). More than 800 soil samples were collected and field observations were recorded at each site. The soil samples were analyzed for physical, chemical, and biological properties by the cooperating laboratories, including the NRCS, Soil Characterization Laboratory in Lincoln, NE, and four ARS and university laboratories located in Ames, IA; Pullman, WA; St. Paul, MN; and Bushland, TX. The soil properties tested by NIRS and the analytical methods used are summarized in Table 1. The basic statistics of soil properties are summarized in Table 2. Some samples did not have complete soil property data, so the real number of samples used in the NIRS analysis could be smaller than 800.

Near-Infrared Reflectance Spectroscopy Measurement

The spectral reflectance of soil samples collected for SQ-NRI Pilot Project was measured in the visible and near-infrared regions with a Perstorp NIRSystems 6500 scanning monochromator. For each sample, ≈100 g of air-dried soil was packed into a natural product cell having a quartz window, and the spectral reflectance, averaged from 25 spectra, was recorded as the logarithm of the inverse of the reflectance, or optical density [$\log(1/R)$], at 2-nm intervals from 400 to 2498 nm. In order to evaluate the application of NIRS for soil testing under more realistic and practical conditions, the <8-mm sieved soil samples were analyzed by NIRS for this study. Although some soil properties were measured using samples passing a 2-mm sieve while the near-infrared spectra were obtained with samples passing an 8-mm sieve, the discrepancy should have little

Table 1. Soil properties measured and analytical methods used in the SQ-NRI Pilot Project.

Soil property	Method	Reference
Total C	Dry combustion using a LECO SC-444 analyzer (g kg^{-1})	USDA-NRCS, 1996
Total N	Dry combustion using a LECO FP-438 analyzer (g kg^{-1})	USDA-NRCS, 1996
Moisture	Gravimetric water content (kg kg^{-1})	
Potentially mineralizable N	35-d incubation at 25°C (mg N kg^{-1})	Drinkwater et al., 1996
Respiration rate: total and basal	35-d incubation at 25°C/CO$_2$ trapping (mg CO$_2$–C kg^{-1} d^{-1})	Drinkwater et al., 1996
Biomass C	Fumigation/extraction (mg kg^{-1})	Tate et al., 1988
CEC	NH$_4$OAc cation-exchange capacity at pH 7 (cmol$_c$ kg^{-1})	USDA-NRCS, 1996
pH-H$_2$O	pH measures at 1:1 soil/water suspension	USDA-NRCS, 1996
pH-CaCl$_2$	pH measured at 1:2 soil/CaCl$_2$ suspension	USDA-NRCS, 1996
1.5 MPa water	15 MPa water (wilting point) on air-dry soil (kg kg^{-1})	USDA-NRCS, 1996
Aggregation: macro, 2, 1, 0.5, and 0.25 mm	Wet aggregate stability, percentage (w/w) of aggregates retained on 2-, 1-, 0.5-, 0.25-mm sieve, respectively. Macro-aggregation is the sum of the percentage (w/w) of 2- to 0.25-mm aggregates	Kemper and Rosenau, 1986
Particle-size distribution: clay, silt, and sand	Pipette method (%)	Kemper and Rosenau, 1986
Mehlich III extractable metals: Ca, Cu, Fe, K, Mg, Mn, Zn, and P	Extraction with Mehlich III followed by ICP-AES analysis (mg kg^{-1})	Mehlich, 1984
Exchangeable bases: Ca, Mg, Na, and K	NH$_4$OAc extraction (cmol kg^{-1}), measured using a Perkin-Elmer AA-5000	USDA-NRCS, 1996
Sum of extractable bases	Sum of NH$_4$OAc extractable bases (cmol kg^{-1})	USDA-NRCS, 1996
Exchangeable acidity	NH$_4$OAc extractable acidity measured by BaCl$_2$–triethanolamine solution buffered at pH 8.2 and back titrated with HCl (cmol kg^{-1})	USDA-NRCS, 1996

affect on the NIRS predictions as long as the procedures were consistent.

Near-Infrared Reflectance Spectroscopy Prediction

A PCR technique was used to relate the near-infrared reflectance spectra to the measured soil properties. The NIRS–PCR technique, summarized in Fig. 1, contained four steps: (i) data pretreatment, (ii) calibration set selection, (iii) principal component regression and prediction, and (iv) prediction ability evaluation.

Data Pretreatment

The spectra in the range of 1300 and 2500 nm were used to correlate with the soil properties. The spectral reflectance data (optical density) were reduced by averaging every five adjacent spectral points to yield 120 new spectral data points, each representing a 10-nm interval. The first derivatives of the reflectance spectra were then calculated using

$$d_i = x_{i+1} - x_i, i = 1, 2,...., 119 \qquad [1]$$

where d_i represents the first derivative of ith data point (x_i) in the reduced NIR spectral data.

The selected range of spectra includes three major reflectance peaks; the intense peak in the 400- and 1300-nm range was not included because it was found to reduce the accuracy of predicted soil properties using this calibration model.

Calibration Set Selection

In order to evaluate the ability of NIRS to predict soil properties, spectral data for all soil samples were treated the same during the computation procedures. A database for all 802 samples containing the transformed and truncated first derivative spectra as well as measured values for the soil properties was established. The calibration strategy used in this study was to first select a small ($n = 30$) subset of similar (having similar spectra) samples for each sample being tested, and then to use PCR to predicted the value of the soil property for the test sample.

The 30 samples used in the calibration set for a given sample were selected from all of those in the database (excluding the sample being tested) based on the shortest squared Euclidean distance to the derivative reflectance spectra for the test sample. Squared Euclidean distance between two reflectance spectra can be obtained by

$$D^2_{AB} = \sum_{i=1}^{p} (d_{Ai} - d_{Bi})^2 \qquad [2]$$

where d_{Ai} and d_{Bi} are the ith points of the first derivative spectra for Soil A and Soil B, respectively; and p is the number of data points in the spectra.

Principal Components Regression and Prediction

For the calibration set, the values of a soil property ($\mathbf{y}_{30\times1}$) can be expressed as a linear function of their spectral reflectance, such as

$$\mathbf{y}_{30\times1} = \mathbf{1}_{30\times1}b_0 + \mathbf{C}_{30\times119}\mathbf{b}_{119\times1} + \mathbf{e}_{30\times1} \qquad [3]$$

where $\mathbf{1}$ is a 30×1 vector containing ones; b_0 and $\mathbf{b}$ are coefficients; $\mathbf{e}$ is the error vector. Because NIR spectral data points have high colinearity, a few principal components obtained from the spectral data matrix are employed to replace the original spectra matrix in Eq. [1]. The new regression

Table 2. Sample size, mean, and the range of soil properties obtained from SQ-NRI Pilot Project.

Soil property	n	Mean	Min.	Max.	SD
Total C, g kg^{-1}	743	26.5	1.3	285.8	25.8
Total N, g kg^{-1}	744	2.3	0.2	16.5	1.8
Moisture, kg kg^{-1}	802	0.02	0.00	0.16	0.01
Potentially mineralizable N, mg N kg^{-1}	764	52.6	0.4	555.1	55.8
Total respiration rate, mg CO$_2$-C kg^{-1} d^{-1}	786	282.1	0.3	2 792.2	418.6
Basal respiration rate, mg CO$_2$-C kg^{-1} d^{-1}	786	45.0	0.3	884.7	69.9
Biomass, mg kg^{-1}	745	602.9	3.7	22 337.6	989.2
CEC, cmol$_c$ kg^{-1}†	744	18.7	1.9	92.2	9.7
15 MPa water, kg kg^{-1}	744	0.11	0.01	0.62	0.06
Ph (H$_2$O)	743	6.3	3.9	8.5	0.8
pH (CaCl$_2$)	743	5.9	3.5	8.0	0.8
Aggregation (%)					
2.00 mm	763	6.5	0.0	38.4	6.4
1.00 mm	764	7.4	0.0	41.6	7.4
0.50 mm	763	10.7	0.0	38.1	7.6
0.25 mm	765	15.6	0.4	54.8	7.9
Macro	765	40.1	2.2	89.8	22.2
Particle-size distribution (%)					
Clay	743	17.0	0.7	35.2	7.0
Silt	743	53.4	3.1	85.3	23.9
Sand	743	29.6	1.2	95.2	27.8
Mehlich III extractable metal, mg kg^{-1}					
Ca	784	2265.1	87.7	12 763.4	1366.9
Cu	761	4.8	0.3	130.6	10.5
Fe	784	209.3	5.0	646.0	110.6
K	784	326.3	16.2	1 757.2	237.2
Mg	784	423.9	17.4	2 886.3	282.9
Mn	784	141.2	0.9	675.9	109.2
P	779	49.4	0.7	507.6	49.8
Zn	762	10.2	0.1	373.3	32.9
Exchangeable base, cmol kg^{-1}					
Ca	744	13.0	0.2	98.7	8.8
Mg	742	3.7	0.1	17.6	2.4
Na	439	0.2	0.1	1.8	0.2
K	742	0.9	0.1	4.8	0.6
Sum of exchangeable bases, cmol kg^{-1}	744	17.7	0.4	115.4	10.9
Exchangeable acidity, cmol kg^{-1}	730	6.5	0.1	54.0	4.6

† Cation-exchange capacity.

coefficients can be determined by a least square method and then used to estimate the properties for a test sample. The details of PCR can be found in *Multivariate Calibration* (Martens and Naes, 1989).

Prediction Ability

Up to 20 principal components were used in the regression calculations, the number of principal components (F) was chosen to give the smallest root mean standard error of cross validation (RMSECV), largest r^2, and a slope close to one for the regression line between predicted and measured values of soil properties. The ability of the NIRS–PCR technique to predict a soil property was evaluated using r^2, RMSECV, and the ratio of standard deviation to RMSECV (RPD).

In this study, eigenvectors were obtained from singular value decompositions, and all calculations were performed using MATLAB (The Mathworks, Inc., Natick, MA).

Property Equation

Near-infrared reflectance spectra are directly influenced by combinations and overtones of fundamental vibrations for organic functional groups and water, particle size, and surface properties. Therefore, total C, total N, moisture content, particle size, and aggregation are considered primary properties, because a theoretical basis for NIRS predictions of these properties is apparent. Other soil properties are considered secondary properties. If a secondary property is correlated with one or more of the primary properties, the secondary property may also be predicted by NIRS even though the secondary property does not respond directly to near-infrared light. For instance, potentially mineralizable N is related to total organic C and total N, and CEC is correlated with clay abundance and organic matter content. Previous studies have found that
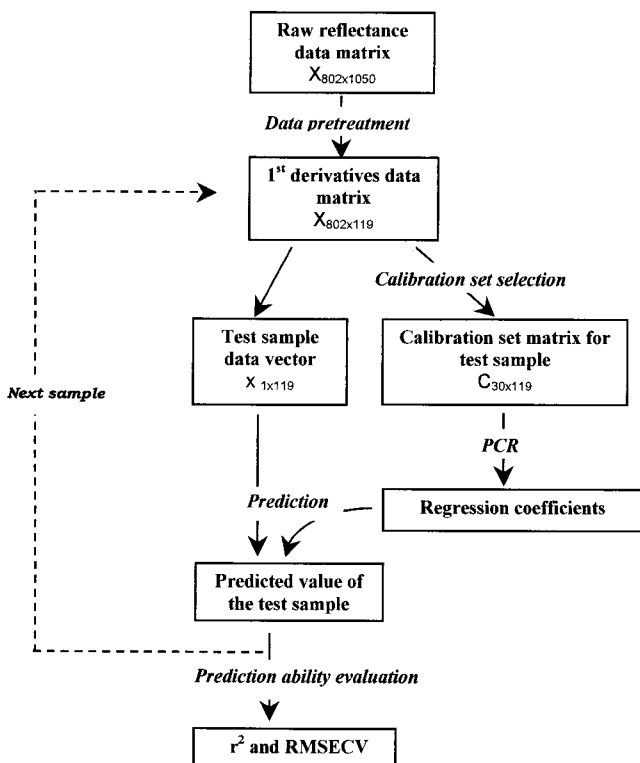
NIRS can predict some soil properties not theoretically related to near-infrared light (Fox et al., 1993; Fritze et al., 1994; Ben-Dor and Banin, 1995). This phenomenon was studied by comparing the predictive ability of the NIRS–PCR technique for secondary soil properties to their level of correlation with primary soil properties.

RESULTS AND DISCUSSION

Near-Infrared Reflectance Spectroscopy of Soil Samples

All soils tested in this study had similar visible and near-infrared reflectance spectra. They had high optical density [$\log(1/R)$] in the visible light region (400~800 nm), and three major absorption peaks (around 1400, 1900, and 2200 nm) in the near-infrared light region. Mean spectra for soils from the four MLRAs shown in Fig. 2 were obtained using only samples with low moisture content (<0.01 kg kg^{-1}) in an effort to reduce the influence of water and to emphasize the essential characteristics of the reflectance spectra. In the near-infrared region, the mean spectra for MLRA 67 and MLRA 77 had more intense peaks than the mean spectra for MLRA 9 and MLRA 105. The mean spectrum of MLRA 77 had the highest optical density in the visible light range. Mean spectra for MLRA 9 and MLRA 105 had different basal lines, but both had weak peaks in the near-infrared light region.

Soils from the different MLRAs could be separated from each other by the distribution of their first three principal components (PCs) generated from the first derivatives of optical density for the 1300- to 2500-nm range. Figure 3 demonstrates both the power of PC analysis as a means of discriminating near-infrared reflectance spectra of soils and a tendency for regional similarity in NIR spectra of soils.

The size discrepancy between samples used for the

Fig. 1. Summary of the near-infrared reflectance spectroscopy–principal components regression (NIRS–PCR) technique. RMSECV is root mean standard error of cross validation.

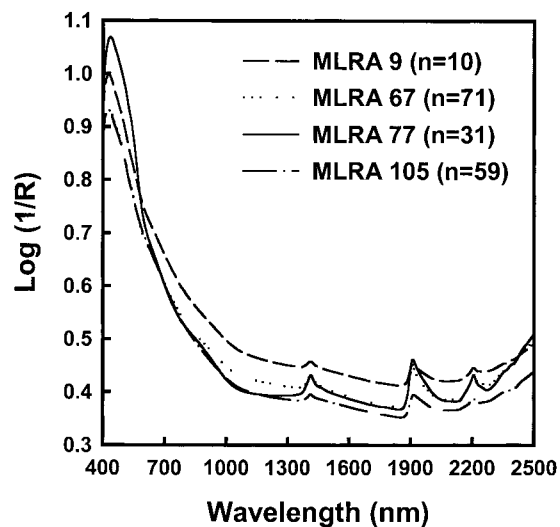Fig. 2. The average near-infrared spectra for soils containing low moisture content (<0.01 kg kg^{-1}) from four Major Land Resource Area (MLRA) groups. The total C (g kg^{-1}) and total N (g kg^{-1}) of these samples are as follows: MLRA 9, total C = 14.5, total N = 1.3; MLRA 67, total C = 7.9, total N = 0.9; MLRA 77, total C = 4.5, total N = 0.6; MLRA 105, total C = 17.3, total N = 1.6.
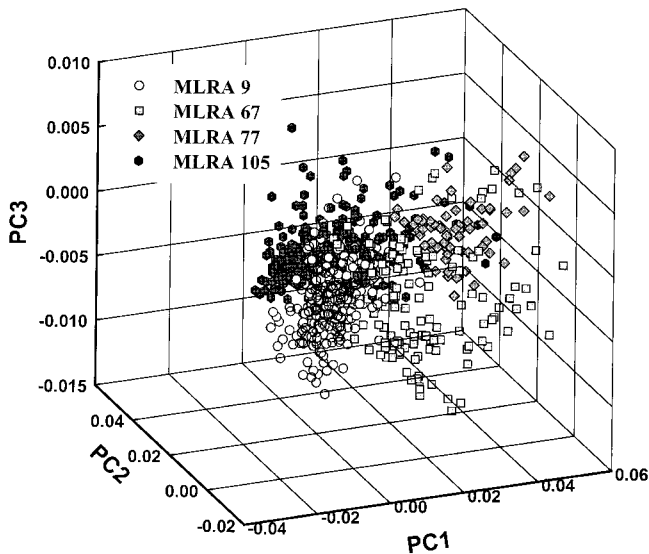
Fig. 3. Distribution of the first three principal components (PC1, PC2, and PC3) for soils from four Major Land Resource Areas (MLRAs). The PCs are based on the first derivatives of the optical density [log(1/R)] from 1300 to 2500 nm.

soil property measurements (sieved to <2 mm for some properties) and NIRS analysis (sieved to <8 mm) should only have a small influence on NIRS predictions. However, using <8-mm soils may have greater variation in surface physical properties (such as size and shape of soil aggregates) than <2-mm soils; surface physical properties affect the NIRS analysis because of their influence on the light scattering and light path lengths. The effect of sample physical properties may be partially but not entirely removed by spectral data pretreatments. As a result, NIRS analysis is usually more accurate when samples have similar particle size.

The first step in the calibration procedure used in the NIRS–PCR technique was the selection of 30 neighboring samples for the calibration set, based on the smallest squared Euclidean distance between the measured first derivative spectra for the first derivative test samples and the spectra for any of the other samples in the database. Generally, soils from the same area have similar soil properties (such as size and shape of aggregates) and similar responses to incident light. Therefore, selecting similar spectra should help to reduce the influence of physical variations. One means of testing the

viability of this approach was to determine the extent of regional bias that occurred during selection of samples for the calibration set. To do so, we defined q as the probability that samples in the calibration set are from the same MLRA as the test sample

$$q = n/30 \qquad [4]$$

where n is the number of samples in the calibration set that are from the test sample's MLRA. We further defined Q as the adjusted population ratio for the test sample's MLRA (i.e., the expected value of q if the samples for the calibration set were randomly selected). Results in Table 3 show that the average q obtained by selecting the calibration set using raw spectral data is smaller than q obtained by selecting the calibration set using the first derivative spectra. The results indicate that the first derivative spectra provide more effective information for selecting similar samples. Furthermore, the large discrepancies between q and Q for MLRA 77 and MLRA 67 obtained using the first derivative data suggest that the spectra for soils from these MLRAs are unique relative to spectra for soils from the other MLRAs. However, the value of q for MLRA 77 is reduced slightly when only spectra data from near-infrared range are considered.

The differences in average spectral features for soils from the four MLRAs are due to differences in soil properties. Samples collected from MLRA 77 (Texas and New Mexico) had very low organic matter content (average total C is 4 g kg^{-1}) and had an orange to reddish color. By contrast, most soils from MLRA 9 (Washington and Idaho) and MLRA 105 (Iowa, Minnesota, and Wisconsin) had higher organic matter content (average total C is 28 g kg^{-1} and 34 g kg^{-1}, respectively) and appeared brownish in color. Other factors, such as soil structure and texture also probably contributed to differences in the average spectra.

Soil Properties and Reflectance Spectra

Soil is a heterogeneous mixture of various chemical compounds, and thus a unique spectral response for soil properties is by no means certain. One method of evaluating the spectral response for a soil property is to study the relationship between wavelength, optical density, and the values of the soil property. The relation-

Table 3. Regional bias of near-infrared reflectance spectroscopy–principal components regression technique in selection of samples for inclusion in calibration sets for the raw spectral data, log(1/R), and the first derivatives of raw spectra.†

	MLRA 9	MLRA 67	MLRA 77	MLRA 105
Q‡, expected value	172/801 = 0.34	127/801 = 0.16	63/804 = 0.08	337/801 = 0.42
q§, based on raw data				
λ = 400~2500 nm	0.58 (0.18)¶	0.41 (0.15)	0.79 (0.21)	0.55 (0.26)
λ = 400~1300 nm	0.54 (0.19)	0.32 (0.14)	0.86 (0.17)	0.52 (0.22)
λ = 1300~2500 nm	0.63 (0.19)	0.35 (0.13)	0.34 (0.16)	0.56 (0.28)
q, based on first derivative data				
λ = 400~2500 nm	0.71 (0.23)	0.76 (0.22)	0.98 (0.13)	0.73 (0.21)
λ = 400~1300 nm	0.64 (0.22)	0.50 (0.16)	0.97 (0.13)	0.65 (0.22)
λ = 1300~2500 nm	0.73 (0.25)	0.61 (0.26)	0.75 (0.22)	0.79 (0.18)

† MLRA is major land resource area.
‡ The Q values are expected if the selection of spectra for inclusion in the calibration set were random.
§ The q values are average proportions of spectra selected for the calibration sets from the same MLRA as the test sample. Regional bias is indicated by q values that are larger than Q values.
¶ Values in the parentheses represent the standard deviations.

ship between the first derivatives of optical density and measured values for some soil properties (total C, total N and moisture content, biochemical properties, and macroaggregation and particle size) in the visible–near-infrared light region are shown in Fig. 4. Total C and total N had similar responses, but the response for soil moisture was unique. Carbon and N are major elements in soil organic matter, and quantities of C and N in soils are highly correlated. Some biochemical soil properties (such as potential mineralizable N, respiration rate, and biomass C) and CEC also have similar spectral responses, though less intense, to that of total organic C and total N. The similarity in spectral response for C, N, and other biological properties is probably due to their high intercorrelation (Brejda et al., 2000a,b). In contrast, water contains -OH groups, which clearly have a unique spectral response in the near-infrared region. Soil particle-size distribution and macroaggregation also exhibited a unique spectral response to incident light.
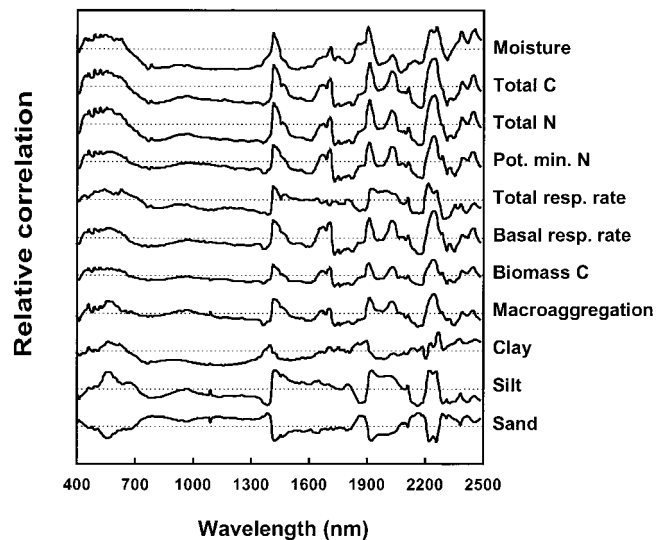

Fig. 4. The influence of wavelength on the correlation between reflectance intensity, $\log(1/R)$, and values of some soil properties. Dotted lines are lines of zero correlation and are separated by ± 1 from each other.

Near-Infrared Reflectance Spectroscopy Prediction for Soil Properties

The NIRS–PCR technique involves data reduction procedures, in which some spectral information is sacrificed. In this study each test sample had a different calibration set, thus the principle components (PCs) of each calibration set account for different amounts of the total variance. The first 10 PCs of all calibration sets account for 97 to almost 100% of the total variance (mean = 99.5%). Therefore, using fewer PCs as independent variables in this NIR–PCR technique is acceptable.

The ability of NIRS to predict the 33 soil properties is summarized in Table 4. Samples having difference between measured and predicted values larger than three times RMSECV were considered outliers and were excluded from the analysis. For most tested soil properties, more than 97.3% of soil samples were included for evaluating the prediction ability of NIRS. Scatter plots of predicted vs. measured values for some properties are shown in Fig. 5.

The ability of NIRS to predict values of soil properties can be grouped into three categories based on RPD values (Table 4). Category A (RPD > 2.0) includes properties with measured vs. predicted r^2 values between 0.80 and 1.00. The soil properties in category A include total C, total N, moisture, silt, sand, CEC, 1.5 MPa water, Mehlich III extractable Ca, and basal respiration rate. Category B (RPD = 1.4~2.0) includes soil properties with measured vs. predicted r^2 values between 0.50 and 0.80. This group includes clay, potentially mineralizable N, total respiration rate, biomass C, pH, some wet aggregation measures (macro-, 1.0, and 0.5 mm), some Mehlich III extractable metals (Fe, K, Mg, Mn), NH$_4$OAc extractable cations (Ca, Mg, K), and titratable acidity. Mehlich III extractable Cu, P, and Zn; NH$_4$OAc extractable Na; and some wet aggregation measurements (2 and 0.25 mm) are in Category C ($r^2 <$ 0.50, RPD < 1.4). The authors believe that prediction of soil properties in Category B can be improved by

using different calibration strategies, but properties in Category C may not be reliably predicted using NIRS.

The results indicate that total C, total N, and moisture content were readily and accurately estimated ($r^2 >$ 0.84, RPD > 2.47). The slopes of the measured vs. predicted regression lines for total C, total N, and moisture content are not significantly different from 1 which indicates that NIRS–PCR did not tend to over- or underestimate these properties. In other studies of NIRS analysis of soils, the r^2 values for measured vs. predicted water content are higher than 0.9 (Dalal and Henry, 1986; Sudduth and Hummel, 1993). The lower r^2 value obtained in this study was probably due to the small range in moisture content of the air-dried samples (range = 0.00~0.17, mean = 0.02 kg kg^{-1}).

The size, shape, and arrangement of particles in a sample can affect the path of light transmission and the reflectance spectra, as such NIR spectra are assumed to exhibit a primary response to soil texture and structure. Based on the results, NIRS can be used to predict measured sand and silt content ($r^2 >$ 0.82; RPD > 2.32) more accurate than measured clay ($r^2 =$ 0.67; RPD = 1.71). Predictions of measured aggregation measurements by NIRS, on the other hand, are less reliable ($r^2 <$ 0.6; RPD < 1.58). The poor performances of NIRS–PCR in aggregation may be due to artifacts related to sample handling, and/or the calibration model used may not be suitable for analyzing these soil properties. The use of first derivatives of near-infrared reflectance data used in this NIRS–PCR technique may remove some of particle size influence. However, use of the original $\log(1/R)$ data with the PCR technique did not improve results. Another possible explanation for the poor performance of the NIRS–PCR technique in predicting measures of aggregation may be that these measures did not accurately represent the shape and size of particles and voids in the samples.

Several soil properties that do not have a primary

Table 4. Statistics of near-infrared reflectance spectroscopy–principal components regression (NIRS–PCR) technique on the prediction of soil properties.

Soil property	$(n/N) \times 100$†	F	r^2‡	RMSECV‡	RPD‡	Category§
Total C	97.7	7	0.87	7.86	2.79	A
Total N	97.9	7	0.85	0.62	2.52	A
Moisture	97.9	8	0.84	0.005	2.47	A
Potentially min. N	98.7	8	0.72	26.05	1.84	B
Total respiration rate	97.5	9	0.66	205.37	1.70	B
Basal respiration rate	98.4	7	0.82	23.55	2.31	A
Biomass C	99.2	9	0.60	389.71	1.50	B
CEC¶	98.0	8	0.81	3.82	2.28	A
1.5 MPa water	97.7	7	0.81	0.025	2.30	A
pH (H$_2$O)	98.4	13	0.55	0.57	1.43	B
pH (CaCl$_2$)	98.3	13	0.56	0.56	1.47	B
Aggregation						
2.00 mm	98.7	6	0.48	4.28	1.38	C
1.00 mm	98.4	5	0.58	4.57	1.55	B
0.50 mm	99.7	9	0.53	5.19	1.44	B
0.25 mm	99.1	8	0.46	5.67	1.34	C
Macro	99.1	9	0.60	14.01	1.58	B
Particle size						
Clay	98.1	12	0.67	4.06	1.71	B
Silt	98.8	8	0.84	9.51	2.52	A
Sand	98.8	8	0.82	11.93	2.32	A
Mehlich III ex. metal						
Ca	98.6	12	0.80	564.62	2.19	A
Cu	98.0	14	0.25	5.37	0.92	C
Fe	97.3	9	0.64	61.40	1.66	B
K	97.8	13	0.64	132.43	1.64	B
Mg	98.2	14	0.71	143.79	1.78	B
Mn	97.7	12	0.70	56.40	1.79	B
P	97.6	13	0.40	32.28	1.18	C
Zn	97.9	9	0.44	15.28	1.07	C
Exchangeable base						
Ca	98.3	12	0.75	4.00	1.94	B
Mg	98.2	9	0.68	1.28	1.75	B
Na	98.2	7	0.09	0.13	0.92	C
K	98.2	13	0.55	0.42	1.44	B
Sum of exch. bases	98.0	11	0.75	4.86	1.98	B
Exchangeable acidity	98.6	11	0.65	2.44	1.68	B

† N is the number of tested samples and n is the number of tested samples with difference between measured and predicted value smaller than three times original RMSECV.
‡ The r^2, RMSECV, and RPD are based on the measured and predicted values of n samples.
§ Categories of the prediction ability of NIRS–PCR technique for soil properties. A: RDP > 2.0; B: RDP = 1.4–2.0; C: RDP < 1.4.
¶ Cation-exchange capacity.

response in the near-infrared region, such as CEC and 1.5 MPa water, were accurately predicted by the NIRS–PCR technique (r^2 > 0.80; RPD > 2.28). Correlation of these properties with soil organic matter and clay contents may explain some of this effect. Near-infrared reflectance–principal components regression analysis is also relatively accurate for predicting levels of Mehlich III extractable cations (Ca, Mg, K, Fe, and Mn). The ability of the NIRS–PCR technique to predict levels of extractable cations may be due to covariance of these properties with soil properties exhibiting a primary response in the near-infrared region. Generally, NIRS–PCR analyses were more accurate for predicting cations extracted by the Mehlich III method than the NH_4OAc method. The reasons for this difference are not clear.

The ability of the NIRS–PCR technique to predict soil biochemical properties varied (potentially mineralizable N, r^2 = 0.72, RPD = 1.84; total respiration rate, r^2 = 0.66, RPD = 1.70; basal respiration rate, r^2 = 0.82, RPD = 2.31; and biomass C, r^2 = 0.60, RPD = 1.50). Values of the biochemical properties and the accuracy of the NIRS–PCR predictions are affected by the levels and nature of soil organic matter. For example, soil samples from MLRA 77 have a lower mean total C content and a narrower range (mean = 4.2, range =

$1.9{\sim}11.1$ g kg^{-1}) than the soils from MLRA 105 (mean = 34.2, range = $5.5{\sim}285.8$ g kg^{-1}). The ability of the NIRS–PCR technique to predict biomass C and basal respiration rate for soils from MLRA 77 was very low (r^2 < 0.02), whereas these predictions were much more accurate for soil samples from MLRA 105 (r^2 = 0.66 and 0.80, respectively). In general, the regression coefficients for other soil properties for samples from MLRA 77 are lower than regression coefficients for samples from other MLRA groups. These differences suggest that organic matter content had a large effect on the accuracy of NIRS–PCR predictions for many soil properties.

In addition to abundance, the nature of soil organic matter also affects biochemical properties of soils. Soil organic matter is composed of large and complex organic polymers. Abundance and distributions of the various organic functional groups may influence both the biochemical properties and the NIR spectra. For example, potentially mineralizable N measures the N that is mineralized during a 35-d incubation. Soil organic matter that contains a large amount of N in readily decomposable proteins will have higher potentially mineralizable N levels than soils with large amounts of N in more recalcitrant heterocycles. Respiration rate may also de-

Table 5. Correlation coefficient, r, between soil properties, and r^2 between measured and predicted values based on Method A, $y = f$ (primary soil properties), and Method B, $y = f$ (near-infrared reflectance spectroscopy, NIRS).

Soil property	Correlation coefficient, r, between measured properties					r^2, based on measured and predicted values	
	Macro aggregation	Clay	Sand	Total C	Moisture	Method A (property)	Method B (NIRS)
Potentially min. N	0.44	0.03	−0.12	0.71	0.32	0.54	0.72
Total resp. rate	−0.15	0.10	−0.26	0.35	0.31	0.25	0.66
Basal resp. rate	0.45	0.10	−0.26	0.68	0.33	0.50	0.82
Biomass C	0.22	0.06	−0.09	0.54	0.29	0.27	0.60
1.5 MPa water	0.23	0.56	−0.47	0.86	0.72	0.88	0.81
CEC†	0.15	0.63	−0.54	0.80	0.79	0.89	0.81
pH (H_2O)	−0.02	0.08	0.19	−0.04	−0.01	0.09	0.55
pH ($CaCl_2$)	0.05	0.07	0.17	0.02	0.00	0.09	0.56
Mehlich III ex. Metal							
Ca	0.12	0.53	−0.35	0.59	0.61	0.53	0.80
Cu	0.12	−0.03	−0.02	0.03	−0.01	0.00	0.25
Fe	0.25	0.12	−0.45	0.40	0.28	0.32	0.64
K	−0.24	0.11	−0.04	0.20	0.24	0.13	0.64
Mg	0.27	0.62	−0.44	0.53	0.55	0.58	0.71
Mn	0.42	−0.02	−0.32	0.28	0.10	0.29	0.70
P	0.15	0.02	−0.10	0.31	0.13	0.07	0.40
Zn	0.17	−0.06	−0.00	0.14	0.01	0.03	0.44
NH_4OAc ex. Base							
Ca	0.16	0.51	−0.36	0.68	0.64	0.61	0.75
Mg	0.29	0.61	−0.44	0.57	0.55	0.60	0.68
Na	−0.05	0.13	−0.03	0.03	0.08	0.00	0.09
K	−0.20	0.12	−0.04	0.24	0.26	0.13	0.55
Sum of ex. bases	0.18	0.56	−0.40	0.70	0.66	0.68	0.75
Ex. acidity	0.09	0.25	−0.41	0.55	0.45	0.36	0.65

† Cation-exchange capacity.

pend on the nature of organic functional groups. Unfortunately, distinguishing soil organic functional groups using NIR spectra at this stage is difficult.

Near-Infrared Reflectance Spectroscopy–Principal Components Regression Predictions of Secondary Properties

Many soil properties with a secondary response in the near-infrared red could be predicted using NIRS–PCR (Table 4). The ability of NIRS–PCR to predict these properties may be due to correlations with soil properties having a primary response in the near-infrared light region. To test this hypothesis, we compared the predictions of secondary properties by multiple linear regression using the primary properties as independent variables (Method A) with predictions using the NIRS–PCR technique (Method B). Total N and silt contents were not included in the multiple linear regression of Method A because of the strong correlation between total C ($r = 0.95$) and sand contents ($r = −0.97$), respectively. The analysis (Table 5) indicates that in general, when high regression coefficients were obtained by multiple linear regression (Method A), high regression coefficients were also obtained using the NIRS–PCR technique. For instance, CEC was readily predicted by multiple linear regression using moisture, total C, particle-size distribution, and macro-aggregation ($r^2 = 0.89$), and therefore it was not surprising that CEC was also accurately predicted by NIRS–PCR ($r^2 = 0.81$). Mehlich III extractable Cu, on the other hand, was not correlated with any of the primary properties, and the regression coefficients for Methods A and B were both low (0.00 and 0.25, respectively). However, the ability of the NIRS–PCR technique to predict some secondary properties cannot

be explained by covariance with primary soil properties. Mehlich III extractable K is one example. The regression coefficient between the measured and predicted values based on Method A was low (0.13), but the regression coefficient for the NIRS–PCR technique was 0.64. The pH measured in H_2O and $CaCl_2$ were also more accurately predicted by NIRS–PCR than would be expected from covariance with primary soil properties. This discrepancy may be due to the inability of multiple linear regression to accurately represent the relationship between primary and secondary soil properties, and/or that other soil properties not measured may influence the NIR spectra.

CONCLUSIONS

Near-infrared reflectance spectroscopy is a simple and nondestructive analytical method that can be used to predict several soil properties simultaneously. The main challenge limiting application of NIRS for the evaluation of soil properties is finding suitable data pretreatment and calibration strategies. In this study, a NIRS–PCR technique was used to correlate reflectance data and values for soil properties. The results show that this technique can be used to simultaneously estimate several primary soil properties (such as total C, total N, moisture content, particle size) and secondary soil properties (such as CEC; extractable Ca, K, and Mn; respiration rate; and potentially mineralizable N) in a very short time. Different calibration strategies may be able to significantly improve NIRS predictions for other soil properties. However, the accuracy of NIRS–PCR is sufficient to support field and pedon-scale analysis of soil properties, and therefore NIRS has the potential to
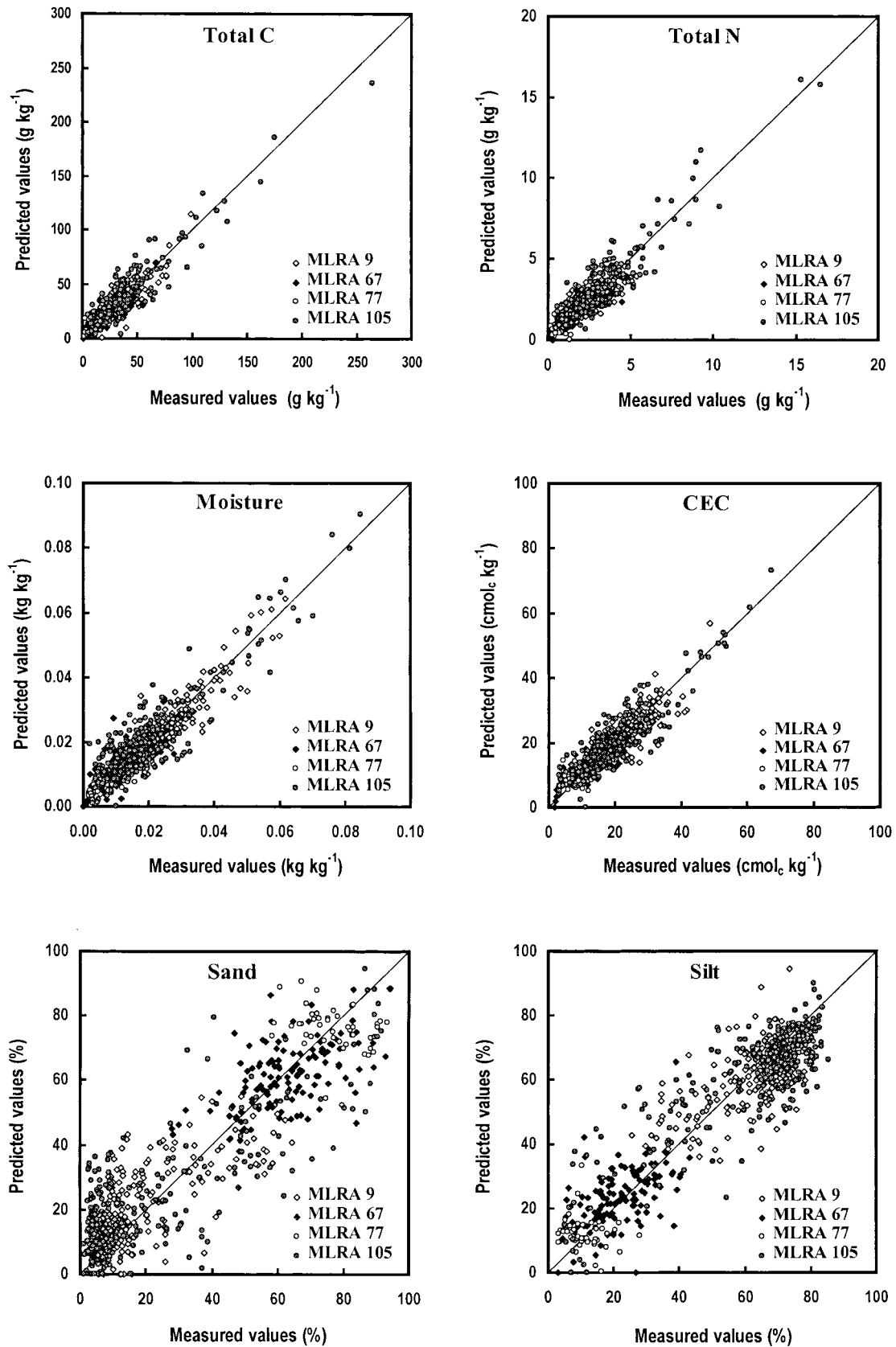
Fig. 5. Relationship between soil properties measured by standard laboratory procedures and predicted by NIRS–PCR technique. The 1:1 line is indicated on each figure. MLRA is Major Land Resource Area.
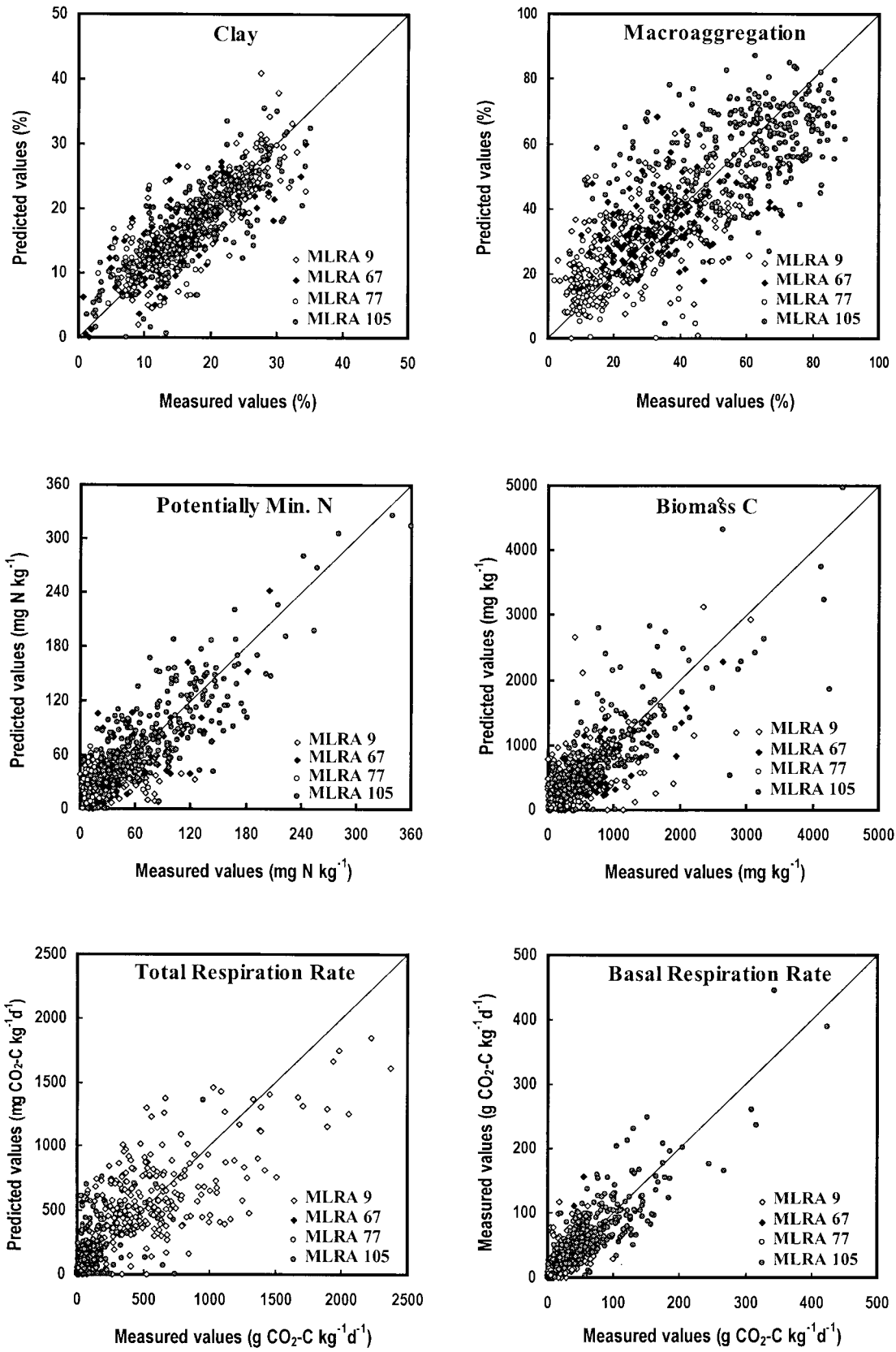
Fig. 5. Continued.

be used as a rapid soil testing technique for precision soil management and assessing soil quality.

REFERENCES

Al-Abbas, A.H., P.H. Swain, and M.F. Baumgardner. 1972. Relating organic matter and clay content to the multispectral radiance of soils. Soil Sci. 114:477–485.

Ben-Dor, E., and A. Banin. 1995. Near infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J. 59:364–372.

Ben-Gera, I., and K.H. Norris. 1968. Determination of moisture content in soybeans by direct spectrophotometry. Isr. J. Agric. Res. 18:124–132.

Bowers, S.A., and R.J. Hanks. 1965. Reflection of radiant energy from soils. Soil Sci. 100:130–138.

Brejda, J.J., T.B. Moorman, J.L. Smith, D.L. Karlen, D.L. Allan, and T.H. Dao. 2000. Distribution and variability of surface soil properties at a regional scale. Soil Sci. Soc. Am. J. 64:974–981.

Brejda, J.J., T.B. Moorman, J.L. Smith, D.L. Karlen, and T.H. Dao. 2000. Identification of regional soil quality factors and indicators: I. Central and southern high plains. Soil Sci. Soc. Am. J. 64: 2115–2124.

Creaser, C.S., and A.M.C. Davies. 1988a. Section 1: Near infrared spectroscopy. p. 1–172. In Analytical applications of spectroscopy. Royal Soc. Chem., London.

Creaser, C.S., and A.M.C. Davies. 1988b. Section 6: Chemometrics and data analysis. p. 403–480. In Analytical applications of spectroscopy. Royal Soc. Chem., London.

Dalal, R.C., and R.J. Henry. 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance. Soil Sci. Soc. Am. J. 50:120–123.

Drinkwater, L.E., C.A. Cambardella, J.D. Reed, and C.W. Rice. 1996. Potentially mineralizable nitrogen as a property of biologically active soil nitrogen. p. 217–229. In J.W. Doran and A.J. Jones (ed.) Methods for assessing soil quality. SSSA Spec. Publ. 49. SSSA, Madison, WI.

Fox, R.H., J.S. Shenk, W.P. Piekielek, M.O. Westerhaus, J.D. Toth, and K.E. Macneal. 1993. Comparison of near-infrared spectroscopy and other soil nitrogen availability quick tests for corn. Agron. J. 85:1049–1053.

Fritze, H., P. Järvinen, and R. Hiukka. 1994. Near-infrared characteristics of forest humus are correlated with soil respiration and microbial biomass in burnt soil. Biol. Fertil. Soils. 18:80–82.

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. 56:865–872.

Holst, H. 1992. Comparison of different calibration methods suited for calibration problems with many variables. Appl. Spec. 46: 1780–1784.

Kemper, W.R., and R.C. Rosenau. 1986. Aggregate stability and sized distribution. p. 425–442. In A. Klute (ed.) Methods of soil analysis. Part 1. 2nd ed. Agron. Monogr. 9. ASA, SSSA, CSSA, Madison, WI.

Krishnan, R., J.D. Alexander, B.J. Butler, and J.W. Hummel. 1980. Reflectance technique for predicting soil organic matter. Soil Sci. Soc. Am. J. 44:1282–1285.

Martens, H., and T. Naes. 1989. Multivariate calibration. John Wiley and Sons, New York.

Mehlich, A. 1984. Mehlich III soil test extractant: A modification of Mehlich II extractant. Commun. Soil Sci. Plant Anal. 15:1409–1416.

Morra, M.J., M.H. Hall, and L.L. Freeborn. 1991. Carbon and nitrogen analysis of soil fractions using near-infrared reflectance spectroscopy. Soil Sci. Soc. Am. J. 55:288–291.

Murray, I., and I.A. Cowe. 1992. Making light work: Advances in near infrared spectroscopy. Weinheim, New York.

Naes, T., and T. Isaksson. 1990. Locally weighted regression and scatter correction for near-infrared reflectance data. Anal. Chem. 62:664–673.

Sudduth, K.A., and J.W. Hummel. 1993a. Potable, near-infrared spectrophotometer for rapid soil analysis. Trans. ASAE 36:185–193.

Sudduth, K.A., and J.W. Hummel. 1993b. Soil organic matter, CEC, and moisture sensing with a portable NIR spectrophotometer. Trans. ASAE 36:1571–1582.

Sudduth, K.A., and J.W. Hummel. 1996. Geographic operating range evaluation of a NIR soil sensor. Trans. ASAE 39:1599–1604.

Tate, K.R., D.J. Ross, and C.W. Feltham. 1988. A direct extraction method to estimate soil microbial C: Effects of some experimental variables and some different calibration procedures. Soil Biol. Biochem. 20:329–335.

USDA-NRCS. 1996. Soil survey laboratory methods manual. Soil Survey Investigations Rep. 42, Ver. 3.0. Natl. Soil Survey Center, Lincoln, NE.

Wetzel, D.L. 1983. Near-infrared reflectance analysis: Sleeper among spectroscopic techniques. Anal. Chem. 55:1165A–1176A.

Workman, J.J., Jr. 1996. Interpretive spectroscopy for near infrared. Appl. Spec. Rev. 31:251–320.