CALIBRATION METHODS FOR SOIL PROPERTY ESTIMATION USING REFLECTANCE SPECTROSCOPY

K. S. Lee, K. A. Sudduth, S. T. Drummond, D. H. Lee, N. R. Kitchen, S. O. Chung

ABSTRACT. *Optical diffuse reflectance sensing is a potential approach for rapid and reliable on-site estimation of soil properties. One issue with this sensing approach is whether additional calibration is necessary when the sensor is applied under conditions (e.g., soil types or soil moisture conditions) different from those used to generate an initial calibration, and if so, how many sample points are required in this additional calibration. In this study, these issues were addressed using data from ten fields in five states in the U.S. Corn Belt. Partial least squares (PLS) regression was used to develop calibrations between soil properties and reflectance spectra. Model evaluation was based on the ratio of standard deviation to RMS error (RPD), a statistic commonly used in spectral analysis. When sample data from the field where calibrations were to be applied (i.e., test field) were included in the calibration stage (full information calibration), RPD values of prediction models were increased by an average of 0.55 (from 1.08 to 1.63) compared with results from models not including data from the test field (calibration without field-specific information). Including some samples from the test field (hybrid calibration) generally increased RPD to 90% of that from full information calibration (average increase = 0.49) by using data from 8 to 20 soil cores, with little further improvement given additional data. Using test field points as a bias adjustment (two-stage calibration) increased RPD by an average of 0.29 with two to six sample points, a finding that was confirmed by Monte Carlo simulation. These results show the importance of including in a calibration set samples similar (i.e., obtained from the same or similar fields) to those in the test set. These similar samples could be included directly in the calibration or could be used to implement a post-calibration bias adjustment. Although results were more accurate with the recalibration approach, the bias adjustment approach was more efficient computationally and required less data. Thus, either might be preferred depending on specific circumstances.*

Keywords. Calibration methods, Near-infrared, Reflectance spectroscopy, Sensors, Soil properties.

P recision agriculture is a management system where application of agricultural chemicals such as fertilizers, pesticides, and herbicides is matched to actual needs point-by-point within fields. This approach can provide economic benefits to farmers and protection of the soil environment from excessive chemical application. For precision agriculture to meet its goals, site-specific quantification of soil physical and chemical properties that affect

The authors are **Kyou Seung Lee, ASABE Member Engineer,** Professor, Department of Bio-Mechatronics Engineering, Sungkyunkwan University, Suwon, Korea; **Kenneth A. Sudduth, ASABE Fellow,** Agricultural Engineer, and **Scott T. Drummond,** Information Technology Specialist, USDA-ARS Cropping Systems and Water Quality Research Unit, Columbia, Missouri; **Dong Hoon Lee,** Graduate Student, Department of Bio-Mechatronics Engineering, Sungkyunkwan University, Suwon, Korea; **Newell R. Kitchen,** Soil Scientist, USDA-ARS Cropping Systems and Water Quality Research Unit, Columbia, Missouri; and **Sun-Ok Chung, ASABE Member Engineer,** Assistant Professor, Department of Biosystems Machinery Engineering, Chungnam National University, Daejeon, Korea. **Corresponding author:** Kenneth A. Sudduth, USDA-ARS Cropping Systems and Water Quality Research Unit, 269 Agricultural Engineering Bldg., University of Missouri, Columbia, MO 65211; phone: 573-882-1114, ext. 303; fax: 573-882-1115; e-mail: Ken.Sudduth@ars.usda.gov.

soil quality and crop production is necessary. Many of these properties can change on a fine spatial scale, making traditional laboratory methods impractical due to time and cost of the sampling and analytical procedures. Another approach is needed to provide accurate, reliable, and timely characterization of within-field variability at a reasonable cost. Diffuse reflectance spectroscopy (DRS) is a promising, nondestructive technique that may fulfill these requirements. Many investigators have successfully estimated soil physical and chemical properties in the laboratory using DRS in the visible (VIS; 400-700 nm), near-infrared (NIR; 700-2500 nm), and mid-infrared (MIR; 2500-25000 nm) wavelength ranges. In addition, several DRS soil sensors have been successfully used in field settings.

To date, most DRS soil sensing research has been carried out in the VIS, NIR, or combined VIS-NIR wavelength ranges (Viscarra Rossel et al., 2006). Total C in arable soils has been estimated with NIR or VIS-NIR spectroscopy (Chang et al., 2001; Confalonieri et al., 2001; McCarty et al., 2002; Mouazen et al., 2007), producing R^2 values from 0.73 to 0.95. This technique has also been used to determine soil organic C (SOC) (Krishnan et al., 1980; Dalal and Henry, 1986; Sudduth and Hummel, 1991; Reeves and McCarty, 2001; Shepherd and Walsh, 2002; Islam et al., 2003; Mouazen et al., 2007) and to estimate soil properties such as cation exchange capacity (CEC), Ca, K, texture (sand, silt, and clay fractions), Mg, pH, and total N (Sudduth and Hummel, 1993; Ben-Dor and Banin, 1995; Shepherd and Walsh, 2002; Cozzolino and Moron, 2003; Islam et al., 2003; Nanni and De-

matte, 2006). In a previous study (Lee et al., 2009), we combined VIS-NIR reflectance sensing with partial least squares (PLS) regression to estimate surface and profile soil properties and to identify wavelength bands important for estimating soil properties. Soil samples were obtained from ten fields in five states of the U.S. Corn Belt. Good estimates of SOC, CEC, Ca, and texture fractions were obtained for both surface and profile datasets. Viscarra Rossel et al. (2006) provided a comprehensive review of soil DRS applications, including accuracy statistics.

Proper design and selection of calibration sets and methods is important for obtaining the most precise and accurate calibration results (Westerhaus et al., 2004). Three types of calibration sets can be used: closed population (entire sample set to be analyzed is available at the time of calibration development), open population (some samples are not available at the time of calibration), and historical databases (calibration developed from previous experiments with variation in instrumentation and methodology). Westerhaus et al. (2004) stated that for an open population, the only method applicable for real-time prediction, one way to ensure accurate estimation of a new sample was to have one or more very similar samples in the calibration database. Sudduth and Hummel (1996) reported that the geographic range over which a set of soil samples was obtained affected the NIR estimation accuracy of SOC, CEC, and soil moisture. Specifically, prediction of SOC and soil moisture became less accurate as soil samples from a wider geographic range were considered. Using one-field-out validation, an approach mimicking the prediction problems encountered in real-time measurement, Christy (2008) obtained best results for organic matter, with an RMSE of 0.52% and an R^2 of 0.67. Prediction accuracy and the percentage of locations with accurate predictions increased as the number of fields in the calibration set increased, but the best one-field-out results had 30% higher error than calibrations that included data from all fields. Brown et al. (2005) applied VIS-NIR PLS regression modeling to soil samples obtained from six sites with similar soils and concluded that a stable, effective PLS calibration could be developed for similar soils from the same physiographic region. However, when they attempted to predict soil C for each of the six sites using a calibration developed from the remaining five sites, the models failed completely at two of the six sites and gave inconsistent results at a third site. This was true despite pre-screening for spectral similarity.

Previous research has shown that the DRS calibration approach can have a significant effect on the accuracy of soil property estimates. Specifically, choice of a calibration dataset requires consideration of the trade-off between the accuracy required and the resources available to develop the calibration. If the highest accuracy is needed and resources are not a limitation, then individual field calibration with a significant number of laboratory-measured calibration samples will likely provide the best results. On the other hand, resource limitations or other practicality issues may mean that it is not possible to obtain any within-field calibration samples, requiring that a global or "factory" calibration be used. In between, there is likely an interplay between how many within-field calibration samples are available and the best methodology for incorporating those samples.

Although NIR or VIS-NIR soil property estimation accuracies have been widely reported in the literature, very few studies have compared multiple calibration approaches.

Thus, our goal was to evaluate several calibration methods along the continuum between within-field and global calibration. Specific objectives were to:
- Evaluate the reduction in accuracy of soil property estimates when test-field samples are not included in developing a calibration.
- Compare different methods for including samples from a test field when developing a calibration.
- Determine the effect of varying numbers of test-field calibration samples.

MATERIALS AND METHODS
CALIBRATION METHODS
Four different calibration methodologies based on PLS regression were compared using a DRS dataset that was subdivided into six parts as described below. These methods were:

Method 1: Full Information Calibration. This method used data from all six datasets to develop the initial calibration equation. The calibration equation was then used to calculate separate prediction statistics for each of the six datasets. This method was included to provide an upper bound on prediction accuracy for the different soil properties, against which the other methods could be compared.

Method 2: Calibration without Field-Specific Information. This method simulated the application of a "factory-calibrated" sensor to soils different from those used to develop the initial calibration. A calibration was developed using five of the six datasets and prediction statistics were calculated for the remaining dataset, which we will refer to as the "test" dataset. This process was repeated six times to obtain prediction statistics for all datasets.

Method 3: Hybrid Calibration. This method was intermediate between methods 1 and 2, simulating the addition of a small number of field-specific calibration points to an initial general calibration. PLS calibrations were developed as in method 2, but also included a number of additional calibration points (soil cores) selected at random from the test dataset. The number of additional points was iteratively increased from zero (equivalent to method 2) to the maximum number of points in the dataset (equivalent to method 1). Prediction statistics were calculated as for method 2.

Method 4: Two-Stage Calibration with Bias Correction. Observing that the difference between method 1 and method 2 was principally a bias shift (fig. 1) led us to develop method 4. This method was similar to method 3, but instead of directly including the points randomly selected from the test dataset in the PLS calibration, they were used to adjust the results of the calibration. The advantage of this method over method 3 was that it did not require a re-calculation of the PLS regression for each new test dataset. The first stage of this method was the PLS calibration of method 2. The second stage was a linear bias correction where the general calibration was shifted using laboratory data from a number of calibration points in the test dataset. Similar to method 3, the number of calibration points used in the bias correction was iteratively increased to create a suite of method 4 estimates, followed by calculation of prediction statistics.

SOIL AND SPECTRAL DATA
Soils used in this study, also used by Lee et al. (2009), were obtained from ten fields managed in a corn-soybean rotation,
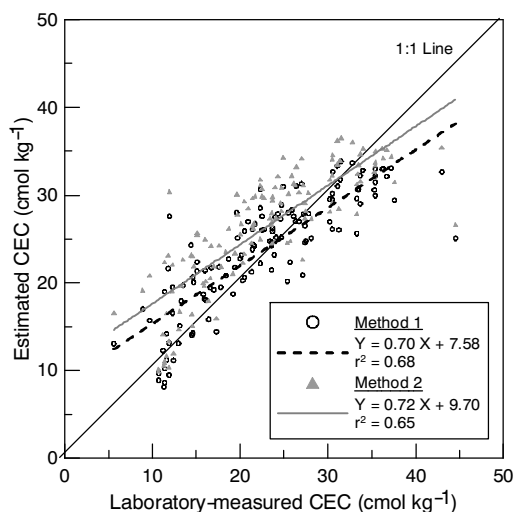
Figure 1. Estimated vs. laboratory-measured CEC for an example dataset, showing that the difference between method 1 and method 2 calibrations could be represented reasonably well by a linear bias shift.

two each in Missouri (MO), Illinois (IL), Michigan (MI), South Dakota (SD), and Iowa (IA), as described by Sudduth et al. (2005). Fields were selected in part due to their relatively high degree of within-field soil and production variability, and to represent the range of climate, soil, and landscape characteristics typical of the north-central U.S. Soils exhibited differences in texture, parent material, and mineralogy. For example, prevailing surface soil texture varied across the research sites as follows: loam (MI), loam to clay loam (IA), silt loam to silty clay loam (IL, MO, and SD). Subsoil texture was even more variable, ranging from loamy sand at the MI fields to clay at the MO fields.

Soil samples were collected from 12 to 20 locations in each field. Locations were chosen with the goal of including samples from all the landscape positions and soil map units present in each field. One 4.0 cm diameter, 120 cm long core sample was obtained at each site using a hydraulic soil coring machine. Cores were examined and segmented by pedogenic horizon. The number of horizon samples per core varied from 1 to 6, with a median of 3. Soil texture (clay, silt, and sand fractions, %), cation exchange capacity (CEC), Ca, and Mg (cmol kg^{-1}), organic C (%), total N (%), and pH were determined by laboratory analysis, using procedures described by Sudduth et al. (2005). Descriptive statistics of the laboratory-measured soil properties are given in table 1.

Spectra of air-dried and 2 mm sieved soil were obtained in the laboratory using a spectrometer (FieldSpec Pro FR,

Analytical Spectral Devices, Boulder, Colo.). Approximately 15 cm^3 of soil was packed in a glass-bottomed sample cup for reflectance determination. The sample was illuminated through the glass by a halogen lamp, and the reflected light from an approximately 12 mm diameter area was transmitted to the spectrometer through a fiber optic bundle. Each soil spectrum was obtained as the mean of ten scans. The spectrometer data collection software automatically adjusted the data for dark current variations using dark current scans obtained at the beginning of each data collection session and at least every 30 min thereafter. A Spectralon (Labsphere, Inc., North Sutton, N.H.) reflectance standard was scanned after every ten soils and used to convert the raw spectral data to decimal reflectance.

Although the spectrometer was capable of collecting data from 350 to 2500 nm, only NIR data from 1770 to 2500 nm were used in this study. We previously determined that soil property estimates using this reduced wavelength range had similar accuracy to those obtained with the full range of the spectrometer (Lee et al., 2009). Other researchers have also reported that wavelengths in the NIR range were most predictive of soil properties (Sudduth and Hummel, 1991; Henderson et al., 1992; Sudduth and Hummel, 1993; Chang et al., 2001).

ANALYTICAL PROCEDURES

Reflectance data were preprocessed to improve stability of the regression. Each spectral scan was (1) transformed from reflectance to absorbance (log$_{10}$[1/reflectance]), (2) mean-normalized (i.e., divided by its mean value), and (3) smoothed with an 8-point (24 nm) moving average to simulate the wider bandwidth likely in a real-time sensor.

PLS regression, implemented in Unscrambler version 9.1 (CAMO, Inc., Oslo, Norway), was used to develop calibration models relating soil properties and the preprocessed reflectance spectra. PLS creates a new set of variables (called factors) that are uncorrelated and that explain variation in both response and predictor variables (Beebe and Kowalski, 1987). A key step in PLS regression is selecting the optimal number of factors to best represent the calibration data without overfitting. In this analysis, a 10-segment cross-validation approach was used to choose the optimum number of PLS factors. Additional details were previously reported (Lee et al., 2009).

Model evaluation was based on coefficient of determination (R^2) and the ratio of standard deviation to root mean square error of prediction (RMSEP), commonly termed RPD. Being a normalized statistic, RPD is suggested instead of RMSEP when comparing results from datasets containing

Table 1. Means and standard deviations (SD) of laboratory-determined soil properties.

Soil Property	Missouri		Illinois		Michigan		South Dakota		Iowa	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Clay fraction (%)	34.7	12.4	29.7	6.3	14.3	6.2	26.4	5.6	25.0	7.1
Silt fraction (%)	60.8	10.9	58.9	9.2	32.9	18.9	51.6	12.3	36.3	8.5
Sand fraction (%)	4.5	3.8	11.3	11.6	52.8	23.1	22.0	16.0	38.6	15.0
Ca (cmol kg^{-1})	11.4	3.8	14.5	6.4	5.0	2.5	32.4	13.6	28.6	16.5
Mg (cmol kg^{-1})	4.0	2.0	5.3	1.9	1.0	0.5	8.5	2.9	4.0	1.6
CEC (cmol kg^{-1})	24.0	7.9	20.7	7.0	9.8	4.2	22.0	7.5	22.6	8.4
SOC (%)	0.65	0.40	0.86	0.71	0.63	0.55	1.09	0.89	1.10	0.89
Total N (%)	0.08	0.03	0.10	0.06	0.06	0.05	0.11	0.08	0.12	0.08
pH	5.0	0.7	6.6	0.8	5.3	0.6	7.6	0.7	6.4	1.0

Table 2. RPD values obtained with method 1. Bold entries indicate good estimation (RPD > 2.0).

Prediction Dataset	Clay	Silt	Sand	Ca	Mg	CEC	SOC	pH	Total N	Mean
MO	1.88	1.41	0.42	0.92	1.58	1.68	1.76	1.46	0.95	1.34
IL	1.38	1.03	1.07	1.35	1.29	1.86	**3.50**	1.68	**2.55**	1.75
MI	1.16	1.94	1.96	0.52	0.40	1.10	**2.80**	1.40	**2.24**	1.50
SD1	1.47	1.16	1.24	1.54	1.23	**2.17**	**3.83**	1.38	**2.69**	1.86
SD2	1.79	0.96	1.10	1.41	1.20	**2.38**	**2.99**	1.16	**2.38**	1.71
IA	1.69	1.12	1.44	**2.00**	0.88	1.75	**2.54**	**2.04**	1.38	1.65
All combined	**2.24**	1.88	**2.11**	**2.22**	1.86	**2.27**	**2.73**	**2.51**	1.71	**2.17**

Table 3. R^2 values obtained with method 1. Bold entries indicate good estimation ($R^2 > 0.8$).

Prediction Dataset	Clay	Silt	Sand	Ca	Mg	CEC	SOC	pH	Total N	Mean
MO	0.74	0.60	0.03	0.32	0.60	0.66	0.78	0.54	0.33	0.51
IL	0.48	0.19	0.30	0.54	0.42	0.71	**0.92**	0.72	**0.85**	0.57
MI	0.57	0.75	0.74	0.48	0.05	0.54	**0.88**	0.49	**0.81**	0.59
SD1	0.57	0.47	0.54	0.59	0.41	**0.81**	**0.96**	0.49	**0.91**	0.64
SD2	0.70	0.32	0.44	0.52	0.55	**0.89**	**0.97**	0.47	**0.95**	0.65
IA	0.67	0.49	0.60	0.75	0.24	0.68	**0.86**	0.76	0.50	0.62
All combined	0.80	0.72	0.78	0.80	0.71	**0.81**	**0.87**	**0.84**	0.66	0.78

different degrees of variability, as is the case with multiple soil properties and different test fields (Malley et al., 2004). Values of RPD above 2.0 and R^2 above 0.80 have been considered indicative of good accuracy in soil analysis (Chang et al., 2001; Lee et al., 2009).

For analysis, data from the two fields in each of MO, IL, IA, and MI, which were located less than 2 km from each other and had similar soils, were pooled. The two SD fields, which were geographically separated and had considerably different soils, were kept separate, resulting in six prediction datasets: MO, IL, IA, MI, SD1, and SD2.

RESULTS AND DISCUSSION
METHOD 1 AND METHOD 2

Tables 2 and 3 show method 1 RPD and R^2 values obtained for each soil property and prediction dataset. Also shown are statistics for the combined dataset, as previously presented by Lee et al. (2009). The generally lower RPD and R^2 values of the individual prediction datasets illustrate the difficulty of obtaining good field-specific estimates, particularly when the within-field variation in the measured property is low (table 1). Across the individual prediction datasets, there were consistent trends when comparing these two accuracy measures, with only one dataset being classified as "good" (RPD > 2.0; $R^2 > 0.8$) by one statistic and not the other. Because of this and because studies in the chemometrics literature (e.g., Blanco et al., 2000; Estienne et al., 2001) consider it appropriate to evaluate different DRS calibration methodologies based only on RPD or RMSEP, RPD was chosen as the evaluation statistic for comparison among calibration methods in the remainder of this study.

Figure 2 compares RPD values for soil property estimates by methods 1 and 2. All RPD values are located below the 1:1 line, indicating that method 1, where specific field information was included, was consistently better than method 2, regardless of field location or soil property. The distance of each point from the 1:1 line indicates how much RPD de-creased from method 1 to method 2. For example, data points from the two SD datasets were close to the line, meaning there was little decrease in accuracy when using method 2.
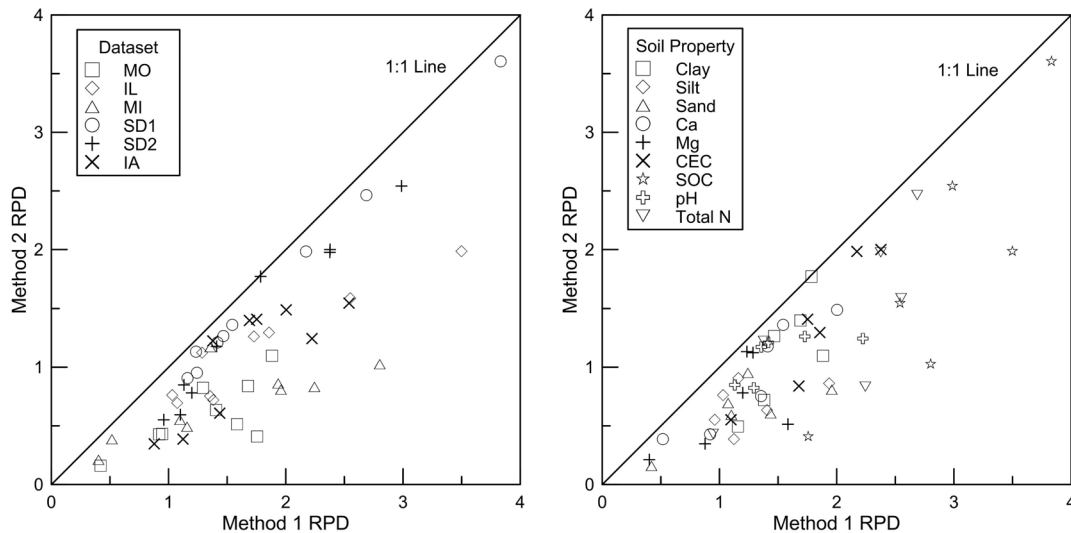

Figure 2. Comparison of RPD for methods 1 and 2 across all datasets (left) and soil properties (right).

Table 4. Ratio of RPD from method 2 to RPD from method 1.

Test Dataset	Clay	Silt	Sand	Ca	Mg	CEC	SOC	pH	Total N	Mean
MO	0.58	0.45	0.38	0.46	0.32	0.50	0.23	0.64	0.45	0.45
IL	0.52	0.74	0.65	0.56	0.87	0.70	0.57	0.73	0.62	0.66
MI	0.43	0.45	0.41	0.75	0.53	0.50	0.37	0.86	0.37	0.52
SD1	0.86	0.78	0.77	0.88	0.92	0.91	0.94	0.85	0.92	0.87
SD2	0.99	0.57	0.54	0.83	0.65	0.84	0.85	0.75	0.83	0.76
IA	0.83	0.34	0.42	0.74	0.40	0.80	0.61	0.56	0.89	0.62
Mean	0.70	0.56	0.53	0.70	0.61	0.71	0.59	0.73	0.68	0.65

To further investigate the increase in RPD, ratios of method 2 RPD to method 1 RPD were calculated (table 4). RPD increases were different for different soil properties, with method 2 providing RPD values from 23% to 99% of the method 1 RPD. As noted above, method 2 performed better for the SD datasets, with mean RPD values 76% and 87% of the method 1 RPD. For the MO, IL, MI, and IA datasets, method 2 provided mean RPD values from 45% to 66% of the method 1 RPD.

RPD differences for SOC were generally large for the MO, IL, MI, and IA datasets, with ratios ranging from 23% to 61%, and small for the SD fields, with ratios of 85% and 94% (table 4). The largest RPD differences (smallest ratios) for clay, CEC, and total N were found for the MO, IL, and MI datasets. The largest RPD differences for silt and sand were in MO, MI, and IA, and for Ca in MO and IL. For other soil properties and states, RPD was more similar between the two methods; however, the prediction accuracy of method 1 was always better than that of method 2.

For the two SD fields, where method 1 was more accurate on average than at the other fields (tables 2 and 3), the method 2 RPD was closer to method 1 compared to the datasets from other states. This could have been due to several factors: (1) variability of most soil properties within the SD fields was within the range encompassed by the other fields (Lee et al., 2009), (2) because the two SD datasets were kept separate for analysis, they contained fewer calibration points than the other datasets, or (3) contrary to our initial assumption that soils were different between the two SD fields, soils were similar enough that having one of the fields in the calibration dataset meant that the second did not provide much improvement.

Based on these results, and consistent with Westerhaus et al. (2004), we conclude that NIR soil property estimates are degraded considerably when calibration samples are not representative of the variability present in the locations (e.g., fields, watersheds, regions) where the sensor is to be operated. The approach used in our analysis, where all but the SD datasets included data from multiple fields, suggested that samples may not need to be from the exact fields under study but should at least come from fields with similar soils and management histories. Previous researchers (Sudduth and Hummel, 1996; Christy, 2008; Brown et al., 2005) reported similar findings.

METHOD 3

With method 3, RPD increased as the number of test field sample points added to the calibration model increased, but the degree of increase was different for different sites and soil properties. Figure 3 shows RPD vs. the number of sample points (soil cores) that were added to method 2 for CEC (left) and SOC (right). Each sample point included all data from the complete soil core obtained at that point, an average of three samples per core. Soil organic carbon was chosen for illustration because it had the highest RPD in method 1 (table 2) and for several datasets exhibited a large difference in RPD between method 1 and method 2 (table 4). CEC was chosen because its RPD values were near the mean for all soil properties in both cases.

RPD values increased rapidly with added points but only up to a certain number (e.g., about 6 for CEC and 9 for SOC in MO fields). The rates of RPD increase then became lower, to near zero in some cases. For example, as the number of sample points increased from 0 (same as method 2) to 16, SOC RPD increased rapidly, from about 1.0 to 2.3 in the MI dataset and from 2.0 to 3.1 in the IL dataset. In the MO and IA datasets, SOC RPD increased from 0.4 to 1.4 and from 1.6 to 2.2, respectively (fig. 3). For CEC, RPD values increased by about 0.7 with 12 additional sample points for the MO dataset, and by about 0.5 with 10 added points for the IL dataset and 19 added points for the MI dataset. Generally, RPD increases for the two SD datasets were lower; however, the initial (method 2) RPD was higher for the SD datasets than for others (fig. 3).
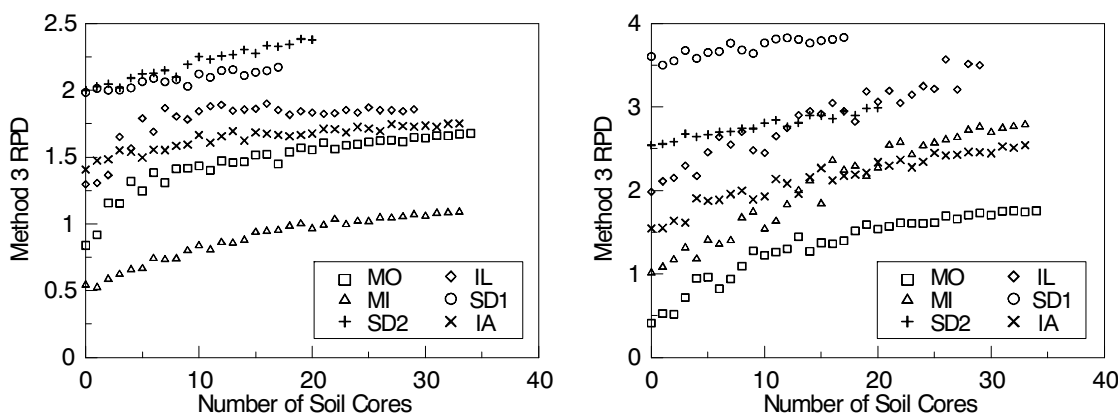

Figure 3. Method 3 RPD vs. number of sample points included from the test dataset for CEC (left) and SOC (right).

Table 5. Number of sample points from the test dataset required to achieve a method 3 RPD 90% of that from method 1.

Test Dataset	Clay	Silt	Sand	Ca	Mg	CEC	SOC	pH	Total N
MO	16	24	24	18	22	14	20	14	12
IL	16	8	14	20	2	4	24	14	16
MI	20	10	16	22	20	20	20	14	20
SD1	8	4	6	1	1	1	1	2	1
SD2	1	12	12	6	8	8	4	6	8
IA	12	20	14	18	22	6	18	10	8

Table 5 shows the number of soil sample points needed with method 3 to obtain 90% of the RPD of method 1. The number of samples required to reach this threshold using method 3 was different depending on sites and soil properties and ranged from 1 to 24. For SOC, 15 additional sample points were sufficient to reach this threshold for the IA dataset, 20 were required for the MO and IL datasets, and 24 were necessary for the MI dataset. In the case of the SD1 and SD2 datasets, where relatively high RPD values were obtained using method 2 and the RPD differences were relatively small between methods 1 and 2, adding only a few sample points (1 for SD1 and 4 for SD2) was sufficient to reach the 90% threshold.

In the case of CEC, adding a small number of sample points often allowed the 90% threshold to be met (1 for SD1, 4 for IL, 6 for IA, and 8 for SD2); however, 14 were required for MO and 20 for MI, the two datasets with the lowest method 2 RPD. Similar patterns were observed for the other soil properties. Those with relatively high method 2 RPD values (e.g., soil properties from SD fields) required only a small number of additional sample points to be added in method 3 to reach the 90% threshold, although the RPD increases were generally small. However, soil properties where method 2 RPD values were relatively low showed gradual RPD increases, with the threshold being reached with somewhere between 8 to 20 added sample points. From these results, we conclude that the addition of several within-field calibration points to the PLS calibration can often improve the accuracy of NIR soil property estimates, and that a large portion of the potential improvement (i.e., reaching 90 % of the method 1 RPD) can be obtained through the addition of 8 to 20 field-specific points, depending on the soil properties to be estimated. Using an approach similar to method 3, Bricklemyer and Brown (2010) found a 20% increase in RPD for clay content when adding nine local samples to a general calibration, but no increase in SOC accuracy. Their lower improvement in RPD may have been because they used a maximum of nine local samples, fewer than the number indicated for most test datasets in this study (table 5).

METHOD 4

Figure 4 shows the maximum RPD values obtained using method 4 with bias correction (i.e., adjusted method 2) vs. method 2 without field-specific information, for all fields and soil properties. RPD values were all located above the 1:1 line, indicating that the bias correction of method 4 consistently improved estimates of soil properties compared with method 2. Datasets with higher method 2 RPD values had little improvement with method 4, as shown by those values being near the 1:1 line. However, method 4 provided a substantial improvement for many datasets with a moderate method 2 RPD (fig. 4).

In method 4, the rate of RPD increase with the number of sample points used for bias correction was different for different field datasets and soil properties. For many datasets and soil properties, RPD increased by about 0.2 to 1.0 when 2 to 6 sample points were used, but then did not increase further. Greater RPD increases of about 0.4 to 1.0 were observed for CEC in the MO and IL datasets and for SOC in the MO, MI, and IA datasets. In the SD1 and SD2 datasets, where method 2 RPD was already relatively high, bias correction did not improve SOC estimates.

Table 6 shows the number of soil sample points required in method 4 to obtain 90% of the method 1 RPD. Small numbers of additional sample points (i.e., 1 to 9) were needed to achieve the 90% threshold for many soil properties in the SD
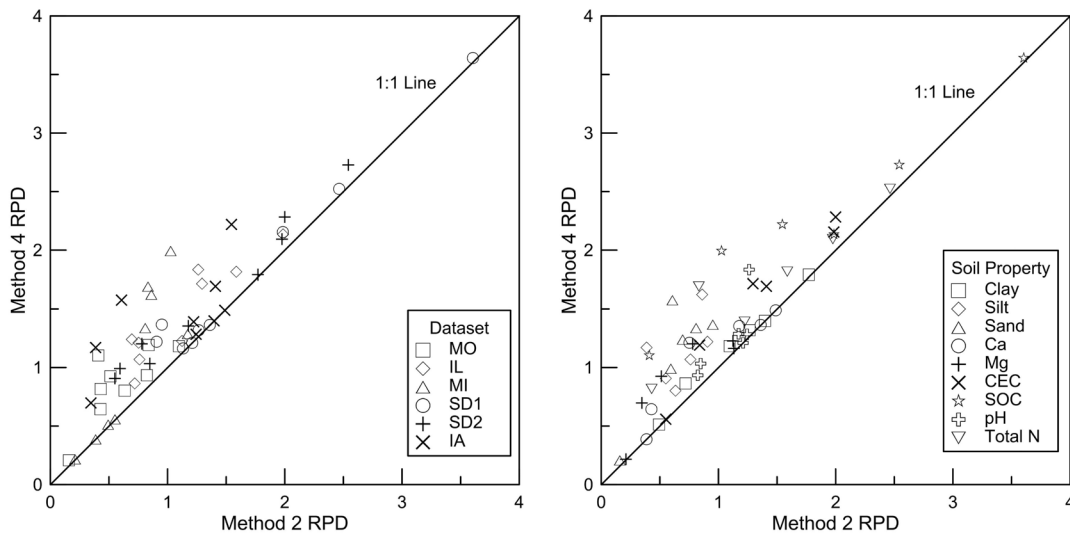

Figure 4. Comparison of RPD statistics for methods 4 and 2 across all datasets (left) and soil properties (right).

Table 6. Number of sample points from the test dataset required in method 4 to achieve an RPD 90% of that from method 1.

Test Dataset	Clay	Silt	Sand	Ca	Mg	CEC	SOC	pH	Total N
MO	--[a]	--	--	--	--	--	--	--	--
IL	--	2	2	--	1	4	--	1	--
MI	--	--	--	--	--	--	--	1	--
SD1	--	1	1	--	5	4	1	--	1
SD2	1	9	-	1	2	1	1	2	--
IA	--	1	3	--	--	6	--	--	11

[a] Indicates that method 4 was not able to achieve an RPD 90% of that from method 1.

Table 7. Ratio of maximum RPD from method 4 to RPD from method 1.

Test Dataset	Clay	Silt	Sand	Ca	Mg	CEC	SOC	pH	Total N	Mean
MO	0.63	0.57	0.49	0.70	0.58	0.71	0.62	0.72	0.86	0.65
IL	0.89	1.04	1.15	0.89	0.95	0.92	0.61	1.06	0.71	0.91
MI	0.75	0.84	0.68	0.75	0.54	0.51	0.71	0.95	0.75	0.72
SD1	0.88	1.05	1.10	0.88	0.94	0.99	0.95	0.85	0.94	0.95
SD2	0.96	0.94	0.90	0.96	1.00	0.96	0.91	0.91	0.88	0.94
IA	0.74	1.04	1.10	0.74	0.79	0.97	0.87	0.57	0.94	0.86
Mean	0.81	0.91	0.90	0.82	0.80	0.84	0.78	0.84	0.85	0.84

datasets, while the threshold was not reached with all sample points for any soil property in the MO dataset and was reached for only one MI soil property. In the IL and IA datasets, some soil properties required only a few (i.e., 1 to 6) additional samples for method 4 to reach the 90% threshold, while the threshold was not met with other properties.

Comparison of method 4 (table 6) with method 3 (table 5) showed that in the cases where method 4 was able to achieve the 90% threshold, it generally did so with fewer sample points. We attributed this to the characteristics of method 4, including a relatively rapid increase in RPD with data from a few soil sample points and very little improvement after that, regardless of how many additional points were used for bias correction. In contrast, method 3, which included additional samples directly in the calibration equation, exhibited a more gradual increase in RPD as more sample points were added (fig. 3).

When including six or fewer additional points from the test dataset, RPD values of method 4 reached 60% to 90% of method 1 RPD values, showing that method 4 was effective in increasing RPD. Overall, it appeared that using two to six sample points for bias correction in method 4 increased RPD above method 2 by 0.2 to 1.0 for most soil properties. Only a few additional sample points were necessary to reach 60% to 90% of the method 1 RPD for many of the datasets. On average, method 4 was least accurate for the MO and MI datasets; for all others, maximum method 4 RPD was 86% of the method 1 RPD or greater (table 7). These same datasets were also lowest in mean method 1 RPD (table 2). While means for many soil properties (e.g., clay, CEC, SOC; table 1) were similar between the IL, IA, and SD datasets, mean values for the MO and MI datasets were often quite different. Evidently these two datasets were different enough in their soil property/DRS relationship that a simple bias adjustment was not sufficient for calibration improvement. In contrast, method 4 was quite effective for the other datasets, even providing RPD values greater than those from method 1 in a few cases (table 7).

For both method 3 (fig. 3) and method 4, the RPD results represent only one possible actualization of calibration set selection. To better understand the number of samples required for bias correction using method 4, a Monte Carlo sim-

ulation was employed. Ideally, this approach would have been applied to method 3 as well, but it was impractical to perform the several thousand additional PLS analyses that would have been required.

The Monte Carlo simulation was performed in the following manner. For each sample size from 1 to 34 (maximum number of cores at any site), a random calibration set was selected 100 times, and the resulting mean and standard deviation of the bias correction for each calibration set size were computed. The 95% confidence intervals were determined for each calibration set size, and the RPD for an observation at the edge of that confidence interval was computed by using a bias value that was 1.96 standard deviations from the mean estimate. Figure 5 displays the RPD values from the simulation. The true RPD values should be as good as or better than these 95 times out of 100.

RPD increased continuously (fig. 5) as the number of soil cores used for bias correction increased, up to about five or six cores, and then the rate of the RPD increase began to plateau. Thus, the maximum improvement due to bias correction (method 4) was obtained with around six soil cores, both for the single actualization of the calibration and with the Monte Carlo simulation. Method 4 results using the Monte Carlo simulation for other soil properties showed trends similar to those of CEC and SOC. Monte Carlo simulation also confirmed that only small numbers of additional sample points (i.e., two to six samples) were required to reach 60% to 90% of method 1 RPD values for many of the sites and soil properties.

COMPARING METHODS

Figure 6 compares RPD values averaged over all fields and soil properties for the four different calibration methods. For methods 3 and 4, results are given for one random selection of varying numbers of calibration cores (6, 10, 14, and all) from the test dataset. These numbers of additional calibration points (soil cores) were chosen for display because examination of the method 3 and method 4 results indicated that most of the improvement in RPD was achieved after the addition of between 6 and 14 cores.

As expected, the mean RPD was highest for method 1 and lowest for method 2. In methods 3 and 4, as the number of in-
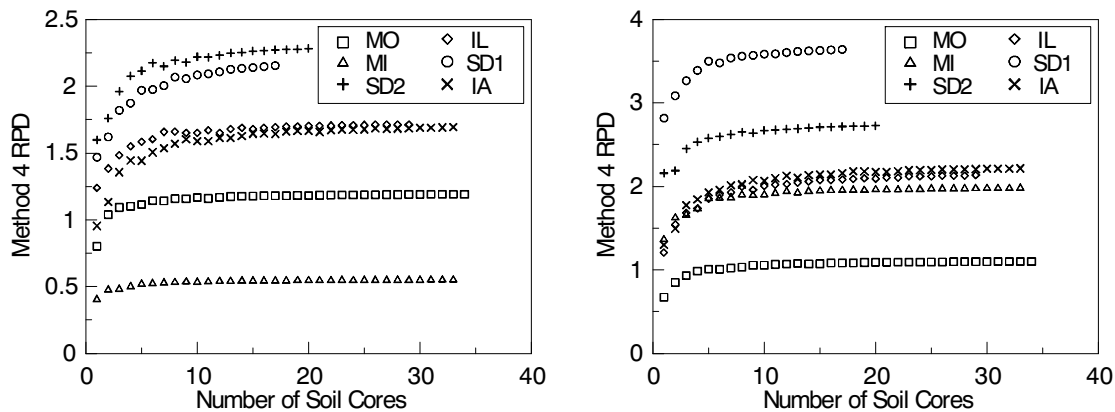
Figure 5. RPD of method 4 model with bias correction based on Monte Carlo simulation vs. number sample points included from the test dataset for CEC (left) and SOC (right).

cluded soil cores increased from 6 to 10, to 14, and to all available cores, mean RPD moved closer to that obtained by method 1. For method 3, as the number of additional soil cores increased, the RPD consistently increased. The rate of increase for the method 3 RPD varied depending on the soil property and site, although for many site/soil property combinations, and on average (fig. 6), there was significant increase in RPD to 14 cores and beyond. In method 4, inclusion of 6 calibration cores provided most of the increase possible, and additional soil cores did not generally increase the RPD value.

In general, method 3 required the addition of 8 to 20 sample points to the calibration model obtained by method 2 before 90% of the method 1 RPD value was achieved. This improved the RPD by 0.2 to 1.1. Method 4 required the addition of 2 to 6 sample points to achieve 60% to 90% of the method 1 RPD, while improving the RPD by 0.2 to 1.0. Results with both methods 3 and 4 depended to some degree on the order in which randomly selected points entered the calibration. Although we sometimes saw somewhat unstable
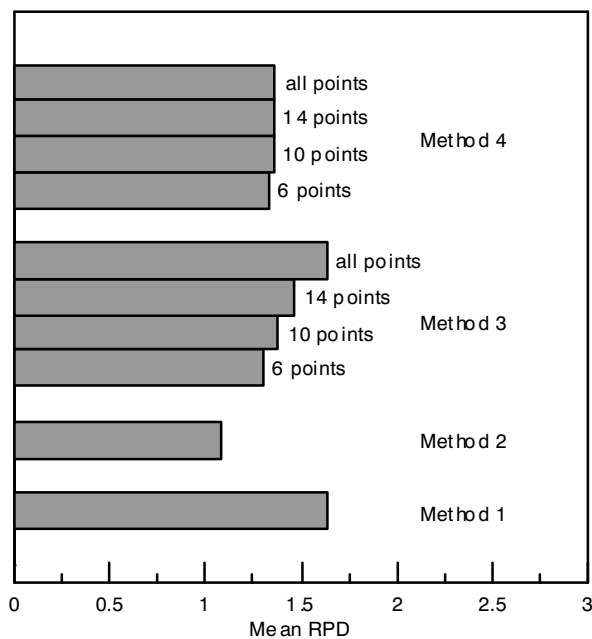

Figure 6. Mean RPD values obtained with each calibration method. Methods 3 and 4 included 6, 10, 14, and all soil cores from the test dataset in the calibration.

fluctuations in RPD when increasing the number of points in method 3 (fig. 3) and method 4, these fluctuations disappeared after more points (generally 6 or fewer) were added and each individual point had less influence on the overall result. More stable results might be expected by implementing calibration point selection based on spectral diversity, as described by Christy (2008).

Averaged across all data, the maximum (i.e., using all samples available in the test field) improvement in RPD by method 4 was 0.29, while the average maximum improvement by method 3 was 0.49 (fig. 6). Although method 3 could achieve higher overall accuracies, we consider method 4 more efficient because it required fewer calibration points within the test field and did not require a complete PLS recalibration for each new situation. Depending on the specific circumstances (e.g., ease and cost of field data collection and analysis vs. accuracy requirements), either method of calibration improvement might be preferred.

Questions remain about the degree of similarity required between calibration and prediction samples and how the required similarity can be achieved. A basic tenet of NIR prediction is that calibration samples must adequately represent the variability in the prediction set (Westerhaus et al., 2004). Although this concept is implemented relatively easily in NIR product analysis, soil heterogeneity, often at multiple spatial scales, makes it more difficult to select the required representative samples. Although variability in some soil properties of interest, such as nutrient levels and to a lesser extent SOC, may be related to farming activities and therefore be specific to fields, other properties such as texture and CEC are more related to mineralogy and soil-forming activities that act over landscapes rather than fields. Brown et al. (2005) reported needing local calibration samples, but their "locations" encompassed multiple fields. Results of this study, where our test datasets (except in the case of SD) consisted of multiple fields located on the same landscape, indicate that accurate predictions require including some calibration samples from that local area. However, similar to Brown et al. (2005), our results are not definitive on whether field-specific calibration is needed.

SUMMARY AND CONCLUSIONS

The broad objective of this research is to develop a reflectance-based soil property sensor for precision agricul-

ture. In this part of the study, different calibration methods were devised and compared to investigate sampling and calibration requirements for such a sensor. The data came from ten fields across the U.S. Corn Belt (MO, IL, MI, IA, and SD), from which 1.2 m deep soil cores were obtained, segmented by horizon, and analyzed in the laboratory for multiple soil properties. These data were compared with soil reflectance spectra from 1770 to 2500 nm obtained using a commercial spectrometer. Major findings were:

- Calibrations that did not include samples from the test dataset (method 2) resulted in lower RPD values than calibrations that included test dataset samples (method 1). Reductions in RPD ranged from 0.01 to 1.78, depending on dataset and soil property. Mean RPD reduction was greatest (1.05) for SOC. Thus, we conclude it is important to include samples similar in soils and crop management to each test field in the calibration data set. Whether calibration samples are required from each specific test field is a question for further investigation.

- When the number of field-specific sample points included in the PLS calibration increased (method 3, hybrid calibration), RPD first increased rapidly but then reached a plateau. In general, initial rates of RPD increase were greater for soil properties that exhibited larger overall RPD increases between methods 1 and 2. We conclude that the addition of several within-field calibration points to an initial (method 2) calibration could improve the accuracy of a NIR soil property sensor, and that a large portion of the potential improvement (e.g., 90% of method 1 RPD) could be obtained by adding about 8 to 20 points, depending on the soil property to be estimated.

- Based on the observation that a major difference between methods 1 and 2 was a bias offset, data from test field points were used to apply a bias correction to method 2 results (method 4, two-stage calibration). In general, bias correction increased RPD by 0.2 to 1.0 when two to six sample points were included. This indicated that a smaller number of additional samples would be required for sensor calibration using method 4 compared to method 3, but with somewhat reduced accuracy. Adding a small number of additional sample points allowed us to obtain 90% of the method 1 RPD values for many soil properties in some states; however, there were some soil properties that never reached 90% of method 1 RPD even when all the additional sample points were used. Method 4 findings, which were based on a single random order entry of samples in the analysis, were confirmed with Monte Carlo simulations.

These results provide guidance on sampling and calibration requirements for NIR soil property estimation. Additional data collection, further investigation using additional model selection criteria, and automation of these procedures are subjects for future study.

REFERENCES

Beebe, K. R., and B. R. Kowalski. 1987. An introduction to multivariate calibration and analysis. *Analytical Chem.* 59(1): 1007A-1017A.

Ben-Dor, E., and A. Banin. 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *SSSA J.* 59(2): 364-372.

Blanco, M., J. Coello, H. Iturriaga, S. Maspoch, and J. Pagès. 2000. NIR calibration in non-linear systems: Different PLS approaches and artificial neural networks. *Chemometrics and Intelligent Lab. Systems* 50(1): 75-82.

Bricklemyer, R. S., and D. J. Brown. 2010. On-the-go VisNIR: Potential and limitations for mapping soil clay and organic carbon. *Computers and Electronics in Agric.* 70(1): 209-216.

Brown, D. J., R. S. Bricklemyer, and P. R. Miller. 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma* 129(3-4): 251-267.

Chang, C. W., D. A. Laird, M. J. Mausbach, and C. R. Hurburgh Jr. 2001. Near-infrared reflectance spectroscopy - Principal component regression analysis of soil properties. *SSSA J.* 65(2): 480-490.

Christy, C. D. 2008. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Computers and Electronics in Agric.* 61(1): 10-19.

Confalonieri, M., F. Fornasier, A. Ursino, F. Boccardi, B. Pintus, and M. Odoardi. 2001. The potential of near infrared reflectance spectroscopy as a tool for chemical characterization of agricultural soils. *J. Near Infrared Spectroscopy* 9(2): 123-131

Cozzolino, D., and A. Moron. 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *J. Agric. Sci.* 140(1): 65-71.

Dalal, R. C., and R. J. Henry. 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near-infrared reflectance spectrophotometry. *SSSA J.* 50(1): 120-123.

Estienne, F., L. Pasti, V. Centner, B. Walczak, F. Despagne, D. Jouan Rimbaud, O. E. de Noord, and D. L. Massart. 2001. A comparison of multivariate calibration techniques applied to experimental NIR datasets: Part II. Predictive ability under extrapolation conditions. *Chemometrics and Intelligent Lab. Systems* 58(2): 195-211.

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. *SSSA J.* 56(3): 865-872.

Islam, K., B. Singh, and A. McBratney. 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Australian J. Soil Research* 41(6): 1101-1114.

Krishnan, P., J. D. Alexander, B. J. Butler, and J. W. Hummel. 1980. Reflectance technique for predicting soil organic matter. *SSSA J.* 44(6): 1282-1285.

Lee, K. S., D. H. Lee, K. A. Sudduth, S. O. Chung, N. R. Kitchen, and S. T. Drummond. 2009. Wavelength identification and diffuse reflectance estimation for surface and profile soil properties. *Trans. ASABE* 52(3): 683-695.

Malley, D. F., P. D. Martin, and E. Ben-Dor. 2004. Chapter 26: Application in analysis of soils. In *Near-Infrared Spectroscopy*

in Agriculture, 729-784. C. A. Roberts, J. Workman Jr., and J. B. Reeves III, ed. Madison, Wisc.: ASA, CSSA, and SSSA.

McCarty, G. W., J. B. Reeves III, V. B. Reeves, R. F. Follett, and J. M. Kimble. 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurements. *SSSA J.* 66(2): 640-646.

Mouazen, A. M., M. R. Maleki, J. D. Baerdemaeker, and H. Ramon. 2007. On-line measurement of some soil properties using a VIS-NIR sensor. *Soil and Tillage Research* 93(1): 13-27.

Nanni, M. R., and J. A. M. Dematte. 2006. Spectral reflectance methodology in comparison to traditional soil analysis. *SSSA J.* 70(2): 393-407.

Reeves III, J. B., and G. W. McCarty. 2001. Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and a fibre-optic probe. *J. Near Infrared Spectroscopy* 9(1): 25-34.

Shepherd, K. D., and M. G. Walsh. 2002. Development of reflectance spectral libraries for characterization of soil properties. *SSSA J.* 66(3): 988-998.

Sudduth, K. A., and J. W. Hummel. 1991. Evaluation of reflectance methods for soil organic matter sensing. *Trans. ASAE* 34(4): 1900-1909.

Sudduth, K. A., and J. W. Hummel. 1993. Soil organic matter, CEC, and moisture sensing with a portable NIR spectrophotometer. *Trans. ASAE* 36(6): 1571-1582.

Sudduth, K. A., and J. W. Hummel. 1996. Geographic operating range evaluation of a NIR soil sensor. *Trans. ASAE* 39(5): 1599-1604.

Sudduth, K. A., N. R. Kitchen, W. J. Wiebold, W. D. Batchelor, G. A. Bollero, D. G. Bullock, D. E. Clay, H. L. Palm, F. J. Pierce, R. T. Schuler, and K. D. Thelen. 2005. Relating apparent electrical conductivity to soil properties across the north-central USA. *Computers and Electronics in Agric.* 46: 263-283.

Viscarra Rossel, R. A., D. J. J. Walvoort, A. B. McBratney, L. J. Janik, and J. O. Skjemstad. 2006. Visible, near-infrared, mid-infrared, or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131(1-2): 59-75.

Westerhaus, M., J. Workman, Jr., J. B. Reeves III, and H. Mark. 2004. Chapter 9: Quantitative analysis. In *Near-Infrared Spectroscopy in Agriculture*, 133-174. C. A. Roberts, J. Workman Jr., and J. B. Reeves III, ed. Madison, Wisc.: ASA, CSSA, and SSSA.