

EVALUATION OF REFLECTANCE METHODS FOR SOIL ORGANIC MATTER SENSING

K. A. Sudduth, J. W. Hummel
ASSOC. MEMBER MEMBER
ASAE ASAE

ABSTRACT

Multivariate calibration methods, including stepwise multiple linear regression, principal components regression, and partial least squares regression, were applied to soil color and spectral reflectance data to predict the organic matter content of Illinois soils. Data were obtained with a visible/near infrared (NIR) spectrophotometer and a commercial colorimeter on a calibration set consisting of 30 Illinois mineral soils at 0.033 MPa and 1.5 MPa moisture tension levels. Good correlation with soil organic matter ($r^2 = 0.91$, standard error of prediction = 0.34% organic matter) was obtained when using as few as 12 NIR reflectance data points in the range from 1720 to 2380 nm. Color coordinate and visible reflectance data provided considerably poorer predictions of soil organic matter content. **KEYWORDS.** Herbicides, Color, Near infrared, Colorimeter, Spectrophotometry, Multivariate calibration

on the cation exchange complex of the organic matter. Meggitt (1970) reported on an experiment in which a linear increase in herbicide dose was required to maintain weed control as organic matter increased up to 15%, when using s-triazines, thiocarbamates, and pyridazinones. Gradual variations in organic carbon as high as 2:1 have been seen within an 80 m transect in a flat, apparently uniform field in Central Illinois (Sudduth, 1989); larger variations might be expected in areas with more variable topography or over longer distances. With this level of variability, significant opportunity exists for cost savings and input optimization of soil-applied herbicides based on soil organic matter content.

It is well known that soils of higher organic matter content are generally less reflective, or darker in color, than those soils with lower organic matter contents. Several reflectance and color based methods of predicting organic matter are evaluated in this article with respect to a representative set of Illinois agricultural soils.

INTRODUCTION

Cropping system management is generally practiced on a field basis, with herbicides, as well as other inputs such as fertilizer and seed, applied at a uniform rate over the entire field. Actual needs for herbicide inputs can commonly vary by a factor of two or more over the field so that a uniform rate based on the spatially-averaged requirements is too high for some areas and too low for others. Over-application results in increased production costs and may also cause environmental damage due to excess chemical contaminating surface water or groundwater. Under-application can result in decreased yields due to poor control of pests. Ideally, application rates should be adjusted to meet the requirements for each part of the field.

With many soil-applied herbicides, the application rate required for effective weed control increases as soil organic matter content increases, due to adsorption of the herbicide

LITERATURE REVIEW

Many researchers have correlated soil color properties with organic matter or organic carbon content.* The color properties used have included the Munsell coordinates of hue, value, and chroma; and coordinates from several of the Commission International de l'Eclairage (CIE) color spaces (Billmeyer and Saltzman, 1981). In general, attempts to extract a single relationship for soils from a wide geographic range have yielded poor correlations (Mc Keague et al., 1971; Karmanov and Rozhkov, 1972).

Studies focused on soils from a single state or other limited geographic areas have been more successful. Alexander (1969) correlated soil color with the organic matter content of over 300 Illinois soils and created a color chart which could be used for field classification of a moist soil into one of five overlapping organic matter ranges. Alexander reported that an experienced technician could determine organic matter content within the correct range more than 95% of the time.

Page (1974) used a commercially available color-difference meter to estimate the organic matter concentration in air dry surface samples of 96 Atlantic Coastal Plain soils. Organic matter content was estimated within $\pm 0.50\%$ of the actual value for 75% of the soils tested ($r^2 = 0.79$). Steinhardt and Franzmeier (1979)

Article was submitted for publication in September 1990; reviewed and approved for publication by the Electrical and Electronics Systems Div. of ASAE in May 1991. Presented as ASAE Paper No. 88-3038.

Trade names are used solely for the purpose of providing specific information. Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply the approval of the named product to the exclusion of other products that may be suitable.

The authors are Kenneth A. Sudduth, Agricultural Engineer, Cropping Systems and Water Quality Research Unit, USDA-Agricultural Research Service, Columbia, MO; and John W. Hummel, Supervisory Agricultural Engineer, Crop Protection Research Unit, USDA-Agricultural Research Service, Urbana, IL.

*In the context of this manuscript, the terms organic matter and organic carbon can be used interchangeably; soil organic matter content is defined as 1.72 times the organic carbon content.

correlated Munsell value and chroma and organic matter content for surface horizon samples of plowed silt loam soils in Indiana. A 90% accuracy rate was reported when classifying 219 soil samples into one of three categories – less than 3% organic matter, 3 to 5%, and over 5% .

Fernandez et al. (1988) correlated Munsell color and organic matter within a given soil landscape, hypothesizing that there would be a closer relationship than those previously reported for wider geographic areas. Samples collected from three soil series in each of two landscapes yielded high correlations between soil organic matter and Munsell value (moist soil $r^2 = 0.92$, dry soil $r^2 = 0.94$). Different calibrations were required at the two moisture levels, and the calibrations developed were applicable only within the landscapes studied, containing only silt loam and silty clay loam soils.

Reflectance measurement has been used extensively to estimate soil organic matter content. Two particular spectral regions have been utilized – the visible region, from 380 to 780 nm, and the near infrared (NIR) region, from the upper boundary of the visible region to about 3000 nm. Schreier (1977) found that the longer visible wavelengths had the highest predictive capability, with a maximum r^2 of 0.67 obtained at 600 nm. Obukhov and Orlov (1964) selected the 750 nm region, while Vinogradov (1981) reported the best correlations in the red range of 640 to 720 nm. Dalal and Henry (1986) used near infrared (NIR) reflectance to predict the organic carbon content ($r^2 = 0.86$) of selected Australian soils simultaneously with moisture and total nitrogen determination.

Baumgardner et al. (1970) correlated spectral reflectance in the 620 to 666 nm band with organic matter for 197 surface soil samples with organic matter contents from 0.7 to 6.3% . Two separate linear relationships, one below and one above 4% organic matter, best described the data. Stoner (1979) correlated average reflectance in 100 nm and wider spectral bands with organic carbon for 481 soils collected from across the continental U.S. They obtained an r^2 of 0.46 when considering all soils, and a maximum r^2 of 0.66 when analyzing correlations separately by climatic zone.

Krishnan et al. (1980) correlated reflectance characteristics in the 400 to 2400 nm range and organic matter content for ten Illinois soils at four moisture levels. Better correlations were obtained with information from the visible range (400 to 780 nm) than with NIR data (780 to 2400 nm). A first derivative model using optical density data yielded an r^2 of 0.85 with the calibration dataset. Other researchers (Pitts et al., 1986) could not obtain satisfactory correlations when using this model with an expanded set of soils. Pitts et al. (1986) successfully predicted organic matter contents of 30 Illinois soils using an exclusion algorithm and polychromatic (white), green, and red reflected light. The width of the prediction range for each soil varied from 1 to 3% organic matter, with an average range width of 1.4%.

Griffis (1985) developed and tested a soil carbon sensor which yielded an r^2 of 0.75 for a set of 18 air dry Arkansas soils. Shonk and Gaultney (1988) developed a real-time soil organic matter sensor intended to be recalibrated for each new soil landscape, rather than for the larger geographic area (such as an entire state) attempted by other

researchers. Laboratory tests yielded good correlations ($r^2 = 0.80$ to $r^2 = 0.98$) within a single landscape at a single moisture content. Field tests showed a curvilinear relationship between sensor output and organic matter, with independent calibrations needed for variations in travel speed or sensing depth.

Prediction of organic matter content from color or reflectance data requires the application of multivariate calibration techniques. The relationship between the dependent variable (organic matter content) and several independent variables (color coordinates or reflectance at various wavelengths) must be estimated over a number of samples (soils) in such a way as to be able to predict the value of the dependent variable for future samples. Several approaches can be used (Martens et al., 1983):

Multiple linear regression (MLR) is the simplest method to use. It is applicable in situations when there are many more samples than independent variables and when the independent variables have low noise and low inter-correlation.

Stepwise multiple linear regression (SMLR) can be used with sets of independent variables that are highly correlated. A search algorithm is used to find the subset of independent variables that provides the best description of the dependent variable. If there are few samples, SMLR may cause overfitting of the data, and the ability to predict to other than the calibration samples may be lost. Hruschka (1987) suggested that a minimum of ten samples were required for each estimated regression constant and for each varied parameter (such as wavelength).

Principal component regression (PCR) is an example of a latent variables regression method (Martens and Naes, 1987). Latent variables methods reduce the set of collinear independent variables to a smaller set of orthogonal components which represents most of the variability in the original data and contains a reduced amount of random measurement noise. The optimum number of components to use in the analysis can be found by the method of cross-validation. PCR consists of first reducing the dimensionality of the data by the use of principal components analysis (PCA), and then independently performing an MLR to relate the PCA factors to the dependent variable. PCA was first applied to soil spectral reflectance analysis by Condit (1970).

Partial least squares regression (PLS), another latent variables method, is similar to PCR but the components extracted from the data are a function of the values of both the independent and the dependent variables. The two steps of reduction in data dimensionality and regression are performed simultaneously so that small but relevant differences that might be ignored in a PCR are included in the solution. Since PLS components are selected to maximize both description of the independent variables and correlation to the dependent variables, the relevant predictive information tends to be concentrated in fewer components, making this method more efficient than PCR (Martens and Naes, 1987). Martens et al. (1983) stated that PLS should yield better results than other methods when the number of samples is low and/or when the calibration data is noisy. For an algorithmic description of the PCR and PLS techniques, the reader is referred to Martens and Naes (1987).

OBJECTIVE

The objective of this study was to select the optimum combination of wavelength range, data type, and calibration method for prediction of soil organic matter content from color coordinate or visible/NIR spectral reflectance data. If adequate predictive capability was obtained using color coordinates (or the related fundamental metamer representation), a tristimulus colorimeter could be used as a simple, relatively inexpensive sensor. Otherwise, a spectrophotometer or similar instrument would be needed to obtain reflectance data at points along the spectral curve.

PROCEDURE

Thirty Illinois mineral soils, selected as representative of the state's important agricultural soils, formed the calibration dataset used in this study (Table 1). The set included 10 soils initially used by Krishnan et al. (1980) and another 20 soils used in later studies by Pitts et al. (1986), Worner (1989), and Smith et al. (1987). The soil samples were prepared for analysis by removal of foreign material and crushing to pass through a 2 mm square mesh sieve.

SOIL PROPERTY DETERMINATION

Visible and NIR reflectance curves for the 30 test soils

TABLE 1. Organic carbon contents and physical characteristics of the 30 Illinois soils used in the study

Soil Name and Textural Class*	ID Number	Dry Combustion Organic Carbon (%)	Walkley Black Organic Matter (%)	Soil Texture			Gravimetric Moisture		
				Sand (%)	Silt (%)	Clay (%)	Air Dry (%)	1.5 MPa (%)	0.033 MPa (%)
Loamy Sand									
Ade	1	0.45	0.67	86.5	7.3	6.2	0.48	1.56	4.19
Plainfield	2	0.59	1.05	83.7	12.7	3.6	0.68	1.11	6.14
Sparta	3	0.69	1.73	85.4	10.4	4.2	0.70	1.41	6.22
Maumee	4	1.04	1.54	84.1	7.6	8.3	1.02	2.12	6.15
Sandy Loam									
Carmi	5	1.14	1.85	67.2	21.7	11.1	1.44	3.94	9.68
Loam									
Ambraw	6	1.26	2.25	48.0	29.2	22.0	2.19	8.74	15.47
Tice	11	0.99	1.33	25.8	50.0	24.2	2.27	8.41	20.45
Clay									
Jacob	7	2.01	2.95	3.8	33.6	62.6	7.55	24.43	36.85
Clay Loam									
Proctor	8	0.82	1.48	25.6	47.1	27.3	2.19	7.66	19.19
Darwin	9	1.35	2.13	34.5	33.9	31.6	2.73	10.97	20.67
Silt Loam									
Wynoose	10	0.94	1.50	6.3	79.0	14.7	1.30	5.28	21.97
Birkbeck	12	1.04	1.70	5.4	77.5	17.1	1.27	4.69	22.73
Shoals	13	0.74	1.67	27.8	59.6	12.6	1.00	4.30	19.76
Cisne	14	1.26	2.43	11.7	68.0	20.3	2.07	8.66	21.79
Bluford	15	0.77	1.77	20.3	66.9	12.8	1.27	4.03	20.61
Saybrook	16	1.26	2.53	12.7	62.8	24.5	2.35	8.49	21.69
Catlin	17	1.86	3.40	5.2	70.6	24.2	2.47	7.68	25.09
Saybrook	18	1.58	3.17	4.8	72.3	26.9	2.35	10.32	25.05
Cisne	19	1.56	3.17	11.5	66.3	22.2	3.18	10.32	22.41
Piopolis	22	1.53	2.40	9.4	65.8	24.8	2.20	9.93	27.06
Silty Clay Loam									
Flanagan	20	2.10	3.30	4.1	68.8	27.1	2.19	7.72	26.38
Jacob	21	1.12	1.72	5.9	57.2	36.9	3.66	16.66	28.46
Flanagan	23	1.84	3.08	6.3	67.1	26.6	2.58	9.34	23.97
Drummer	24	1.79	3.38	9.0	63.4	27.6	3.37	11.47	23.07
Flanagan	25	2.29	4.28	9.2	60.0	30.8	3.63	12.03	21.77
Drummer	26	2.91	5.70	8.7	61.0	30.3	4.05	13.44	23.33
Proctor	27	2.29	4.13	6.7	64.2	29.4	3.32	11.04	24.45
Flanagan	28	2.37	4.30	6.2	66.4	27.4	2.56	9.76	22.06
Drummer	29	3.16	5.32	12.6	55.9	31.5	3.96	13.64	26.23
Plano	30	1.81	3.20	7.9	65.6	26.5	2.79	9.21	21.29

* Textural classification, Walkley-Black organic matter, soil texture, and moisture content from Worner (1989).

were recorded at the Instrumentation and Sensing Research Laboratory (ISRL) of the USDA Agricultural Research Service in Beltsville, MD for Worner (1989). The primary set of 270 samples consisted of three replications of each of the 30 soils at each of three moisture tension levels – 0.033 MPa (1/3 bar, field capacity), 1.5 MPa (15 bar, wilting point), and air dry. Another set of 56 samples was prepared to provide additional information on the effects of soil moisture on reflectance. This supplementary set consisted of two replications of six soils (ID Nos. 4, 9, 16, 20, 23, and 27; Table 1) at each of four additional moisture tension levels (0.1 MPa, 0.4 MPa, 0.7 MPa, and 1.0 MPa). Two 895-point spectra were obtained for each sample: visible/NIR data from 400.8 to 1116.8 nm on a 0.8 nm bandpass, and NIR data from 802.0 to 2592.0 nm on a 2.0 nm bandpass.

Data were transformed to decimal reflectance (1.00 = a perfect reflector) by comparison to a powdered halon reflectance standard and recorded to four decimal places. A five-point moving average routine reduced the number of points per spectral curve from 895 to 297, producing datasets with an effective 2.4 nm (visible/NIR) or 6.0 nm (NIR) bandpass which could be more easily manipulated for analysis. Visual comparison of plotted data verified that no information was lost in this averaging operation, due to the smooth nature of the soil reflectance curves (fig. 1).

Walkley-Black organic matter, soil moisture content, and texture were determined by Worner (1989) for the samples analyzed at the USDA, ISRL. Total organic carbon content was determined by dry combustion of duplicate samples of each soil in a LECO Model HF10 induction furnace.

A Minolta Model CR-110 Chroma Meter was used to collect color coordinate data from a separate preparation of the 270 primary soil samples. The Minolta CR-110 was a handheld tristimulus colorimeter which used diffuse sample illumination and a 50 mm diameter sampling head to integrate surface texture and color variation effects. Data were obtained in the CIE Yxy format using the D₆₅ standard illuminant to simulate the spectral power distribution of natural daylight (Billmeyer and Saltzman, 1981). Readings were taken on three replicates of the 30 Illinois test soils at three moisture tension levels (air dry, 0.033 MPa, and 1.5 MPa). The samples were prepared with

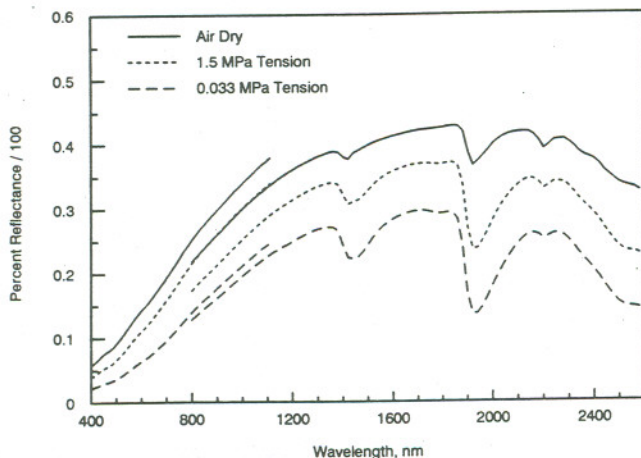


Figure 1—Mean spectral reflectance curves for Ade Loamy Sand (Soil 1) at three moisture tension levels.

a smoothed viewing surface in a 50 mm diameter × 10 mm deep flat black sample cup. Periodic recalibration of the meter against a reflectance standard showed no apparent drift in the x and y chromaticity coordinates, but in some cases a measurable drift (up to 1%) in the Y tristimulus value. Meter output was relatively insensitive to irregularities in the soil surface, as long as the distance between sampling head and soil surface was held constant.

DATA REDUCTION AND ANALYSIS

Unless specifically noted, all analyses used the carbon values obtained by dry combustion, rather than the Walkley-Black data. Calibrations were done with organic carbon (rather than organic matter) as the dependent variable, since carbon was the actual quantity measured with the induction furnace.

The regression dataset of primary interest consisted of all 30 test soils at the 0.033 MPa and 1.5 MPa moisture tension levels, to bracket the soil moisture conditions likely to be encountered during field operation of a soil organic matter sensor. The air dry samples were not included in the primary analyses as they would bias the regression toward the drier end of the moisture content range. Separate analyses of the data collected at each moisture tension were also completed, to eliminate that part of the variability due to soil moisture differences and to provide an indication of the upper limit of predictive capability.

Color Coordinate Data. Raw color coordinate data obtained from the Minolta Chroma Meter consisted of the tristimulus value, Y, and the chromaticity coordinates, x and y. Three other standard color coordinate sets – the tristimulus values (X, Y, Z) and the CIELAB and CIELUV uniform color coordinates (L*, a*, b*, and L*, u*, v*, respectively) – were calculated using standard equations (Billmeyer and Saltzman, 1981). A separate set of spectral tristimulus values was calculated from the spectral reflectance curves of the 270 primary soil samples, using the weighted-ordinate method with points spaced every 5 nm from 380 to 780 nm (Billmeyer and Saltzman, 1981). The CIE 1931 2° observer function and D₆₅ illuminant were used in the calculations, to correspond to the test conditions provided by the Minolta Chroma Meter. The spectral tristimulus values were used to obtain spectral sets of chromaticity coordinates and CIELAB and CIELUV coordinates.

Multiple regression of the colorimeter-derived and spectra-derived color coordinates against soil organic carbon content was accomplished using the PC-SAS computer package (SAS Institute, Inc., Cary, NC). Multiple linear and quadratic regressions were performed using the color coordinates in each of the systems described above. Logarithmic transformations of the color coordinate data were also used in an attempt to discover any multiplicative effects.

Spectral Reflectance Data. The results reported by Smith et al. (1987) were subjected to further analysis and validation tests. They had used SMLR on these same reflectance data to select the best wavelengths for prediction in four-wavelength and eight-wavelength models. The independent variables in these models consisted of the optical density [OD, defined as $\log_{10}(1/\text{reflectance})$] at wavelengths selected by an iterative scanning program to maximize the correlations obtained.

Smith et al. (1987) used the mean of the three replicate spectral reflectance curves for analysis. Their application of the four-wavelength model to the sixty averaged 1.5 and 0.033 MPa reflectance curves yielded an r^2 of 0.81, while the eight-wavelength model produced an r^2 of 0.86.

In this study, the calibration equations obtained by Smith et al. (1987) were applied to linearly independent validation datasets (the spectral reflectance curves of the individual replicate soil samples) to verify their predictive capabilities. Also, their iterative scanning program was modified by the addition of an outer loop to provide for iteration of the starting wavelengths for the scan. This allowed documentation of any variation in the prediction wavelengths and correlations due to the initial conditions for the iteration.

Partial least squares regression (PLS), implemented in the Unscrambler package (CAMO A/S, Trondheim, Norway), was used to analyze both the visible/NIR and the NIR datasets. Because limitations in the analysis programs required that the number of points per dataset be reduced from 297 to less than 100, cubic spline interpolation was used to produce datasets with 5, 10, or 20 nm point spacings.

Since soil color was thought to be a viable indicator of organic matter content, major emphasis was placed on analysis of the visible data (380-780 nm on a 5 nm spacing) as a separate entity. A supplementary analysis of the complete visible/NIR dataset (380 to 1110 nm on a 10 nm spacing) was also done to see if extending the sensing range would result in a better correlation to soil organic carbon. Both the raw reflectance data and the following transformations of that data were used:

Optical density (OD) transformation:

$$\log_{10}(1/\text{reflectance})$$

Kubelka-Munk transformation:

$$(1-\text{reflectance})^2/2*\text{reflectance}$$

Square root transformation.

The OD transformation is an extrapolation of the Beer-Lambert law, developed to quantify constituents present in transmissive samples. The Kubelka-Munk transformation is based on the theory of diffuse reflectance of a scattering medium, and its proportionality to concentration is more rigorously defined than that of OD. In many practical cases both transformations have given equally good results, particularly for NIR reflectance data (Birth and Hecht, 1987). The square root transformation, although not physically based, had yielded good results in preliminary tests.

The NIR dataset (800-2580 nm on a 20 nm spacing) was analyzed with PLS using raw reflectance data, OD transformed data, and Kubelka-Munk transformed data. Analyses were also done on subsets of the NIR data to determine what reduction in predictive capability would be seen if fewer individual reflectance readings with a wider bandwidth were sensed. Supplementary analyses of the NIR dataset documented performance of the prediction equations at additional moisture tension levels intermediate to the 0.033 MPa and 1.5 MPa calibration datasets.

In all PLS analyses, cross-validation techniques integral to the Unscrambler analysis program were used to verify the validity of the calibration equation. Outliers identified by the program were iteratively eliminated from the

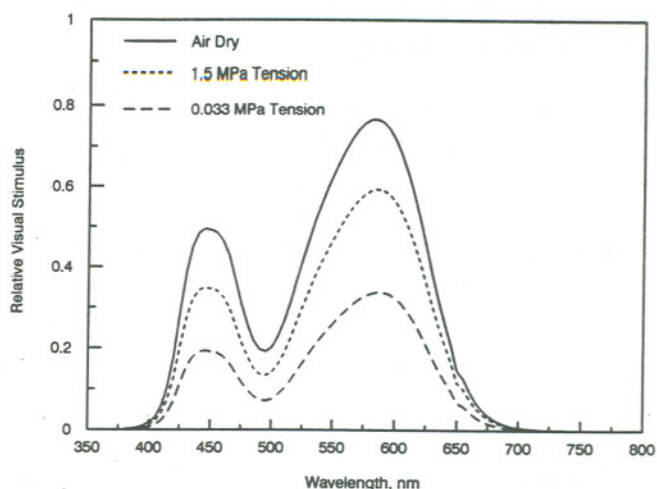


Figure 2—Fundamental metamer transformation of the mean spectral reflectance curves for Ade Loamy Sand (Soil 1).

calibration step to obtain an improved fit; however, these outliers were retained in the prediction step.

Principal Components Regression (PCR) was used to analyze the optical density transformed visible and visible/NIR datasets for the combined moisture tensions of 0.033 MPa and 1.5 MPa. The cross-validation and outlier removal techniques used for PLS were also used in the PCR analysis.

Fundamental Metamer Data. Both the visible reflectance data and the color coordinate data from the Minolta Chroma Meter were transformed to the fundamental metamer representation (fig. 2) using the methods of Cohen and Kappauf (1985). The fundamental metamer (that part of the spectral function which stimulates the human color-sensing mechanism) is in some sense intermediate between the spectral reflectance and color coordinate representations of a visual stimulus. The fundamental metamer is unique for each color stimulus (as are color coordinates), yet it contains information at each wavelength in the spectrum (as does the spectral reflectance function).

PLS was used to analyze the fundamental metamer data (380 to 780 nm on a 5 nm spacing) for the combined dataset with soil moisture tensions of 0.033 MPa and 1.5 MPa. Cross-validation and outlier elimination were used to obtain the most predictive model.

RESULTS AND DISCUSSION

SOIL PROPERTY DETERMINATION

The spectral reflectance curves (fig. 1) obtained at the USDA, ISRL exhibited a difference in reflectance between the overlapping portions of the visible and NIR spectra for each soil moisture tension, due to sample presentation. An open sample cup was used for the visible spectra, while a quartz window covered the cup used for the NIR spectra, resulting in lower readings.

Mean organic carbon obtained by dry combustion for the 30 test soils is reported in Table 1. The duplicate samples analyzed by dry combustion agreed within 0.1% carbon in all cases. Walkley-Black organic matter, soil texture and moisture content data determined by Worner

TABLE 2. Most predictive models selected by multiple regression analysis of Minolta Chroma Meter color coordinate data

Model	Sample moisture tension	r ²	SEE*
%C = f(L*, b*, L* ² , b* ²)†	0.033 and 1.5 MPa	0.72	0.36
%C = f(Y, x, y, y ²)	0.033 and 1.5 MPa	0.70	0.37
%C = f(x, x ²)	0.033 and 1.5 MPa	0.68	0.38
%C = f(x, x ²)	0.033 MPa	0.85	0.26
%C = f(x, x ²)	1.5 MPa	0.62	0.42
%C = f(x, x ²)	air dry	0.56	0.45

* SEE = standard error of the estimate, in percent organic carbon

† Model terms: L*, b* = coordinates from CIELAB color space
 % C = percent organic carbon
 Y = tristimulus value
 x, y = chromaticity coordinates

(1989) for the samples analyzed at the USDA, ISRL are also presented in Table 1.

EXPLORATORY DATA ANALYSIS

Color Coordinate Data. An analysis of variance (ANOVA) procedure detected no significant differences in color coordinates or moisture due to the three replications. Multiple regression models (Table 2) were identified as having the best predictive capability for the combined 0.033 and 1.5 MPa color coordinate data, and for the individual moisture tensions. A multiple quadratic regression using the L* (lightness) and b* (yellowness-blueness) coordinates from the CIELAB uniform color space yielded the highest coefficient of determination (r² = 0.72) and the lowest standard error of the estimate (SEE = 0.36% carbon), while maintaining significance of all terms in the model. The best statistically valid model obtained for the Yxy system of color coordinates was linear in Y and x, and quadratic in y. A quadratic regression using the x

TABLE 3. Most predictive models selected by multiple regression analysis of color coordinate data calculated from soil spectral reflectance curves

Model	Sample moisture tension	r ²	SEE*
%C = f(L*, b*, L* ² , b* ²)†	0.033 and 1.5 MPa	0.62	0.42
%C = f(L*, a*, L* ² , a* ²)	0.033 and 1.5 MPa	0.64	0.41
%C = f(Y, x, y, y ²)	0.033 and 1.5 MPa	0.65	0.41
%C = f(x, x ²)	0.033 and 1.5 MPa	0.59	0.44
%C = f(x, x ²)	0.033 MPa	0.84	0.28
%C = f(x, x ²)	1.5 MPa	0.51	0.48
%C = f(x, x ²)	air dry	0.58	0.44

* SEE = standard error of the estimate, in percent organic carbon

† Model terms: L*, a*, b* = coordinates from CIELAB color space

% C = percent organic carbon

Y = tristimulus value

x, y = chromaticity coordinates

chromaticity coordinate was the best significant model at a single moisture tension. Logarithmic transformations of the color coordinate data gave poorer results than the untransformed data, suggesting the absence of pure multiplicative effects. Results obtained with coordinates from the CIELUV uniform color space were similar in predictive capability and significance of model terms to the results obtained with coordinates from the CIELAB system.

Similar prediction models (Table 3) were applied to the color coordinate data obtained from the spectral reflectance curves. With the spectral data, a 4-term Yxy model (linear in Y and x, quadratic in y) was the best predictor of organic carbon. The best CIELAB model was quadratic in L* and a* (the redness-greenness dimension), but the quadratic L*b* model used for the colorimeter data was also a good predictor. The quadratic model in the x chromaticity coordinate was again the best predictor at a single moisture tension.

Comparison of the predictive capability of the two data sources (Tables 2 and 3) showed that higher correlations were generally obtained with the color coordinates from the handheld colorimeter than with the color coordinates calculated from spectral reflectance curves. However, even the best correlations found by color coordinate multiple regression analysis were not high enough to pursue this sensing method for development of a field organic matter sensor.

Spectral Reflectance Data. The 4-wavelength and 8-wavelength models developed by Smith et al. (1987) were used to predict the Walkley-Black organic matter content for each replication of the data individually, as well as for the entire dataset (Table 4). Although the 8-wavelength model had the better predictive capability for the calibration data, the 4-wavelength model was equally as good or better for all measures of prediction in the validation data. This indication of overfitting in the 8-wavelength model would probably have been more

TABLE 4. Validation results for stepwise multiple linear regression on 0.033 and 1.5 MPa moisture tension data

Prediction Measure	8-Wave-length Model	4-Wave-length Model
Calibration Step		
Coeff. of Determination (r ²)	0.86	0.81
Std. Error of Calibration (SEC)*	0.28	0.32
Validation Step		
SEP - all data†	0.36	0.34
SEP - rep 1 data	0.36	0.33
SEP - rep 2 data	0.39	0.36
SEP - rep 3 data	0.34	0.34
bias - all data‡	0.01	-0.01
bias - rep 1 data	-0.06	-0.04
bias - rep 2 data	0.05	-0.01
bias - rep 3 data	0.02	0.01

* SEC (standard error of calibration) is the standard error of the estimate in the calibration data, in percent organic carbon.

† SEP (standard error of prediction) is the standard error of the estimate in the validation data, in percent organic carbon.

‡ Bias is the difference between the predicted percent carbon mean and the measured percent carbon mean.

apparent if a completely independent validation dataset had been used.

The poorer predictions obtained with the 8-wavelength model could be explained by examination of the calibration equation. The 8-wavelength model was:

$$OMC = k_0 + k_1 * \frac{OD_{424.0} - OD_{808.0}}{OD_{575.2} - OD_{548.8}} + k_2 * \frac{OD_{1021.6} - OD_{1019.2}}{OD_{817.6} - OD_{820.0}} \quad (1)$$

where

- OMC = soil organic matter content, %,
- k_n = multiple regression coefficients,
- OD_n = optical density at wavelength n, specified in nm.

The numerator and denominator in the first quotient term were composed of the difference in OD at two distinct parts of the curve. However, the numerator and denominator in the second quotient term were both composed of the difference in OD between two adjacent points on the curve. The OD at these adjacent points never differed by more than 11 times the digital resolution to which the data was recorded, so the second quotient term had a low signal-to-noise ratio and poor predictive capability to other than the calibration data. The 4-wavelength model, described by equation 1 with k_2 set equal to zero, avoided the noise problems caused by the second quotient term in the eight-term model.

The prediction wavelengths selected by the iterative scanning program and the correlations obtained were found to be dependent on the seed (initial condition) wavelengths. Program runs using the 0.033 and 1.5 MPa visible/NIR dataset with 20 different sets of initial conditions chose 17 different sets of wavelengths as the optimum, with correlations ranging from $r^2 = 0.67$ to $r^2 = 0.81$. Some degree of robustness in the wavelength selection process was evidenced, as the wavelengths selected for the five best prediction equations were very similar. The numerator wavelengths were chosen near the two ends of the visible range, and the denominator wavelengths were chosen in

TABLE 5. Best sets of four prediction wavelengths for stepwise multiple linear regression on 0.033 and 1.5 MPa moisture tension data

r^2	Wave-length 1* (nm)	Wave-length 2 (nm)	Wave-length 3 (nm)	Wave-length 4 (nm)
0.81	424.0	808.0	575.2	548.8
0.79	872.8	424.0	563.2	575.2
0.78	436.0	800.8	553.6	575.2
0.77	431.2	764.8	541.6	575.2
0.76	817.6	407.2	532.0	575.2

* Wavelengths of absorbances used in the prediction equation:

$$OMC = k_0 + k_1 * \frac{OD_1 - OD_2}{OD_3 - OD_4}$$

where: OMC = soil organic matter content, percent
 OD_n = optical density at wavelength n

TABLE 6. Summary of partial least squares and principal components regression results with the visible-NIR dataset (380-1110 nm) at moisture tensions of 0.033 and 1.5 MPa

Dataset	Total Valid	Number of Factors*	SEP within 0.01 of min.	r^2	SEC†	SEP‡
Visible-NIR, 0.033 & 1.5 MPa						
Optical density transform§	10	6	0.81	0.30	0.32	
Optical density, PCR	11	7	0.76	0.33	0.33	
Visible, 0.033 & 1.5 MPa						
Reflectance	10	9	0.72	0.36	0.39	
Optical density transform	9	5	0.80	0.30	0.33	
Kubelka-Munck transform	9	5	0.77	0.32	0.37	
Square root transform	10	8	0.79	0.31	0.35	
Optical density, PCR	10	6	0.76	0.33	0.33	
Fundamental metamer from spectral data	3	3	0.50	0.48	0.48	
Fundamental metamer from color coordinate data	3	3	0.50	0.47	0.47	
Visible, 0.033 MPa						
Reflectance	9	8	0.87	0.24	0.30	
Optical density transform	5	3	0.85	0.27	0.27	
Visible, 1.5 MPa						
Reflectance	9	5	0.76	0.33	0.40	
Optical density transform	7	7	0.82	0.29	0.30	

* The maximum number of factors valid by cross-validation, and the number of factors necessary to achieve an SEP within 0.01 of the minimum.

† SEC (standard error of calibration) is the standard error of the estimate in the calibration data, in percent organic carbon.

‡ SEP (standard error of prediction) is the standard error of the estimate in the validation data, in percent organic carbon.

§ All analyses by partial least squares regression unless denoted by PCR (principal components regression).

|| Visible analyses included data from 380-780 nm.

the yellow-green range from 530 to 580 nm (Table 5).

Partial least squares (PLS) regression on the visible/NIR (380-1110 nm) reflectance data for 0.033 and 1.5 MPa moisture tensions yielded cross-validated correlations (Table 6) slightly better than the SMLR results obtained with the iterative scanning program. Correlations obtained when considering the visible (380-780 nm) data alone were slightly lower. As was the case with the color coordinate data, the 0.033 MPa moisture tension data were better correlated with organic carbon than were the 1.5 MPa moisture tension data. In general, the optical density (OD) transformed spectral data were more highly correlated with organic carbon than were the raw reflectance data or the other transformations. If similar correlations were obtained, the OD data required the extraction of fewer factors. In many of the datasets analyzed, the later factors extracted by the model were not major contributors to reducing the standard error of prediction (SEP, the standard error of the estimate in the validation data), and a smaller number of factors could provide an SEP within 0.01% organic carbon of the minimum (Table 6).

TABLE 7. Summary of partial least squares regression results with the NIR dataset (800-2580 nm) at moisture tensions of 0.033 and 1.5 MPa

Dataset	Number of points used	Number of Factors*		r^2	SEC†	SEP‡
		Total valid	SEP within 0.01 of min.			
20 nm spacing						
Reflectance	90	11	11	0.88	0.23	0.24
Optical density	90	12§	10	0.91	0.20	0.22
Kubelka-Munck	90	11	10	0.79	0.31	0.33
40 nm spacing						
Optical density	43	14	12	0.92	0.18	0.20
60 nm spacing						
Optical density	28	17§	12	0.92	0.18	0.20
Optical density, 11 points deleted	17	13	11	0.91	0.20	0.21
Optical density, 13 points deleted	15	13	11	0.91	0.20	0.21
100 nm spacing						
Optical density	16	13	12	0.90	0.22	0.22
Optical density, 3 points deleted	13	12	11	0.89	0.22	0.23

* The maximum number of factors valid by cross-validation, and the number of factors necessary to achieve an SEP within 0.01 of the minimum SEP.

† SEC (standard error of calibration) is the standard error of the estimate in the calibration data, in percent organic carbon.

‡ SEP (standard error of prediction) is the standard error of the estimate in the validation data, in percent organic carbon.

§ Maximum number of factors which could be extracted by the program, although additional factors may be valid. Results presented are based on this number of factors.

The combination of the NIR (800-2580 nm) dataset and PLS analysis yielded the best correlations between organic carbon content and soil reflectance properties of all the dataset/calibration method combinations tested (Table 7).

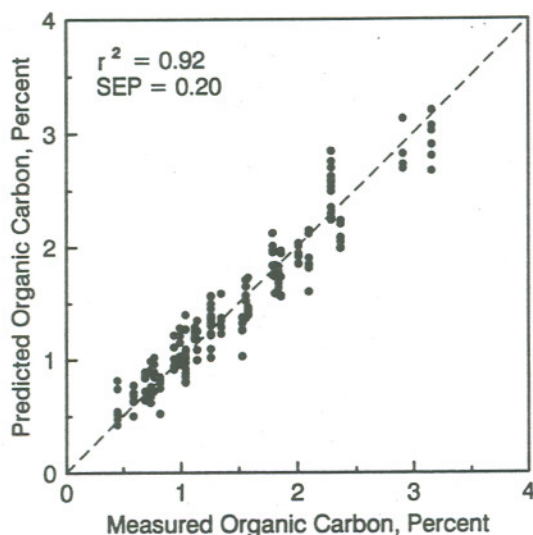


Figure 3—Predicted vs. observed organic carbon content obtained with partial least squares regression on 0.033 and 1.5 MPa moisture tension OD transformed data in the NIR wavelength range.

A maximum r^2 of 0.92 and a minimum SEP of 0.20% carbon were obtained (fig. 3) using OD transformed NIR datasets with either a 60 nm or a 40 nm spacing, created by averaging the adjacent two or three spectral points, respectively, in the initial 20 nm spacing PLS dataset. Most of the predictive capability of the data was retained after widening the data point spacing to 100 nm and eliminating points at wavelengths contributing little to the regression (Table 7). An r^2 of 0.91 and an SEP of 0.21 were obtained when only 15 of the original 90 independent variables were retained in the model.

The NIR spectra provided better predictions of organic carbon than did the visible spectra by removing some of the effects of soil moisture on soil reflectance properties. Spectral reflectance in the 1700-2000 nm region has been shown to be strongly correlated with soil moisture content (Dalal and Henry, 1986), due to the presence of a strong water absorption band centered at approximately 1940 nm (fig. 1). Examination of the PLS output showed that points in the 1940 nm water absorption band had the most influence on the factors extracted from the data, thus compensating for decreases in soil reflectance due to increased soil moisture. Another advantage of the NIR data was that noise due to the quantization of the data was a smaller fraction of the NIR spectra than of the visible spectra, due to the increased soil reflectance at the longer wavelengths.

Linear principal components regression on the OD transformed visible/NIR dataset provided slightly less predictive capability than did PLS regression on the same data (Table 6), and required the extraction of more regression factors.

Fundamental Metamer Data. PLS analysis of the fundamental metamer data showed this representation to have less predictive capability than reflectance or optical density datasets for the same wavelength and moisture tension range (Table 6). Only three factors were valid for the fundamental metamer data from either source (colorimeter or spectral reflectance curves), because the fundamental metamer has a rank of only three (Cohen and Kappauf, 1985). Visual comparison of the fundamental metamer curves for the 30 soils (for example, fig. 2), also revealed the low dimensionality of the data. All the curves had the same basic shape, with the only differences being in the heights of the two peaks and of the intermediate trough.

SELECTION AND REFINEMENT OF PREDICTION METHOD

Based upon the above results, the combination of NIR optical density transformed data and PLS calibration was selected as having the most promise for the prediction of soil organic carbon content. The NIR data exhibited a reduction in standard error of over 30% when compared to visible spectral reflectance and color data. PLS generally yielded slightly better calibrations than SMLR and PCR where all three methods were used on the same data. Additionally, the PLS software allowed calibrations to be generated with the least computational effort.

Accuracy of the selected NIR/PLS method appeared acceptable for the practical measurement of soil organic carbon. Published data quantifying infield organic carbon variations are rare, but carbon variations measured within 80 m in an apparently uniform Central Illinois field

TABLE 8. Summary of partial least squares regression results for reducing the range of the NIR (800-2550 nm) optical density transformed dataset at moisture tensions of 0.033 and 1.5 MPa

Range of center wavelengths	Number of points used	Number of Factors*		SEP within 0.01 of min.	r ²	SEC†	SEP‡
		Total valid					
40 nm data spacing							
850-2530 nm	43	14	12	0.92	0.18	0.20	
1650-2410 nm	20	11	9	0.91	0.20	0.21	
1730-2410 nm	18	11	9	0.91	0.20	0.21	
1730-2370 nm	17	12	9	0.91	0.20	0.21	
1730-2330 nm	16	11	9	0.90	0.22	0.22	
1770-2330 nm	15	13	10	0.86	0.25	0.25	
1730-2290 nm	15	10	10	0.88	0.23	0.23	
60 nm data spacing							
880-2500 nm	28	17§	12	0.92	0.18	0.20	
1300-2380 nm	19	13	10	0.92	0.19	0.20	
1480-2380 nm	16	14	10	0.92	0.19	0.20	
1660-2380 nm	13	11	8	0.91	0.20	0.20	
1720-2380 nm	12	11	10	0.91	0.20	0.20	
1780-2380 nm	11	9	9	0.87	0.24	0.25	
1720-2320 nm	11	10	9	0.89	0.22	0.22	

* The maximum number of factors valid by cross-validation, and the number of factors necessary to achieve an SEP within 0.01 of the minimum SEP.

† SEC (standard error of calibration) is the standard error of the estimate in the calibration data, in percent organic carbon.

‡ SEP (standard error of prediction) is the standard error of the estimate in the validation data, in percent organic carbon.

§ Maximum number of factors which could be extracted by the program, although additional factors may be valid. Results presented are based on this number of factors.

(Sudduth, 1989) were approximately ten times as large as the SEP's obtained with this method.

A sequence of analyses was completed to investigate the effect of a reduced wavelength range on the predictive capability of the regression. Since the optical density transformed NIR datasets with 40 nm and 60 nm data spacings were the most predictive over the full wavelength range (Table 7), they were chosen for further analysis. As data points were iteratively eliminated from the ends of the data range, the reduction in predictive capability was recorded (Table 8). For both datasets, the minimum range of wavelengths possible before the prediction degraded (in terms of an increase in the SEP of more than 0.01% organic carbon from that of the full-range data) was similar. For the data on a 40 nm spacing, that range was 1730 to 2370 nm (17 data points), with an SEP of 0.21. For the 60 nm data, the range was 1720 to 2380 nm (12 data points), with an SEP of 0.20.

The predictive capability of the 60 nm dataset ranging from 1720 to 2380 nm was verified for a range of additional moisture tensions. The prediction equation developed using the 0.033 and 1.5 MPa data was applied to a supplementary set of data collected on six of the test soils at intermediate moisture tensions of 0.1, 0.4, 0.7, and 1.0 MPa, yielding an SEP of 0.18% carbon. Extrapolation of the prediction equation developed with the 0.033 and 1.5 MPa data to the air dry data for the 30 soil samples did not give good results, with an SEP of 0.35 percent carbon. Inclusion of the data for air dry soil along with the 0.033 and 1.5 MPa moisture tension data in the calibration step

produced better correlations, with an r² of 0.86 and an SEP of 0.25.

CONCLUSIONS

1. Color coordinate data were not highly predictive of the organic carbon content of Illinois soils, with a maximum r² of 0.72 and a minimum standard error of 0.36% carbon obtained for combined 0.033 and 1.5 MPa moisture tension data obtained with a handheld colorimeter.
2. The fundamental metamer transformation of soil reflectance or color coordinate data was not predictive of organic carbon content.
3. An 8-wavelength model obtained by stepwise multiple linear regression (SMLR) was found to be inappropriate, due to overfitting of the calibration data. A valid 4-wavelength SMLR model yielded an r² of 0.81 and a standard error of prediction (SEP) of 0.36.
4. Partial least squares (PLS) regression on visible and near infrared (NIR) data from 380 to 1110 nm yielded a maximum validated r² of 0.81 and a minimum SEP of 0.32 for the combined 0.033 and 1.5 MPa data.
5. Data obtained at 0.033 MPa moisture tension were more predictive of carbon content than data at 1.5 MPa or the two datasets combined.
6. The visible/NIR dataset (380-1110 nm) was only marginally more predictive of organic carbon than was the visible (380-780 nm) dataset.
7. Optical density (OD) transformation of the soil reflectance data enhanced its organic carbon predictive capability.
8. NIR OD transformed data (800-2580 nm) and PLS analysis yielded the best predictions of organic carbon with the combined 0.033 and 1.5 MPa data.
9. Excellent correlation (r² = 0.91, SEP = 0.20) was retained when the NIR data were smoothed to a 60 nm data point spacing and the wavelength range reduced to 1720-2380 nm, for a total of only 12 data points used. Similar correlations were obtained with a 40 nm data spacing and a slightly smaller wavelength range.

REFERENCES

- Alexander, J. D. 1969. A color chart for organic matter. *Crops Soils* 21(8):15-17.
- Baumgardner, M.F., S.J. Kristof, C.J. Johannsen and A.L. Zachary. 1970. Effect of organic matter on the multispectral properties of soils. Ag. Expt. Station Journal No. 3939. West Lafayette, IN: Purdue University.
- Billmeyer, F.W. and M. Saltzman. 1981. *Principles of Color Technology*, 2nd ed. New York: John Wiley & Sons.

- Birth, G.S. and H.G. Hecht. 1987. The physics of near-infrared reflectance. In *Near-infrared Technology in the Agricultural and Food Industries*, eds. P.C. Williams and K.H. Norris, 1-15. St. Paul, MN: American Association of Cereal Chemists.
- Cohen, J.B. and W.E. Kappauf. 1985. Color mixture and fundamental metamers: Theory, algebra, geometry, application. *Am. J. Psych.* 98(2):171-259.
- Condit, H.R. 1970. The spectral reflectance of American soils. *Photogrammetric Eng.* 36:955-966.
- Dalal, R.C. and R.J. Henry. 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. *Soil Sci. Soc. Am. J.* 50:120-123.
- Fernandez, R.N., D.G. Schulze, D.L. Coffin and G.E. Van Scoyoc. 1988. Color, organic matter, and pesticide adsorption relationships in a soil landscape. *Soil Sci. Soc. Am. J.* 52(4):1023-1026.
- Griffis, C.L. 1985. Electronic sensing of soil organic matter. *Transactions of the ASAE* 28(3):703-705.
- Hruschka, W.R. 1987. Data analysis: Wavelength selection methods. In *Near-infrared Technology in the Agricultural and Food Industries*, eds. P.C. Williams and K.H. Norris, 35-55. St. Paul, MN: American Association of Cereal Chemists.
- Karmanov, I.I. and V.A. Rozhkov. 1972. Experimental determination of quantitative relationships between the color characteristics of soils and soil constituents. *Soviet Soil Sci.* 4(6):666-674.
- Krishnan, P., 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.
- Mc Keague, J.A., J.H. Day and J.A. Shields. 1971. Evaluating relationships among soil properties by computer analysis. *Can. J. Soil Sci.* 51(1):105-111.
- Martens, H., S. Wold and M. Martens. 1983. A layman's guide to multivariate data analysis. In *Food Research and Data Analysis*. London: Applied Science Publishers.
- Martens, H. and T. Naes. 1987. Multivariate calibration by data compression. In *Near-infrared Technology in the Agricultural and Food industries*, eds. P.C. Williams and K.H. Norris, 57-87. St. Paul, MN: American Association of Cereal Chemists.
- Meggitt, W.F. 1970. Herbicide activity in relation to soil type. In *Pesticides in the Soil: Ecology, Degradation, and Movement*, 139-141. East Lansing: Michigan State University.
- Obukhov, A.I. and D.S. Orlov. 1964. Spectral reflectivity of the major soil groups and possibility of using diffuse reflection in soil investigations. *Soviet Soil Sci.* 2(2):174-184.
- Page, N.R. 1974. Estimation of organic matter in Atlantic coastal plain soils with a color-difference meter. *Agron. J.* 66:652-653.
- Pitts, M.J., J.W. Hummel and B.J. Butler. 1986. Sensors utilizing light reflection to measure soil organic matter. *Transactions of the ASAE* 29(2):422-428.
- Schreier, H. 1977. Quantitative predictions of chemical soil conditions from multispectral airborne, ground, and laboratory measurements. In *Proc. 4th Canadian Symposium on Remote Sensing*, 106-112. Ottawa: Canadian Aeronautical and Space Institute.
- Shonk, J.L. and L.D. Gaultney. 1988. Spectroscopic sensing for the determination of organic matter content. ASAE Paper No. 88-2142. St. Joseph, MI: ASAE.
- Smith, D.L., C.R. Worner and J.W. Hummel. 1987. Soil spectral reflectance relationship to organic matter content. ASAE Paper No. 87-1608. St. Joseph, MI: ASAE.
- Steinhardt, G.C. and D.P. Franzmeier. 1979. Comparison of organic matter content with soil color for silt loam soils of Indiana. *Commun. in Soil Sci. and Plant Anal.* 10(10):1271-1277.
- Stoner, E.R. 1979. Physicochemical, site, and bidirectional reflectance factor characteristics of uniformly moist soils. Ph.D. diss., Purdue University, West Lafayette, IN (Libr. Congr. Card No. Mic. 80-15525).
- Sudduth, K.A. 1989. Near infrared reflectance soil organic matter sensor. Ph.D. diss., University of Illinois at Urbana-Champaign. Public. No. 90-15736. Ann Arbor MI: University Microfilms, Inc. (*Diss. Abstr. Int. B* 51(1):308).
- Vinogradov, B.V. 1981. Remote sensing of the humus content of soils. *Soviet Soil Sci.* 13(6):103-113.
- Worner, C.R. 1989. Design and construction of a portable spectrophotometer for real time analysis of soil reflectance properties. Unpublished M.S. thesis, Library, University of Illinois at Urbana-Champaign.