

미국 중부 토양의 이화학적 특성 추정을 위한 광 확산 반사 신호 전처리 및 캘리브레이션

나우정 Kenneth A. Sudduth 정선옥 김학진

Preprocessing and Calibration of Optical Diffuse Reflectance Signal for Estimation of Soil Physical and Chemical Properties in the Central USA

W. J. La K. A. Sudduth S. O. Chung H. J. Kim

Abstract

Optical diffuse reflectance sensing in visible and near-infrared wavelength ranges is one approach to rapidly quantify soil properties for site-specific management. The objectives of this study were to investigate effects of preprocessing of reflectance data and determine the accuracy of the reflectance approach for estimating physical and chemical properties of selected Missouri and Illinois, USA surface soils encompassing a wide range of soil types and textures. Diffuse reflectance spectra of air-dried, sieved samples were obtained in the laboratory. Calibrations relating spectra to soil properties determined by standard methods were developed using partial least squares (PLS) regression. The best data preprocessing, consisting of absorbance transformation and mean centering, reduced estimation errors by up to 20% compared to raw reflectance data. Good estimates ($R^2 = 0.83$ to 0.92) were obtained using spectral data for soil texture fractions, organic matter, and CEC. Estimates of pH, P, and K were not good ($R^2 < 0.7$), and other approaches to estimating these soil chemical properties should be investigated. Overall, the ability of diffuse reflectance spectroscopy to accurately estimate multiple soil properties across a wide range of soils makes it a good candidate technology for providing at least a portion of the data needed in site-specific management of agriculture.

Keywords : Precision agriculture, Soil sensor, Soil property, Reflectance spectroscopy, Partial least squares (PLS) regression

1. INTRODUCTION

Precision agriculture (PA), also called site-specific crop management (SSCM), has been widely used to improve farm management practices and thus reduce management costs, increase yield, and protect the environment from excessive application of agricultural chemicals such as fertilizers, pesti-

cides, and herbicides. When implementing PA, site-specific measurement of soil physical and chemical properties that affect soil quality and crop production is important. However, when using conventional sampling and analysis techniques, the time and cost required for the intensive sampling needed may be impractical. In this situation, sensor-based in-field data collection may be desirable, allowing the collec-

This research was conducted during the sabbatical leave of Professor Woo Jung La at the University of Missouri Biological Engineering Department and the USDA-ARS Cropping Systems and Water Quality Research Unit, Columbia, Missouri, USA in 2007-2008. The article was submitted for publication on 2008-09-01, reviewed on 2008-11-07, and approved for publication by editorial board of KSAM on 2008-11-28. The authors are Woo Jung La, Professor, Dept. of Bioindustrial Machinery Engineering and Institute of Agriculture & Life Sciences, Gyeongsang National University, Jinju, Republic of Korea, Kenneth A. Sudduth, Agricultural Engineer, Cropping Systems and Water Quality Research Unit, USDA-ARS, Columbia, Missouri, USA, Sun-Ok Chung, Full Time Instructor, Chungnam National University, Daejeon, and Hak-Jin Kim, Assistant Professor, School of Biosystems Engineering, Pusan National University, Miryang, Republic of Korea. Corresponding author: S. O. Chung, Dept. of Bioindustrial Machinery Engineering, Chungnam National University, 220 Gung-Dong, Yuseung-Gu, Daejeon, 305-764, Republic of Korea; E-mail: <sochung@cnu.ac.kr>.

tion of geographically referenced data on a much finer spatial resolution than is currently feasible with manual and/or laboratory methods (Sudduth et al., 1997).

One approach to rapidly quantify soil properties for site-specific management is optical diffuse reflectance sensing in visible and near-infrared wavelength ranges. Diffuse reflectance spectroscopy (DRS) is based on interaction between incident light and soil surface properties and variation in the characteristics of the reflected light due to the soil physical and chemical properties (Mouazen et al., 2005). Investigators have estimated soil physical and chemical properties using DRS in visible (VIS; 400-700 nm), near-infrared (NIR; 700-2500 nm), and mid-infrared (MIR; 2500-25000 nm) wavelengths. Mid-infrared spectra generally produce more accurate results, but the technology is more complex and expensive than for VIS-NIR measurement (Viscarra Rossel et al., 2006). Because of this added cost and complexity, most DRS soil sensing has been accomplished in the VIS, NIR, or VIS-NIR wavelength ranges.

One major advantage of DRS for soil analysis is that the same spectra can provide information about several soil physical and chemical properties. These include texture (sand, silt, and clay fractions) and pH, P, and K. Chang et al. (2001) evaluated the ability of near-infrared reflectance spectroscopy (NIRS) to predict multiple soil properties. They estimated CEC, sand, and silt with $R^2 > 0.80$, while exchangeable K, clay and pH were estimated with less accuracy ($0.50 < R^2 < 0.80$). Moron and Cozzolino (2003) also used NIRS to predict the content of sand and clay in soils of Uruguay. They obtained calibration R^2 values of 0.81, 0.83 and 0.92 for sand, silt, and clay fractions, respectively. Lee et al. (2003) related spectral characteristics and chemical properties of soil samples from major soil orders in Florida. Their models accounted for more than 72% of the variation observed in the validation set for soil pH, P, Ca, and Mg, but less than 50% of the variation in K and soil organic matter. The accuracy of reported spectral estimates of soil properties has varied widely. For example, R^2 for soil organic matter estimates has ranged from under 0.5 (e.g., Lee et al., 2003) to over 0.9 (e.g., Sudduth and Hummel, 1991). The accuracy of numerous past studies in estimating soil properties was reviewed by Viscarra Rossel et al. (2006).

Preprocessing of spectral data before application of the calibration model can improve accuracy (Duckworth, 2004). A commonly used transformation is from reflectance to absorbance, or $\log_{10}(1/\text{reflectance})$, to comply with Beer's Law. Another common transformation is mean centering, which generally enhances differences between samples and leads to more accurate calibration models. Additionally, various types of filtering may be used to increase the signal-to-noise ratio of the data. Because diffuse reflectance measurements may be complicated by non-homogeneous distribution of particles (soil in this case), several scattering corrections have been developed, including the standard normal variate (SNV), with or without detrending. Baseline shift due to instrument drift or other effects can also significantly affect spectra. Baseline issues are often addressed by a derivative transformation of the spectral data (Duckworth, 2004).

This research was part of an overall project to develop and apply soil sensing technology, including DRS, for precision agriculture management. In this specific study, the focus was on preprocessing and calibration methodology for DRS estimation of soil physical and chemical properties. Specific objectives were (1) to investigate several data preprocessing methods for application to the data and (2) to develop and determine the accuracy of calibrations relating spectra to laboratory-measured soil physical and chemical properties using partial least squares (PLS) regression.

2. MATERIALS AND METHODS

A. Soil Samples

A total of 37 surface soils, 17 from Illinois and 20 from Missouri, were used in this study. These soils were selected by Kim et al. (2007) to represent the diverse soil characteristics of important agricultural areas of Illinois and Missouri, and to provide a range of P and K concentration levels. The Illinois samples were sub-samples of soils used in previous studies (Sudduth and Hummel, 1991; Birrell and Hummel, 2001; Price et al., 2003) and the Missouri soils included 16 soils used by Coggeshall et al. (2005) and 4 soils collected from a long-term cropping system research site (Kitchen et al., 2005).

B. Laboratory Analysis

The soil samples used for laboratory analysis were ground and screened using a 2-mm sieve and then oven-dried at 105°C before extraction. Soil pH, organic matter (OM) by loss-on-ignition, and cation exchange capacity (CEC) by sum of bases were determined in the University of Missouri Soil and Plant Testing Laboratory. Clay, silt and sand fraction were obtained by the hydrometer method. Procedures for these analyses are documented in Nathan et al. (2006). Sub-samples of the 37 soils were provided to a commercial soil testing laboratory for P and K determination with an ICP (Inductively Coupled Plasma) spectrophotometer using methods described by Kim et al. (2007). Basic statistics of the laboratory measurements are given in table 1. For all except pH, the variation among the samples was quite large, as shown by the wide ranges and high CVs. Thus, these samples were expected to represent well the variability in these soil properties seen within cropped fields across the states of Missouri and Illinois.

C. Spectral Data Collection

Soil spectral data were obtained in the laboratory with a FieldSpec Pro FR¹) spectrometer (Analytical Spectral Devices, Boulder, Colo.). Spectra were recorded between 350 and 2500 nm and output on a 1-nm interval. The spectrometer used three detector systems: 1) for 350-1000 nm a silicon photodiode array, 2) for 1000-1830 nm an InGaAs detector, and 3) for 1830-2500 nm, an enhanced InGaAs detector. For reflectance data collection, subsamples of the soils collected in the field were air dried and sieved with a 2-mm screen.

Approximately 15 cm³ of soil was packed in a glass-bottomed sample cup for reflectance determination. The sample was illuminated through the glass by a halogen lamp (“MugLite” accessory of the spectrometer) and the reflected light was transmitted to the spectrometer through a fiber optic bundle. Each soil spectrum was obtained as the mean of 10 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 once every 30 minutes thereafter. A Spectralon (Labsphere Inc., North Sutton, N.H.) reflectance standard was scanned after every 10 soils and used to convert the raw spectral data to decimal reflectance. Three replicate measurements were obtained for each soil samples, resulting of 111 spectra (Fig. 1).

Reflectance data were preprocessed to improve stability of the regression. The first 100 readings (350-459 nm) and the

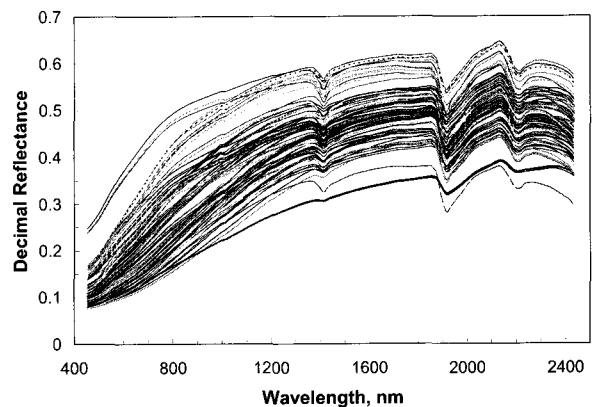


Fig. 1 Reflectance spectra for the soil samples (37 samples × 3 replicate measurements).

Table 1 Statistics of laboratory-measured soil properties

Soil property and units	Mean	Std. Dev.	Min.	Max.	CV, %
pH	5.87	0.70	4.3	7.1	12
OM, %	2.49	1.01	0.3	4.4	41
CEC, meq 100 g ⁻¹	13.2	5.5	3.4	27.5	41
Clay, %	17.2	6.7	4.0	33.0	39
Silt, %	55.4	19.7	9.0	83.0	36
Sand, %	27.3	24.1	4.0	87.0	88
P, mg L ⁻¹	74.0	51.9	21.3	235.8	70
K, mg L ⁻¹	174.5	109.8	51.9	445.5	63

1) Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by Gyeongsang National University, USDA-ARS, or their cooperators.

last 50 readings (2451–2500 nm) were deleted due to their low signal-to-noise ratio; thus the revised spectra began at 450 nm and ended at 2450 nm. (Fig. 1).

D. Analytical Procedures

Partial least squares (PLS) regression implemented in ParLes version 3.1 (Viscarra Rossel, 2008) was used to develop calibrations between soil properties and reflectance spectra. PLS has been widely used in chemometrics, remote sensing, and spectral data processing to deal with large datasets with highly correlated variables. PLS is a full-spectrum method, in that it uses information from all wavelengths in the original spectrum to develop a calibration algorithm. In PLS calibration, a new set of variables (factors) is created. The factors are statistically independent from one another and are constructed such that they capture variation in both the response (soil) and predictor (spectral) variables (Beebe and Kowalski, 1987). As in all multivariate regression, it is important to select the number of factors that best represents the calibration data without overfitting. To do this, we applied a leave-one-out-cross-validation procedure within the ParLes software. As suggested by Viscarra Rossel (2008), we selected the number of factors for each model at the minimum value of the Akaike Information Criterion (AIC), first given by Akaike (1969). ParLes calculated the validation dataset variance, coefficient of determination (R^2), root mean square error (RMSE) and RPD, the ratio of the standard deviation to the RMSE of cross-validation predictions. RPD is a useful measure of fit when comparing results from datasets containing different degrees of variability, with a higher RPD indicating a more accurate prediction (Williams, 1987).

What values of these fit parameters constitute a “good” model is somewhat subjective. Saeys et al. (2005) stated that R^2 values from 0.50 to 0.65 indicated the ability to discriminate between high and low concentrations, while approximate quantitative prediction, good prediction, and excellent prediction were possible from models with R^2 of 0.66 to 0.81, 0.82 to 0.90, and greater than 0.91, respectively. Their corresponding categories for RPD were 1.5 to 2.0, 2.0 to 2.5, 2.5 to 3.0, and above 3.0. Chang et al. (2001) classified the ability of NIR-DRS to estimate soil properties into three categories based on RPD ranges (< 1.4 ,

1.4–2.0, and > 2.0). After considering the criteria used by previous researchers, we defined the following categories: poor estimation when $R^2 \leq 0.65$ or $RPD \leq 1.5$, fair estimation with $0.65 < R^2 < 0.80$ or $1.5 < RPD < 2.0$, and good estimation when $R^2 \geq 0.80$ or $RPD \geq 2.0$ as suggested by Lee et al. (2007).

Many different preprocessing methods that may be appropriate in PLS spectral analysis (Duckworth, 2004) were available in the ParLes software (Viscarra Rossel, 2008). From these available methods, it was necessary to select the best method or methods for use with the specific dataset of this study. Based on past experience and preliminary data analysis, we investigated the following combinations of preprocessing steps on a subset of the dependent variables:

1. Raw reflectance data with no transformation
2. Absorbance transformation
3. Absorbance + mean centering
4. Absorbance + mean centering + SNV
5. Absorbance + mean centering + SNV + detrending
6. Absorbance + mean centering + median filter (rank 2)
7. Absorbance + mean centering + 1st derivative

Analysis of the complete dataset was then done using the preprocessing approach which provided the best results in this preliminary analysis.

3. RESULTS AND DISCUSSION

A. Selection of Preprocessing Method

The various preprocessing methods described above were applied for estimation of OM and pH (table 2). Considering both variables, two combinations provided the best results: absorbance transformation plus mean centering, and absorbance plus mean centering plus rank 2 median filter. Therefore, in the interest of simplicity, we chose the combination of absorbance plus mean centering for the estimation of all soil properties. This combination reduced RMSE by 5 to 20% compared to the untransformed reflectance data.

B. Calibration Model Development

Table 3 shows the leave-one-out cross-validated PLS results

Table 2 Comparison of data manipulation combinations for pH and OM by leave-one-out cross-validation. Optimum number of PLS factors was selected based on minimizing AIC

Data manipulation	Soil property	Optimum number of PLS factors	R ²	RMSE	RPD
Raw reflectance data with no transformation	pH	13	0.457	0.553	1.26
	OM	12	0.871	0.362	2.79
Absorbance transformation	pH	12	0.595	0.445	1.56
	OM	12	0.859	0.379	2.66
Absorbance + mean centering	pH	12	0.611	0.435	1.60
	OM	12	0.884	0.343	2.94
Absorbance + mean centering + SNV	pH	11	0.647	0.412	1.69
	OM	10	0.842	0.40	2.52
Absorbance + mean centering + SNV + detrending	pH	10	0.647	0.413	1.69
	OM	7	0.809	0.440	2.29
Absorbance + mean centering + rank 2 median filter	pH	12	0.613	0.434	1.61
	OM	12	0.885	0.341	2.95
Absorbance + mean centering + 1st derivative	pH	2	0.125	0.668	1.04
	OM	4	0.620	0.621	1.62

for all soil properties. Using the R² and RPD criteria defined above, there were good estimations of OM, CEC, clay, silt and sand ($0.833 \leq R^2 \leq 0.922$ and $2.45 \leq RPD \leq 3.60$), and fair estimation of P ($R^2 = 0.683$, $RPD = 1.77$). Estimations of pH and K were poor ($0.611 \leq R^2 \leq 0.612$), or fair ($1.60 \leq RPD \leq 1.61$) by these criteria, and therefore did not provide accurate estimates.

The estimates of texture fractions, OM, and CEC are comparable to some of the better results reported in the literature (Sudduth and Hummel, 1991; Chang et al., 2001; Moron and Cozzolino, 2003; Cozzolino and Moron, 2003). The lower accuracies for pH, P, and K may be due to the anthropogenic effects of fertilizer and lime additions in the agricultural fields from which the samples were collected.

On the other hand, soil texture, OM, and CEC are more long-term natural factors of soil development and weathering. It may be that the VIS-NIR reflectance approach can not successfully estimate soil chemical properties where the majority of the variation is due to the addition of chemical fertilizers. This observation is supported by Lee et al. (2007) who found that VIS-NIR data provided good estimates of pH across a soil profile, while pH estimates in surface soil, where the direct effect of lime addition was greater, were not good. Another issue may be that the laboratory P and K data used in this analysis was from standard methods calibrated to provide estimates of plant-available nutrients. Although reflectance analysis may be sensitive to chemical concentrations, it is likely not able to discriminate between

Table 3 PLS leave-one-out-cross-validation statistics

Soil Property	Optimum number of PLS factors	R ²	RMSE	RPD
pH	12	0.611	0.435	1.60
OM	12	0.884	0.343	2.94
CEC	12	0.833	2.229	2.45
clay	17	0.881	2.325	2.89
silt	17	0.901	6.185	3.19
sand	18	0.922	6.700	3.60
P	12	0.683	29.312	1.77
K	8	0.612	68.23	1.61

P and K that are plant available and those that are not.

Figure 2 shows scatter plots of measured vs. predicted values of the eight soil properties studied. For OM, CEC and clay, good agreement was obtained between the observed and predicted values for the full range. In the case of silt, a better prediction was displayed at lower and medium

values. On the other hand, in the case of sand, a better prediction was displayed at medium and higher values. This variation may have been due to the differences in texture characteristics of the samples, with many more having high silt and low sand than the opposite. Plots of pH, P, and K showed that the low accuracy for these variables was due

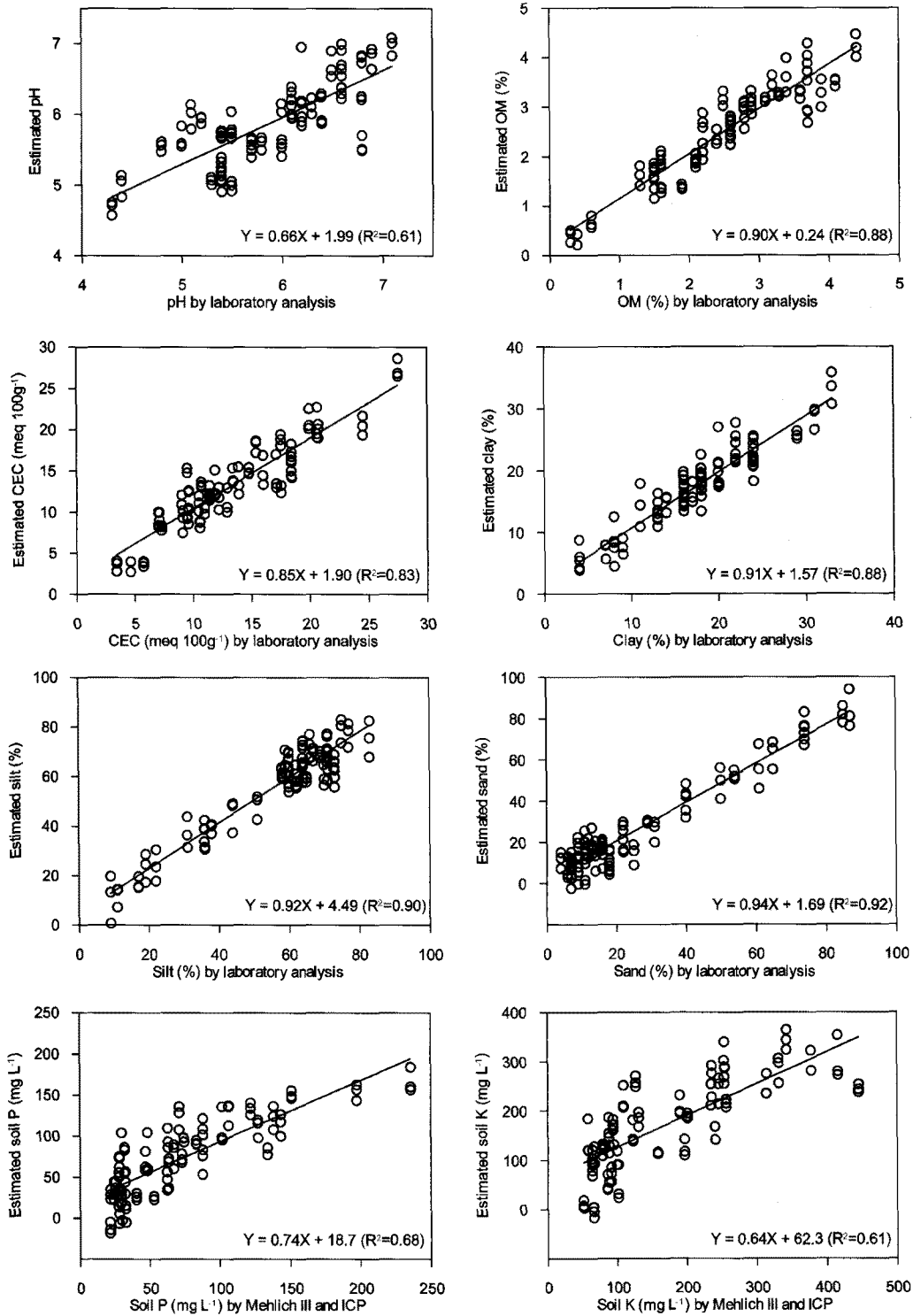


Fig. 2 Soil properties estimated by PLS analysis of spectral data as compared to those obtained by standard laboratory analyses.

to a large amount of scatter over the full range.

Overall, diffuse reflectance spectroscopy provided good results for direct estimation of many soil properties. One major advantage of diffuse reflectance spectroscopy for soil analysis is that many soil properties may be estimated from the same spectra. Also, the technique is rapid and requires relatively little sample preparation, making it possible to analyze many samples in a practical and timely manner. Combined with appropriate analysis methods such as PLS regression, this offers the possibility for considerable cost savings and increased efficiency over conventional laboratory analysis methods, an advantage particularly important with the level of analysis required for precision agriculture.

4. SUMMARY AND CONCLUSIONS

The purpose of this study was to investigate the reflectance approach for estimating physical and chemical properties of surface soil samples obtained from across the states of Missouri and Illinois. Diffuse reflectance spectra of air-dried, sieved samples were obtained from 350 to 2500 nm in the laboratory. Calibrations relating spectra to laboratory-measured soil properties were developed using partial least squares (PLS) regression. Major findings were:

- (1) The best data preprocessing method was found to be a combination of absorbance transformation and mean centering.
- (2) Good estimates ($R^2 = 0.83$ to 0.92) were obtained using spectral data for soil texture fractions, organic matter, and CEC. Estimates of pH, P, and K were not good ($R^2 < 0.7$).
- (3) For OM, CEC and clay, there was good agreement between observed and estimated values for the full range. For silt, a better estimate was produced at lower and medium values. On the other hand, for sand, a better estimate was produced at medium and higher values.

Diffuse reflectance spectroscopy, combined with appropriate preprocessing and analysis methods, successfully estimated many soil properties in this study. However, because this methodology was not able to provide good estimates of

specific chemical constituents (i.e., pH, P, and K), other technologies should be investigated for more accurate sensor-based measurement of soil chemical properties. In this study, a limited number of data (i.e., 37) was used for model development and leave-one-out-cross-validation, and further investigation would be necessary with a greater number of data and other validation methods.

ACKNOWLEDGEMENTS

This research was part of an international joint project, "Development of Site-specific Optimum Soil Management Strategies and In-situ Soil Physical and Chemical Property Sensors for Realization of Environment-friendly Precision Agriculture", between the Rural Development Administration, Republic of Korea and USDA-ARS, USA. We gratefully acknowledge the contributions of S. T. Drummond in data processing and analysis.

REFERENCES

1. Akaike, H. 1969. Fitting autoregressive models for prediction. *Annals Inst. Stat. Math.* 21:243-247.
2. Beebe, K. R. and B. R. Kowalski. 1987. An introduction to multivariate calibration and analysis. *Anal. Chem.* 59(1):1007A-1017A.
3. Birrell, S. J. and J. W. Hummel. 2001. Real-time multi ISFET/FIA soil analysis system with automatic sample extraction. *Comp. Elect. Agric.* 32:45-67.
4. Chang, C. W., D. A. Laird, M. J. Mausbach and C. R. Hurburgh. 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* 65:480-490.
5. Coggeshall, B., J. Lory and P. C. Scharf. 2005. Soil specific phosphorus and potassium buildup and critical values. In *ASA, CSSA, and SSSA Annual Meeting Abstracts [CDROM]*. Madison, Wisc.: ASA, CSSA, and SSSA.
6. Cozzolino, D. and A. Moron. 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *J. Agric. Sci.* 140:65-71.
7. Duckworth, J. 2004. Mathematical data preprocessing. *Near-Infrared Reflectance Spectroscopy, Agronomy Monograph* 44. pp.115-132. ASA, CSSA, and SSSA, Madison, Wisc.
8. Kim, H. J., J. W. Hummel, K. A. Sudduth and P. P. Motavalli. 2007. Simultaneous analysis of soil macronutrients using ion-selective electrodes. *Soil Sci. Soc. Am. J.* 71(6):1867-1877.

9. Kitchen, N. R., K. A. Sudduth, D. B. Myers, R. E. Massey, E. J. Sadler, R. N. Lerch, J. W. Hummel and H. L. Palm. 2005. Development of a conservation-oriented precision agriculture system: crop production assessment and plan implementation. *J. Soil Water Conserv.* 60:421-430.
10. Lee, K. S., D. H. Lee, K. A. Sudduth, S. O. Chung and S. T. Drummond. 2007. Wavelength identification for reflectance estimation of surface and subsurface soil properties. ASABE Paper No. 071046. ASABE, St. Joseph, Mich.
11. Lee, W. S., J. F. Sanchez, R. S. Mylavarapu and J. S. Choe. 2003. Estimating chemical properties of Florida soils using spectral reflectance. *Trans. ASAE* 46(5):1443-1453.
12. Moron, A. and D. Cozzolino. 2003. Exploring the use of near infrared reflectance spectroscopy to study physical properties and microelements in soils. *J. Near Infrared Spect.* 11:145-154.
13. Mouazen, A. M., J. D. Baerdemaeker and H. Ramon. 2005. Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. *Soil Till. Res.* 80: 171-183.
14. Nathan, M., Stecker, J. and Sun, Y. 2006. Soil Testing in Missouri: A Guide for Conducting Soil Tests in Missouri. Extension Circular 923, 4/06 revision. Missouri Cooperative Extension Service, Columbia, Mo. (available online at: <http://soilplantlab.missouri.edu/soil/soiltestinginmissouri.pdf>).
15. Price, R. R., J. W. Hummel, S. J. Birrell and I. S. Ahmad. 2003. Rapid nitrate analysis of soil cores using ISFETs. *Trans. ASAE* 46:601-610.
16. Saeys, W., A. M. Mouazen and H. Ramon. 2005. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosys. Eng.* 91(4): 393-402.
17. Sudduth, K. A. and J. W. Hummel. 1991. Evaluation of reflectance methods for soil organic matter sensing. *Trans. ASAE* 34:1900-1909.
18. Sudduth, K. A., J. W. Hummel and S. J. Birrell. 1997. Sensors for site-specific management. *The State of Site-Specific Management for Agriculture*, F. J. Pierce and E. J. Sadler (eds.), pp.183-210. ASA, CSSA, and SSSA, Madison Wisc.
19. Viscarra Rossel, R. A. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics Intelligent Lab. Sys.* 90:72-83.
20. 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* 131:59-75.
21. Williams, P. C. 1987. Variables affecting near-infrared reflectance spectroscopic analysis, *Near-infrared Technology in the Agricultural and Food Industries*, P. C. Williams and K. H. Norris (eds.), pp.143-167. American Association of Cereal Chemists, St. Paul, Minn.