ISSN 0367-6315 (Print) / ISSN 2288-2162 (Online)
Korean J. Soil Sci. Fert. 47(5), 319-323 (2014)
http://dx.doi.org/10.7745/KJSSF.2014.47.5.319

Short Communication

Predicting Soil Chemical Properties with Regression Rules from Visible-near Infrared Reflectance Spectroscopy

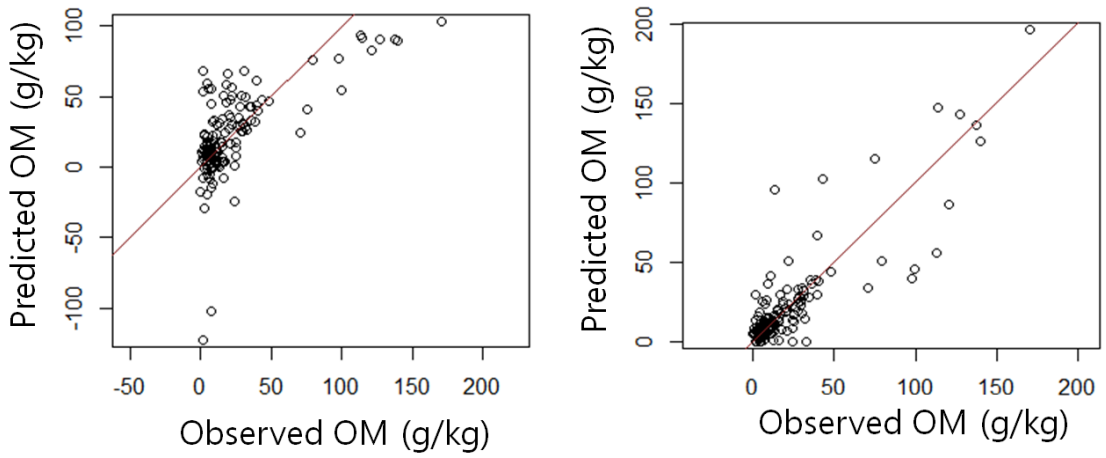
Suk Young Hong*, Kyungdo Lee, Budiman Minasny[1], Yihyun Kim[2], and Byung Keun Hyun

Department of Agricultural Environment, National Academy of Agricultural Science, RDA, Wanju, Korea
[1]Faculty of Agriculture and Environment, The University of Sydney, Sydney, Australia
[2]Research Policy Bureau, Rural Development Administration(RDA), Jeonju, Korea

(Received: October 13 2014, Revised: October 23 2014, Accepted: October 23 2014)

This study investigates the prediction of soil chemical properties (organic matter (OM), pH, Ca, Mg, K, Na, total acidity, cation exchange capacity (CEC)) on 688 Korean soil samples using the visible-near infrared reflectance (VIS-NIR) spectroscopy. Reflectance from the visible to near-infrared spectrum (350 to 2500 nm) was acquired using the ASD Field Spec Pro. A total of 688 soil samples from 168 soil profiles were collected from 2009 to 2011. The spectra were resampled to 10 nm spacing and converted to the 1st derivative of absorbance (log (1/R)), which was used for predicting soil chemical properties. Principal components analysis (PCA), partial least squares regression (PLSR) and regression rules model (Cubist) were applied to predict soil chemical properties. The regression rules model (Cubist) showed the best results among these, with lower error on the calibration data. For quantitatively determining OM, total acidity, CEC, a VIS-NIR spectroscopy could be used as a routine method if the estimation quality is more improved.

Key words: Soil chemical properties, visible-near infrared reflectance, spectroscopy, PCA, PSLR, Cubist



Scatter plot of PLSR (left) and Cubist (right) models with validation dataset to predict OM.

*Corresponding author : Phone: +82632382510, Fax: +82632383823, E-mail: syhong67@korea.kr
[§]Acknowledgement: This work was carried out by the support of Cooperative Research Program for Agriculture Science & Technology Development (PJ00997801), Rural Development Administration, Republic of Korea.

Introduction

Soil chemical properties are important indicators for soil quality, soil fertility, and soil health. However, soil chemical properties are expensive to measure or predict continuously over a region because they are highly heterogeneous across spatial scales (Pozdnyakova et al., 2005). Traditional measurement of soil chemical properties is time consuming and expensive (Bellon-Maurel and McBratney, 2011) and produces harmful waste (Ryu et al., 2001). To cover the large variation of soil chemical properties in a region, we need to analyze a large number of soil samples, thus traditional methods cannot be used to carry out for regional soil assessment. Reeves (2010) suggested that there is a need for a new rapid method which can give good quality data needed to cover soil's variation in a region.

The infrared spectroscopic techniques are promising, because they are low cost and are easy to use, which can be feasible for acquiring data for a large region. The use of infrared spectroscopy in agriculture started in 80s for measuring fruits and vegetations qualities. Near-infrared (NIR) spectroscopy has become well established in agricultural field (Wetzel, 1983). And more recently, NIR spectroscopy techniques have been developed as a useful quantitative tool for the prediction of various soil properties; including soil moisture, soil organic carbon, and total soil nitrogen content (Dalal and Henry, 1986; Morra et al., 1991; Reeves et al., 2002).

While spectroscopic techniques are easy to implement, they can produce a huge amount of data which can be difficult to handle using traditional statistical methods. Researchers have resolved this problem by applying data reduction methods, such as principal component analysis, partial least squares or rule-based regression (Bellon-Maurel and McBratney, 2011).

The objectives of this study are to predict soil chemical properties including OM, pH, Ca, Mg, K, Na, total acidity, CEC for Korean soils using visible-near infrared spectra, to develop prediction models using three methods (PCR, PLS and Cubist), and to validate accuracy between the models for finding the optimal prediction model.

Materials and methods

Soil samples Soil samples were taken from all over the South Korean region based on the soil series information. A total of 688 samples from 168 soil profiles which were taken as part of Soil Classification Project during 2009~2011 carried out by RDA. For each soil, about 4 kg of soil was taken from each horizon by a small shovel. All samples were transferred to soil testing laboratory in National Academy of Agricultural Sciences, Suwon (Korea). The samples were air dried at 60°C for 5 hours. From all samples, soil material was then used for measurement of soil chemical properties (OM, pH, Ca, Mg, K, Na, Total Acidity, Al, CEC) content in the laboratory with Korean soil analysis method (NIAST, 2000) as shown in Table 1. All soil samples were sieved (<2 mm) and ground with an agate mortar and pestle to reduce aggregated particles in the samples of reflectance spectra.

After laboratory measurement, the soil samples were placed into a 5 cm wide and 3 cm height sample holder without compression and leveled for the spectra reading. The visible

Table 1. Statistics of soil properties of 688 soil samples from 168 soil profiles used in this study.

Staistics	pH (1:5)	OM (g kg^{-1})	Exch. Ca ($cmol_c$ kg^{-1})	Exch. Mg ($cmol_c$ kg^{-1})	Exch. K ($cmol_c$ kg^{-1})	Exch. Na ($cmol_c$ kg^{-1})	Total Acidity (g L^{-1})	CEC ($cmol_c$ kg^{-1})
Average	5.8	19.5	4.8	2.0	0.38	0.32	17.4	18.7
Max	8.8	190.7	49.2	16.0	6.40	13.2	86.6	72.3
Min	3.6	0.1	0.1	0.1	0.1	0.02	0.3	0.9
Stdev	0.9	28.3	5.7	2.3	0.64	0.88	15.4	11.5

Fig. 1. Spectral reflectance measurement from 350 nm to 2,500 nm using ASD FieldSpec Pro(Applied Spectral Devices, Boulder, CO).

near infrared spectra from 350 nm to 2500 nm were measured with ASD FieldSpec Pro (Applied Spectral Devices, Boulder, CO) as shown in Fig. 1.

Spectral preprocessing and soil chemical properties transformation Averaging multiple measurements of reflectance were conducted so that the spectral artifacts can be removed. Therefore, 50 scans were taken for each sample and the average value was used. The original spectrum with 1 nm interval bands was resampled at every 10 nm from 500–2450 nm. The 1st derivative of the absorbance spectra (log[1/reflectance]) was calculated with the Savitsky-Golay algorithm for smoothing. Standard normal variate transformation (SNV) was applied for base line correction. 75 percent of the data were used for calibration and the rest of them for validation.

Prediction & validation In this study, three prediction methods were applied: principal components regression (PCR), partial least-squares regression (PLSR) and a regression rule model:—Cubist. All these methods were employed to quantify soil chemical properties from VIS/IR spectra.

PCR and PLSR are standard methods used in chemometrics and soil research to predict the soil properties from spectra. PCR simply takes the principal components (PCs) of the spectra and built a multivariate linear regression based on PCs. PLSR is a technique that attempts to combine PCA and multiple regression. It aims to predict a set of dependent variables (soil properties) by extracting from the spectra a set of 'orthogonal' factors (latent variables) which give the best prediction. The components in partial least squared are determined by not only by the predictor variables (as in PCR) but also the response variables. Meanwhile the Cubist model is a regression rule approach which consists of a collection of rules, that relates the independent variables (spectra) to a dependent variables (soil properties) of the form of:

If A[w_c1] > c1 and A[w_c2] > c2
Then
y = b0 + b1 * A[w_1] + b2 * A[w_2] + ⋯

where A[w] refers to the 1st derivative absorbance value at wavelength w, b are parameters of a linear model, c are the value of the conditions, and y is the target variable (e.g. OM content). A rule indicates that, whenever a case satisfies all the conditions, the linear model is appropriate for predicting the value of the target attribute.

Each rule is a linear model of the absorbance spectra, which is similar to a piecewise linear function introduced by Minasny and McBratney (2008). The collected soil data were split randomly into two parts: 75% samples were used for developing the prediction model, while the rest samples (25%) were used for validation of the prediction accuracy. All calculations were done in the R, an open source statistical software.

Results and Discussion

Cubist mostly showed the best prediction when compared to PCR and PLSR (Table 2). In the calibration data, Cubist shows good prediction for OM, exchangeable Ca, exchangeable Mg, total acidity, and CEC with R^2 values > 0.80., while it had a lower accuracy on the validation dataset for exchangeable Ca, and exchangeable Mg. Cubist could built good models for prediction of OM, total acidity and CEC, with R^2 values > 0.70 on the validation dataset. These values showed a good prediction capability.

Table 2. Goodness of fit for the prediction of Korean soil chemical properties using PCR, PLSR and Cubist (n. samples =708, where the calibration is 75% and validation is 25%. RMSE is the root mean squared error, and RPIQ is the ratio of prediction to inter-quartile range.

Soil chemical properties	Prediction methods		R^2	RMSE	bias	RPIQ
OM ($g\ kg^{-1}$)	PCR	Calibration	0.49	20.72	0.00	0.21
		Validation	0.45	22.36	2.19	0.23
	PLSR	Calibration	0.65	17.32	0.00	0.25
		Validation	0.44	24.65	-0.04	0.20
	Cubist	Calibration	0.96	5.51	-0.58	0.80
		Validation	**0.74**	**16.07**	0.74	**0.31**
pH (1:5)	PCR	Calibration	0.18	0.77	0.00	0.78
		Validation	0.13	0.91	-0.03	0.66
	PLSR	Calibration	0.41	0.65	0.00	0.92
		Validation	**0.49**	**0.70**	0.02	**0.86**
	Cubist	Calibration	0.66	0.50	-0.03	1.21
		Validation	0.43	0.75	-0.10	0.80

Table 2. Goodness of fit for the prediction of Korean soil chemical properties using PCR, PLSR and Cubist (n. samples =708, where the calibration is 75% and validation is 25%. RMSE is the root mean squared error, and RPIQ is the ratio of prediction to inter-quartile range (continue).

Soil chemical properties	Prediction methods		R^2	RMSE	bias	RPIQ
Exchangeable Ca ($cmol_c\ kg^{-1}$)	PCR	Calibration	0.11	5.40	0.00	0.43
		Validation	0.04	5.73	-0.36	0.52
	PLSR	Calibration	0.40	4.44	0.00	0.52
		Validation	**0.37**	6.09	0.47	0.49
	Cubist	Calibration	0.83	2.37	-0.13	0.97
		Validation	0.25	**5.46**	-0.22	**0.55**
Exch. Mg ($cmol_c\ kg^{-1}$)	PCR	Calibration	0.21	2.24	0.00	0.31
		Validation	0.14	1.47	0.38	0.48
	PLSR	Calibration	0.71	1.35	0.00	0.52
		Validation	0.09	2.49	-0.18	0.28
	Cubist	Calibration	0.83	1.06	-0.06	0.66
		Validation	**0.50**	**1.22**	0.08	**0.57**
Exch. K ($cmol_c\ kg^{-1}$)	PCR	Calibration	0.13	0.59	0.00	0.17
		Validation	0.01	0.80	-0.05	**0.13**
	PLSR	Calibration	0.71	0.34	0.00	0.29
		Validation	0.05	0.99	-0.04	0.10
	Cubist	Calibration	0.79	0.29	-0.01	0.34
		Validation	**0.12**	**0.80**	-0.07	0.12
Exch. Na ($cmol_c\ kg^{-1}$)	PCR	Calibration	0.06	0.86	0.00	0.00
		Validation	0.07	1.08	0.08	0.00
	PLSR	Calibration	0.61	0.55	0.00	0.00
		Validation	**0.29**	1.05	-0.06	0.00
	Cubist	Calibration	0.78	0.42	-0.05	0.00
		Validation	0.21	**0.99**	-0.02	0.00
Total acidity ($g\ L^{-1}$)	PCR	Calibration	0.43	11.91	0.00	0.35
		Validation	0.32	14.10	0.45	0.34
	PLSR	Calibration	0.75	7.94	0.00	0.53
		Validation	0.47	14.79	0.04	0.32
	Cubist	Calibration	0.94	3.80	-0.15	1.11
		Validation	**0.75**	**8.55**	0.31	**0.56**
CEC ($cmol_c\ kg^{-1}$)	PCA	Calibration	0.51	8.46	0.00	0.56
		Validation	0.39	9.24	0.70	0.47
	PLSR	Calibration	0.74	6.19	0.00	0.76
		Validation	0.62	7.51	-0.16	0.57
	Cubist	Calibration	0.94	2.99	0.02	1.57
		Validation	**0.79**	**5.51**	-0.28	**0.78**

Results of the Cubist model prediction of OM were plotted to compare the actual measured values and the predicted ones. Cubist model displayed a better prediction following the 1:1 line when compared to the PLSR (Fig. 2). In addition, the RMSE for Cubist were almost half of the PLSR prediction, indicating a higher accuracy. Cubist appeared to be the best model showing the highest R^2 values, lowest RMSE and highest RPIQ on the validation dataset. Minasny and McBratney (2008) stated that one of the advantages and reasons for its success in Cubist models is that they provide data grouping. It is possible for Cubist to separate data into more detailed groups to improve the accuracy of the prediction.

Since Korean soils used in this study display extreme variety over relatively small spatial scale, it would be better

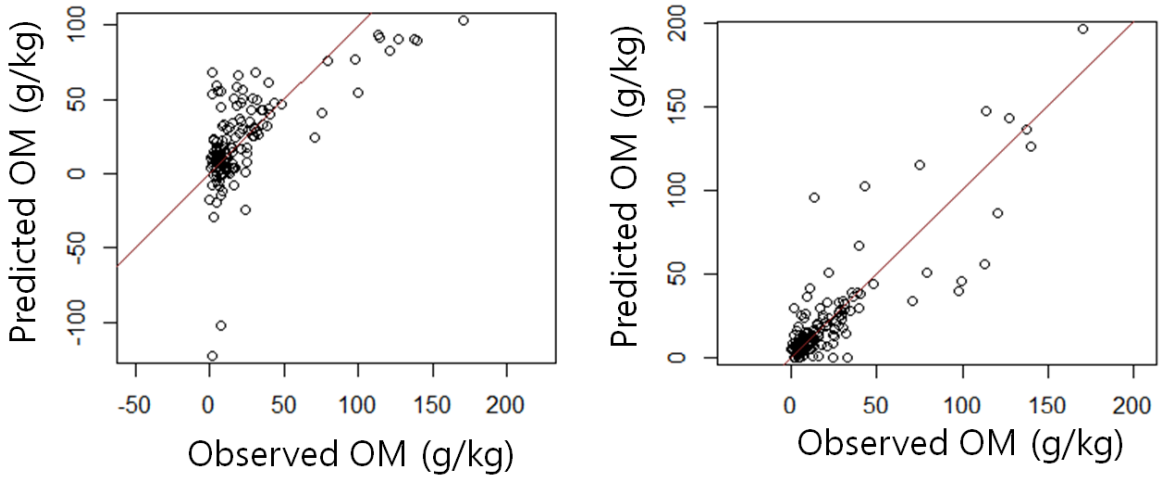


Fig. 2. Scatter plot of PLSR (left) and Cubist (right) models with validation dataset to predict OM.

using a rule-based regression model, which is easy to classify large datasets and provide clear linear relationships among data. The outputs of this study will be useful information to create digital mapping products of OM, Total Acidity, and CEC in Korea.

Conclusions

This study concluded that the rule-based regression model performed the best among others to predict soil chemical properties for the soil samples used in this study. In especially, for quantitatively determining OM, total acidity, CEC, soil analysis using visible-near infrared reflectance spectroscopy could be used as a routine method if the estimation quality is more improved.

References

Bellon-Maurel, V. and A. McBratney. 2011 Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils − Critical review and research perspectives. Soil Biology and Biochemistry. 43: 1398-1410.

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. of America J. 50: 120-123.

Minasny, B., and A. B. McBratney. 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemometrics and Intelligent Laboratory Systems. 94:72-79.

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

NIAST. 2000. Methods of soil and crop plant analysis. National Institute of Agricultural Science and Technology. Suwon, Korea.

Pozdnyakova, L., D. Gimenez,. and P. Oudemans. 2005. Spatial analysis of cranberry yield at three scales. Agronomy J. 97:49-57.

Reeves III, J. B. 2010. Near versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? Geoderma. 158:3-14.

Reeves III, J. B., G. W. McCarty,. and T. Mimmo. 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soil. Environmental Pollution. 116: 264-277.

Ryu, K. S., R. K. Cho, W. C. Park, and B. J. Kim. 2001. Use of NIR analyzer for measuring chemical properties of field soil. Korean J. of Soil Science and Fertilizer. 34(4):278-283.

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