COMPARING THE ARTIFICIAL NEURAL NETWORK WITH PARCIAL LEAST SQUARES FOR PREDICTION OF SOIL ORGANIC CARBON AND pH AT DIFFERENT MOISTURE CONTENT LEVELS USING VISIBLE AND NEAR-INFRARED SPECTROSCOPY[1]

Yücel Tekin[2], Zeynal Tümsavas[3] & Abdul Mounem Mouazen[4]

SUMMARY

Visible and near infrared (vis-NIR) spectroscopy is widely used to detect soil properties. The objective of this study is to evaluate the combined effect of moisture content (MC) and the modeling algorithm on prediction of soil organic carbon (SOC) and pH. Partial least squares (PLS) and the Artificial neural network (ANN) for modeling of SOC and pH at different MC levels were compared in terms of efficiency in prediction of regression. A total of 270 soil samples were used. Before spectral measurement, dry soil samples were weighed to determine the amount of water to be added by weight to achieve the specified gravimetric MC levels of 5, 10, 15, 20, and 25 %. A fiber-optic vis-NIR spectrophotometer (350-2500 nm) was used to measure spectra of soil samples in the diffuse reflectance mode. Spectra preprocessing and PLS regression were carried using Unscrambler® software. Statistica® software was used for ANN modeling. The best prediction result for SOC was obtained using the ANN (RMSEP = 0.82 % and RPD = 4.23) for soil samples with 25 % MC. The best prediction results for pH were obtained with PLS for dry soil samples (RMSEP = 0.65 % and RPD = 1.68) and soil samples with 10 % MC (RMSEP = 0.61 % and RPD = 1.71). Whereas the ANN showed better performance for SOC prediction at all MC levels, PLS showed better predictive accuracy of pH at all MC levels except for 25 % MC. Therefore, based on the data set used in the current study, the ANN is recommended for the analyses of SOC at all MC levels, whereas PLS is recommended for the analysis of pH at MC levels below 20 %.

Index terms: modeling, prediction, vis-NIR.

---

[1] Received for publication on April 8, 2014 and approved on July 23, 2014.

[2] Associate Professor, Uludag University, Vocational School of Technical Science. 16059, Bursa, Turkey. E-mail: ytekin@uludag.edu.tr

[3] Associate Professor, Uludag University, Agricultural Faculty. Department of Soil Science and Plant Nutrition. 16059, Bursa, Turkey. E-mail: zeynal@uludag.edu.tr

[4] Professor, Environmental Science and Technology Department, Cranfield University. MK43 0AL, United Kingdom. E-mail: a.mouazen@cranfield.ac.uk

RESUMO: *COMPARAÇÃO ENTRE REDE NEURAL ARTIFICIAL E REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS NA PREDIÇÃO DO CARBONO ORGÂNICO DO SOLO E DO pH, EM DIFERENTES NÍVEIS DE UMIDADE DO SOLO UTILIZANDO ESPECTROSCOPIA NO VIS-IVP*

A espectroscopia de infravermelho-visível próximo (vis-IVP) é amplamente usada para detectar propriedades do solo. O objetivo deste estudo foi avaliar o efeito combinado da umidade do solo e do algoritmo de modelagem na predição do carbono orgânico do solo (COS) e pH. Para tanto, foram utilizadas 270 amostras de solo para comparar a eficiência da predição da regressão por mínimos quadrados parciais (PLS) e da rede neural artificial (RNA) na modelagem do COS e pH, em diferentes níveis de umidade. Antes da determinação espectral, as amostras de solo secas foram pesadas para discriminar a quantidade de água a ser adicionada para que essa atingisse os níveis de umidade gravimétrica de 5, 10, 15, 20 e 25 %. Para obter a resposta espectral das amostras de solo, no modo de reflectância difusa, foi utilizado um espectrofotômetro de fibra ótica, na faixa do infravermelho-visível próximo (350-2500 nm). O pré-processamento dos dados espectrais e a análise pela regress do PLS foram implementados no software *Unscrambler®; enquanto na modelagem pela RNA, foi utilizado o software Statistica® (Version 11, StatSoft Inc. USA). Para o COS, o melhor resultado da predição foi obtido utilizando a RNA (RMSEP = 0.82 % e RPD = 4.23) para amostras de solo com 25 % de umidade. Para o pH, as melhores predições foram obtidas com a PLS para as amostras de solo seco (RMSEP = 0.65 % e RPD = 1.68) e para as de solo com 10 % de umidade (RMSEP = 0.61 % e RPD = 1.71). A RNA apresentou melhor desempenho para predição do COS em todos os níveis de umidade do solo; a PLS evidenciou melhor acurácia na predição do pH para todos os níveis de umidade, exceto para a umidade de 25 %. Dessa forma, a RNA é recomendada para análise de COS em todos os níveis de umidade do solo, enquanto a PLS é indicada para a análise do pH nos níveis de umidade inferiores a 20 %.*

Termos de indexação: modelagem, predição, vis-IVP.

INTRODUCTION

One of the most rapid and promising techniques of soil analysis for precision agriculture (PA) applications is visible and near infrared (vis-NIR) spectroscopy. It is a simple and non-destructive analytical method that can be used to enhance or replace conventional methods of soil analysis. It is particularly useful for overcoming some of the limitations of conventional laboratory methods and may be used to predict several soil properties simultaneously (Gholizadeh et al., 2013). Vis-NIR spectroscopy is becoming more and more attractive for end-users of PA, as recent research (Mouazen et al., 2007; Viscarra-Rossel & Chen, 2011; Tekin et al., 2013; Kodaira & Shibusawa, 2013) proves that it provides accurate quantification of the main physical and chemical soil properties useful for digital soil mapping. Although vis-NIR spectroscopy allows for rapid, cost effective, and intensive sampling, researchers admit shortcomings associated with instability of instrumentation from ambient conditions (e.g., light, temperature, etc.), transferability of calibration between different instruments, difficulties associated with the scale of the model (global, continental, regional, country, local and field) versus accuracy, and others (Stenberg et al., 2010; Mouazen et al., 2010).

Many researchers have successfully measured soil organic carbon (SOC) using vis-NIR spectroscopy (Mouazen et al., 2007; Gomez et al., 2008; Vasquez et al., 2008; Leone et al., 2012; Tekin et al., 2012). A comprehensive analysis of the literature was carried out by Stenberg et al. (2010), confirming the possibility of successful measurement of SOC with vis-NIR, which was attributed to the direct spectral response of C in the NIR range. In contrast, pH does not have a spectral response and it is a difficult element to measure through the vis-NIR technique. However, there are some reports exhibiting some degree of success. Marin-Gonzalez et al. (2013) reported that predictive accuracy for the laboratory and on-line measurements was classified as excellent/very good for pH (Residual prediction deviation - RPD) = 2.69 and 2.14 and R^2 = 0.86 and 0.78, respectively). In another study, a model for prediction of pH based on the RPD showed moderate accuracy ($1.5 < RPD < 2.0$) (Cohen et al., 2007).

For spectroscopy analysis, there are many factors affecting the diffuse reflectance spectra of soils (Stenberg et al., 2010), such as texture (Mouazen et al., 2005), color, and MC (Mouazen et al., 2006). In fact, one of the major influences on the accuracy of prediction of soil properties through vis-NIR spectroscopy is MC. Mouazen et al. (2006) found that variable soil MC decreased the predictive accuracies of several soil properties, including total C and pH. Morgan et al. (2009) arrived at the same conclusion for SOC and inorganic C. Likewise, Tekin et al. (2012),

using the same data set as that of the current study, reported significantly improved results for prediction of SOC using dry soil samples, as compared to wet ones. These authors concluded that MC significantly affects the predictive performance of both SOC and pH, although this effect was found to be greater for the former than for the latter.

In order to model the complex relationship between spectral signatures and a soil property, multivariate regression methods have an advantage over simple bivariate relationships, based on, for example, peak intensity measurements (Soriano-Disla et al., 2014). PLS is the most common technique currently adopted to model the relationships between infrared spectral intensities characteristic of the soil components and soil properties through derived PLS loadings, scores, and regression coefficients (Janik et al., 2009). Although it is a linear regression method, it can be forced to adopt nonlinearity either by using additional PLS factors or a nonlinear preprocessing function (Janik et al., 2009). PLS regression finds a series of components or latent vectors that provide a simultaneous reduction or decomposition of X and Y such that these components explain, as much as possible, the covariance between X and Y (Summers et al., 2011). One of the advantages of PLS regression compared to other chemometric methods, e.g., principal component regression analysis, is the possibility of interpreting the first few latent variables because they show the correlations between the property values and the spectral features (Yang & Mouazen, 2012). Bellinaso et al., (2010) concluded that principal component analysis grouped soils originating from similar parent materials, with some differentiation caused by altitude. Different researchers showed that PLS regression can provide high modeling performance (Bogrekci & Lee, 2005; McCarty et al., 2002; Vasquez et al., 2012). However, other researchers reported that other modeling techniques, e.g., ANN, can provide higher predictive accuracy than PLS regression (Viscarra-Rossel & Behrens, 2010; Kodaira & Shibusawa, 2013). The ANN can potentially deal better with non-linear spectral responses than PLS and it has been proposed as a means of achieving better predictive accuracy (Zhao et al., 2006). Mouazen et al. (2010) concluded that the best predictive accuracy was obtained for SOC with a back propagation neural network based on data of PLS latent variables (LVs). Authors have not attempted to study the combined effect of MC and the modeling technique on predictive accuracy. The hypothesis of this study is that the ANN provides better results for the prediction of SOC and pH than PLS at any MC level studied.

The objective of this study is to evaluate the combined effect of MC and the modeling algorithm on the prediction of SOC and pH. We will compare the PLS and the ANN for modeling of SOC and pH at different MC levels.

MATERIALS AND METHODS

Soil samples and laboratory analysis

A total of 270 soil samples were used in this study - 150 samples were collected from the Bursa region of Turkey and 120 samples from different counties across the United Kingdom. They were collected from the top layer (0-20 cm) of arable, fruit and vegetable fields. Soils were put in an airtight nylon bag, labeled, and placed in cold storage at -4 °C. Soil pH was measured in a 2:1 water-soil slurry (deionized water:air dried soil) (McLean, 1986). The SOC was measured through the Walkley-Black method (Nelson & Sommers, 1982). Statistics of the Turkish and UK soil samples are shown in table 1.

Optical measurements

Each soil sample was carefully mixed, and surface material, plant residues, and stones were removed. All samples were dried in a laboratory oven at 65 °C for 24 h, ground, and sieved in a 2 mm sieve. Before spectral measurement, dry samples were weighed to determine the amount of water by weight to be added to achieve the specified gravimetric MC levels of 5, 10, 15, 20, and 25 % (Tekin et al., 2012). To achieve homogeneous MC distribution, samples in cups after wetting were closed with plastic covers overnight until optical measurements. This method was repeated for

Table 1. Statistics of soil organic carbon (SOC) and pH in water of the Turkish and UK soil samples

	SOC		$pH(H_2O)$	
	Turkish soil	UK soil	Turkish soil	UK soil
	%			
Max	3.98	14.20	8.02	8.10
Min	0.16	3.10	4.14	3.60
Mean	1.30	7.79	6.61	5.03
Standard deviation[(1)]	0.65	2.50	0.76	1.11

[(1)] Derived from Tekin et al. (2012).

every 5 % MC interval until the MC in the soil samples reached 25 %.

A fiber-optic vis-NIR spectrophotometer (350-2500 nm) (LabSpec2500 Near Infrared Analyzer, Analytical Spectral Devices, Inc, USA) was used to measure spectra of soil samples in the diffuse reflectance mode. It has one Si array (350-1000 nm) and two Peltier cooled InGaAs detectors (1000-1800 and 1800-2500 nm). The sampling interval of the instrument was 1 nm. However, spectral resolution was 3 at 700 nm and 10 at 1400 and 2100 nm. A high intensity probe with a built-in light source was used. A quartz-halogen bulb constituting a 3000 K light source and a detection fiber were gathered in the high intensity probe enclosing a 35º angle. Before scanning, soil in a cup was gently pressed before leveling with a spatula. This resulted in a smooth soil surface, which ensured a maximum diffuse reflection and, thus, a good signal-to-noise ratio (Mouazen et al., 2005). Soil samples were placed in direct contact with the high intensity probe. Three reflectance spectra were measured from each soil sample by rotating the cup at a 120º angle. In order to achieve a representative spectrum of a soil sample, three soil sample replicates were considered in this study. A total of 10 scans were measured from each spot, and these were averaged in one spectrum. The three spectra were averaged in one spectrum, which was used for data analysis.

Processing of soil spectra and development of calibration models

Spectral preprocessing and PLS regression were carried using Unscrambler® software (Version 9.8, 1986-2003, Camo A/S, Oslo, Norway). Statistica® software (Version 11, StatSoft Inc., USA) was used for ANN modeling. The main aim of spectral preprocessing is to remove the noisy part in the spectrum or to eliminate some sources of variation not related to the measured value. The noisy part of the spectrum was found to be at the 305-401 and 2423-2500 nm ranges due to low reflectance of the soil and lower sensitivity of the instrument at these wavelengths. Spectra at wavelengths of 401-2423 nm were used for calibration, whereas the remaining parts were cut from the spectra. To achieve better result for calibration models, different data pre-processing options were used. A trial and error process was performed to discover the best pre-processing procedure, and the final selection of a pre-processing method was based on comparing the results (e.g., the root mean square error of prediction (RMSEP) and the RPD of the different models). The best pre-processing method of soil spectra included reducing the wavelengths by averaging four adjacent wavelengths into one wavelength for SOC, and by averaging three wavelengths for pH. The wavelength average was followed successively by maximum normalization for SOC and mean normalization for pH, the Savitzky & Golay 1st derivative (Savitzky & Golay, 1964), and, finally, smoothing to remove the noise.

After spectral preprocessing, PLS regression was carried out to develop calibration models for SOC and pH at different MC. The PLS regression is a bilinear regression method that extracts a small number of factors, which are a combination of the independent variables, and uses these factors as a regression generator for the dependent variables or chemically-measured values (Maleki et al., 2006). Although PLS regression has many advantages, such as its simplicity, robustness, predictability, precision, and clearly quantitative explanations, the main disadvantage is that PLS regression does not provide a quantitative explanation for the relationship between predictor variables and response variables, and it does not support re-use of model algorithms between different instrumentations (Li et al., 2012).

The ANN is typically organized in layers, and these layers are made up of a number of interconnected nodes which contain an activation function (Ramadan et al., 2005). The network used in this study was a feed-forward network, consisting of an input layer (vis-NIR spectra), hidden layer, and output layer (SOC or pH). Extremely long training time and over-fitting are two major difficulties of ANN calibration when using raw infrared spectral data points as inputs (Mouazen et al., 2010). Data is pre-processed by scaling to (0-1) using a linear transformation. Statistica® offers Multilayer Perceptron (MLP) function network types. The network uses the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, which is a powerful second-order training algorithm with very fast convergence (Statsoft, 2011). Transfer functions used for hidden and output layers vary and they are hyperbolic tangent (Tanh), logistic sigmoid (Logistic), and negative exponential (Exp) functions (Quraishi & Mouazen, 2013). The neural network used for this study is a multilayer perceptron (MLP) neural network, with a sum of squares error function. The number of input, hidden, and output neurons were 504, 9, and 1 (e.g., SOC or pH), respectively. A total of 20 networks were trained and the five best performing networks were retained. Before each analysis, the entire spectra (270) were randomly split into 80 % (220 spectra) and 20 % (50 spectra) for the calibration sets and independent validation sets, respectively.

Assessment of model performance

For evaluation of model performance, the RMSEP (Williams & Norris, 2001) and RPD were used. The accuracy of cross-validation is estimated by the RMSEP (Viscarra-Rossel & Behrens, 2010). The RPD is the ratio of standard deviation of the measured values to the RMSEP. It is a good index for comparison of the different calibration models developed (Stenberg et al., 2004). Generally, in spectroscopic analysis, it is desirable to have $R^2 > 0.95$, and RPD > 5. For samples of complex material with variable composition, such as soil, this is an ambitious requirement. Viscarra-Rossel et al. (2006) classified the RPD values for soil

analysis as follows: RPD < 1.0 indicates very poor model/predictions and their use is not recommended; RPD between 1.0 and 1.4 indicates poor model/predictions, where only high and low values are distinguishable; RPD between 1.4 and 1.8 indicates fair model/predictions, which may be used for assessment and correlation; RPD values between 1.8 and 2.0 indicates good model/predictions, where quantitative predictions are possible; RPD between 2.0 and 2.5 indicates very good quantitative model/predictions; and RPD > 2.5 indicates excellent model/predictions. This classification system was adopted in this study.

RESULTS AND DISCUSSION

Prediction of soil organic carbon

Three samples were eliminated from the calibration sample set when developing the calibration model for SOC, whereas no outliers were removed from the independent validation set. Laboratory chemical analysis shows that the values for SOC for two of these samples were too low, compared to the other samples in the calibration set. Table 2 shows independent validation of the PLS and ANN calibration models for prediction of SOC. Results indicate that the ANN outperformed PLS in prediction of SOC at all MC. The best prediction result was obtained for 25 % MC soil samples with the ANN (RMSEP = 0.82 % and RPD = 4.23) (Table 2). An RPD value of 4.32 indicates excellent model prediction performance (Viscarra-Rossel et al., 2006). However, the best prediction result for PLS regression was obtained for dry soil samples (RMSEP = 1.17 % and RPD = 2.66) (Table 2), which was also classified as excellent model performance. However, it is clear that the ANN outperformed PLS, which is in line with the findings of Mouazen et al. (2010) and Viscarra-Rossel & Behrens (2010). The superior performance of the ANN can be attributed to the ability of the ANN to deal with the non-linear behavior of SOC known in NIR spectroscopy (Stenberg et al., 2010). The lowest predictive accuracy when the ANN was adopted was found for 15 % MC soil samples (RMSEP = 1.28 % and RPD = 2.73), whereas the lowest prediction result obtained with PLS regression was for a MC of 5 % (RMSEP = 1.58 % and RPD = 2.08) (Table 2). The worst prediction performance for the ANN is still classified as excellent model prediction performance. However, performance was classified as very good for PLS.

Figure 1 shows the scatter plots of measured versus predicted SOC at different MC levels obtained with PLS and the ANN. The predictive accuracy achieved in this study with PLS is evaluated as very good for models developed for 5, 15, 20, and 25 % MC samples, whereas the accuracy is evaluated as excellent for models developed with dry and wet samples of 10 % MC PLS. With the ANN, predictive accuracy is evaluated as excellent for models developed with dry and wet samples of all MC (Viscarra-Rossel et al., 2006). It was difficult to explain why PLS provided the best results for dry soil samples, while the ANN provided the best results for wet samples of 25 % MC. While the ANN showed a better performance for SOC prediction at all MC levels, PLS showed a better

Table 2. Partial least squares (PLS) and artificial neural network (ANN) independent validation results for prediction of SOC and pH

MC	RMSEP[1]		RPD[2]		$R^2_{calibration}$		$R^2_{validation}$	
	PLS	ANN	PLS	ANN	PLS	ANN	PLS	ANN
%	%							
				SOC				
0	1.17	0.91	2.66	3.86	0.80	0.90	0.87	0.93
5	1.58	1.04	2.08	3.36	0.78	0.82	0.80	0.92
10	1.40	1.13	2.56	3.09	0.79	0.87	0.85	0.91
15	1.35	1.28	2.40	2.73	0.82	0.86	0.84	0.86
20	1.55	1.17	2.28	2.99	0.82	0.82	0.81	0.90
25	1.59	0.82	2.42	4.23	0.80	0.90	0.83	0.95
				pH				
0	0.65	0.87	1.68	1.28	0.64	0.38	0.64	0.40
5	0.86	0.78	1.53	1.41	0.57	0.44	0.58	0.45
10	0.61	0.74	1.71	1.50	0.62	0.46	0.64	0.52
15	0.89	0.75	1.56	1.48	0.56	0.42	0.59	0.51
20	0.77	0.74	1.54	1.48	0.52	0.39	0.57	0.52
25	0.92	0.75	1.26	1.50	0.52	0.40	0.42	0.50

[1] RMSEP: root mean square error of prediction. [2] RPD: residual prediction to deviation.

predictive accuracy for pH at all MC levels, except for 25 % MC.

The histogram of normal distribution of error plots was calculated by subtracting predicted SOC from measured values using the 50 samples of the independent validation set scanned under laboratory (Figure 2) conditions. These plots show that errors are normally distributed around 0 values.

Prediction of pH

Table 2 shows the independent validation results of PLS and ANN calibration models for prediction of pH using the independent validation set. The best prediction result obtained with PLS was for dry (RMSEP = 0.65 % and RPD = 1.68) and 10 % MC soil samples (RMSEP = 0.61 % and RPD = 1.71) (Table 2). The prediction performance of these models is classified as fair. It is obvious that the predictive accuracy of pH is much lower than that of SOC, which is in line with the findings of other researchers (Stenberg et al., 2010). This confirms that the reason for better model performance for SOC compared to pH is associated with the presence of direct spectral responses of SOC in the NIR spectra. It is surprising to observe that for pH, PLS performed better than

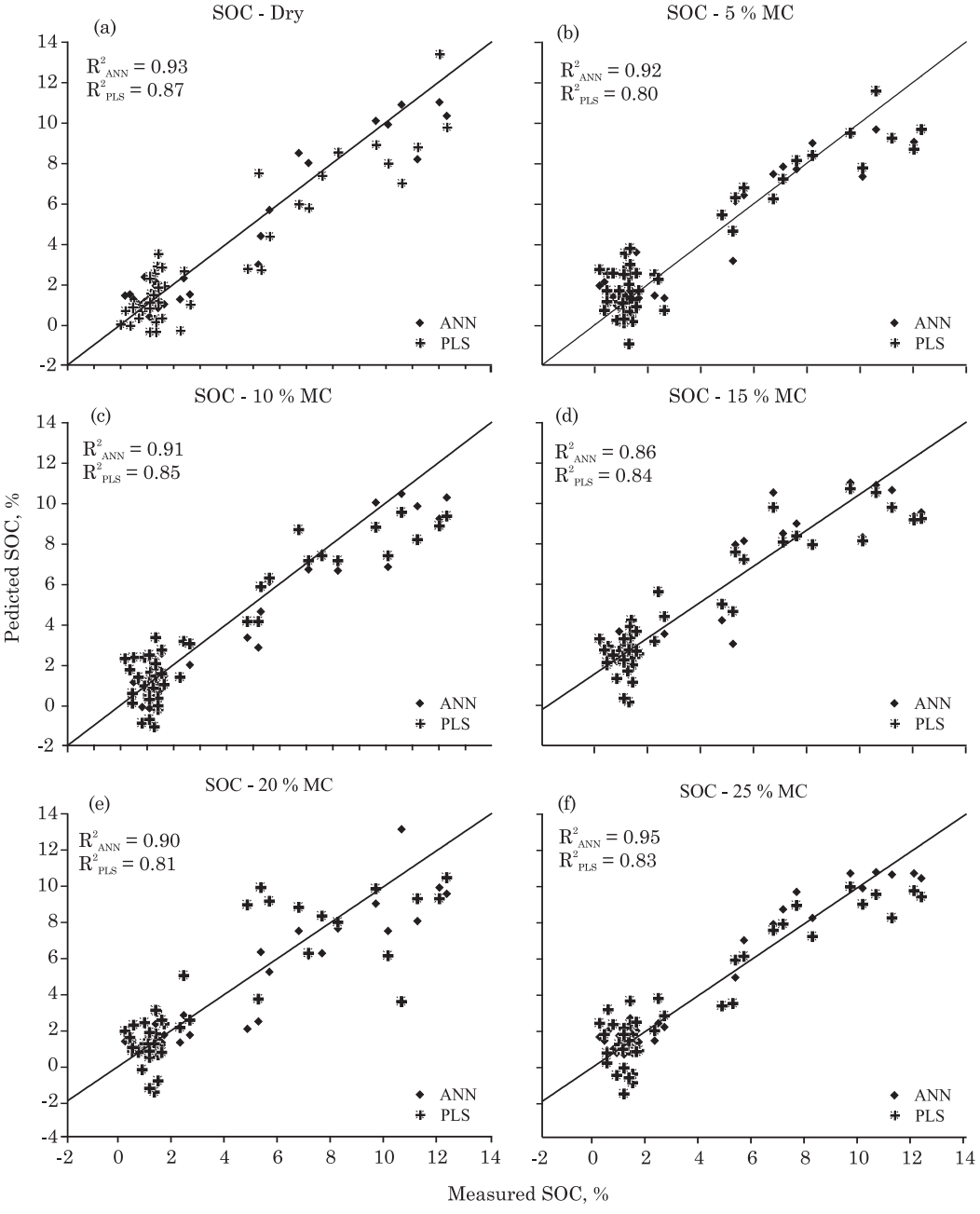

Figure 1. Effect of moisture content (MC) and calibration technique on R^2 of soil organic carbon (SOC) prediction with partial least squares (PLS) and the artificial neural network (ANN).

the ANN. This may be attributed to the nonlinear response of SOC when measured with vis-NIR spectroscopy (Stenberg et al., 2010). The best prediction result for the ANN was obtained for wet soil samples (RMSEP = 0.74-0.78 % and RPD = 1.41-1.50). The performance of these models was classified as fair (Viscarra-Rossel et al., 2006). Once more, the lowest accuracy for the ANN model is obtained for dry soil samples (RMSEP = 0.87 % and RPD = 1.21), which may be classified as poor model prediction.

It can be concluded that the performance of all pH models, both PLS and the ANN, can be classified as poor to fair model/predictions, since RPD values range from 1.26 to 1.71 for PLS and from 1.28 to 1.50 for the ANN. This is also shown by the poor quality of the scatter plots of measured pH versus predicted pH at different MC levels, as shown in figure 2.

The histogram of normal distribution of error plots was calculated by subtracting predicted pH from measured values using the 50 samples of the independent validation set scanned under laboratory conditions (Figure 3). As in the SOC histogram plots of error, the pH plots show that errors are normally distributed around 0 values.

With PLS, clear linear decreases in RPD values with MC were found for both SOC and pH (Figure 4). This is expected due to the linear nature of PLS. For the ANN, nonlinear polynomials are found to best fit the variation in both accuracy indices with MC (Figure 4).

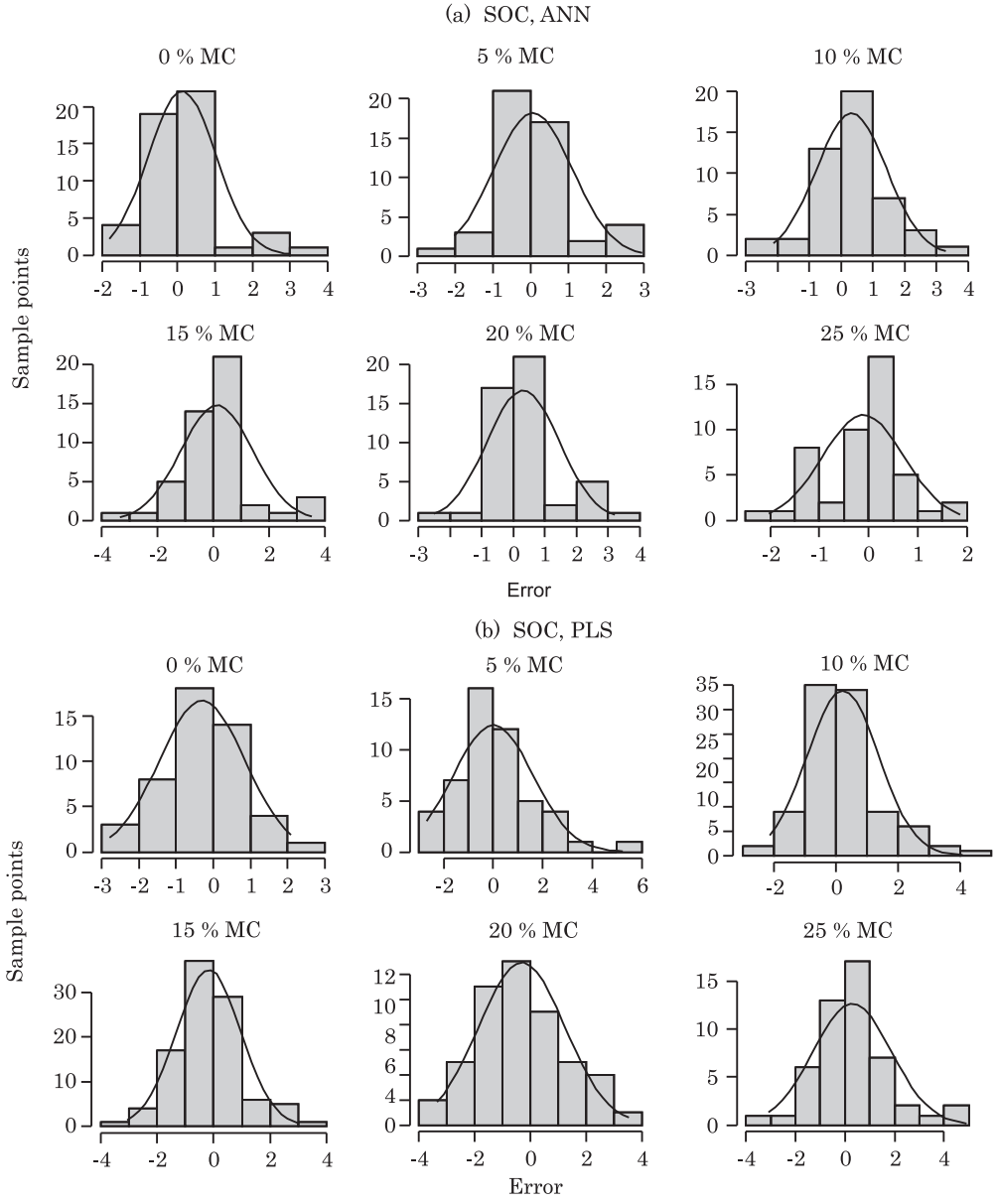

Figure 2. Histogram of normal distribution of error for the artificial neural network (ANN) predictions (a) and partial least squares (PLS) predictions (b) of soil organic carbon (SOC).

An opposite trend for variation in RMSEP with MC can be observed in figure 5, compared to variation in RPD with MC (Figure 6). Again second-order and third-order polynomials best fit the variation of RMSEP with MC for SOC and pH, respectively. From variation of RPD and RMSEP with MC of the data set used in this study, it may be concluded that for pH, use of PLS for dry soils, and the ANN for wet soils with a MC > 20 % is recommended. However, for SOC, the ANN is strongly recommended, as it clearly outperformed PLS. Further investigation is needed with different data sets to confirm this conclusion.

Second-order and third-order polynomials are fitted to the variation of RPD with MC for SOC and pH, respectively. With the ANN, the lowest RPD values for SOC and pH are obtained with 15 % MC soils and dry soils, respectively. With PLS, the highest RPD values for both properties are for dry samples.

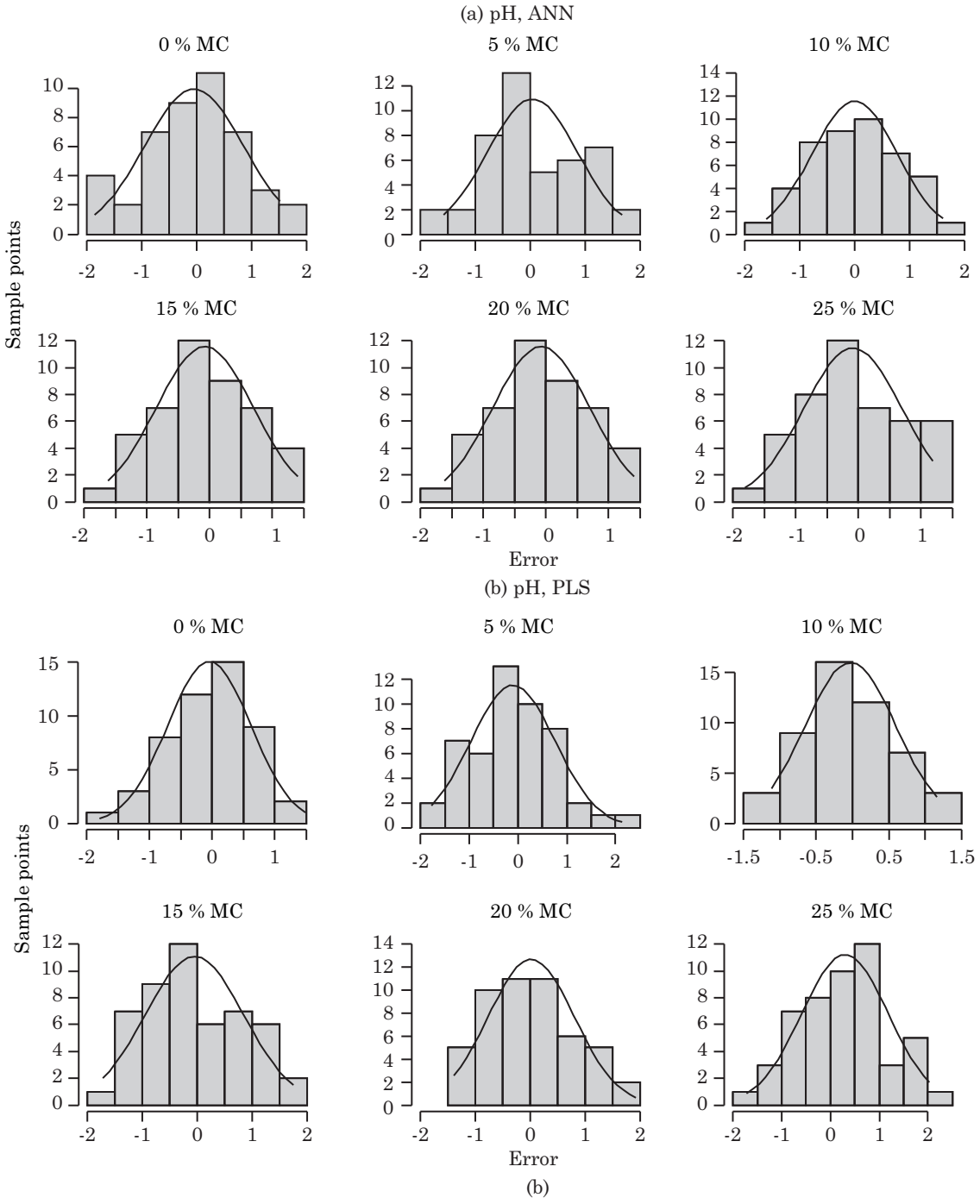

Figure 3. Histogram of normal distribution of error for the artificial neural network (ANN) predictions (a) and partial least squares (PLS) predictions (b) of pH.

CONCLUSIONS

1. Predictive accuracy of SOC and pH with vis-NIR spectroscopy depends on the modeling algorithm used, whether PLS or the ANN. It is not true to assume that the ANN always outperforms PLS.

2. The performance of these two techniques varies with the MC of soil samples and the property to be analyzed. The ANN outperformed PLS in the prediction of SOC for all the MC levels studied, whereas the ANN outperformed PLS only at high MC levels. The ANN models for SOC and pH resulted in non-linear variation of RPD and RMSEP with MC, whereas linear decrease in RPD and linear increase in RMSEP with MC were observed for PLS regression.

3. Therefore, based on the data set used in the current study, the ANN is recommended for analyses of SOC at all MC levels, whereas PLS is recommended for analysis of pH at all MC levels below 20 %. Further investigation with different data sets is needed to confirm this conclusion.

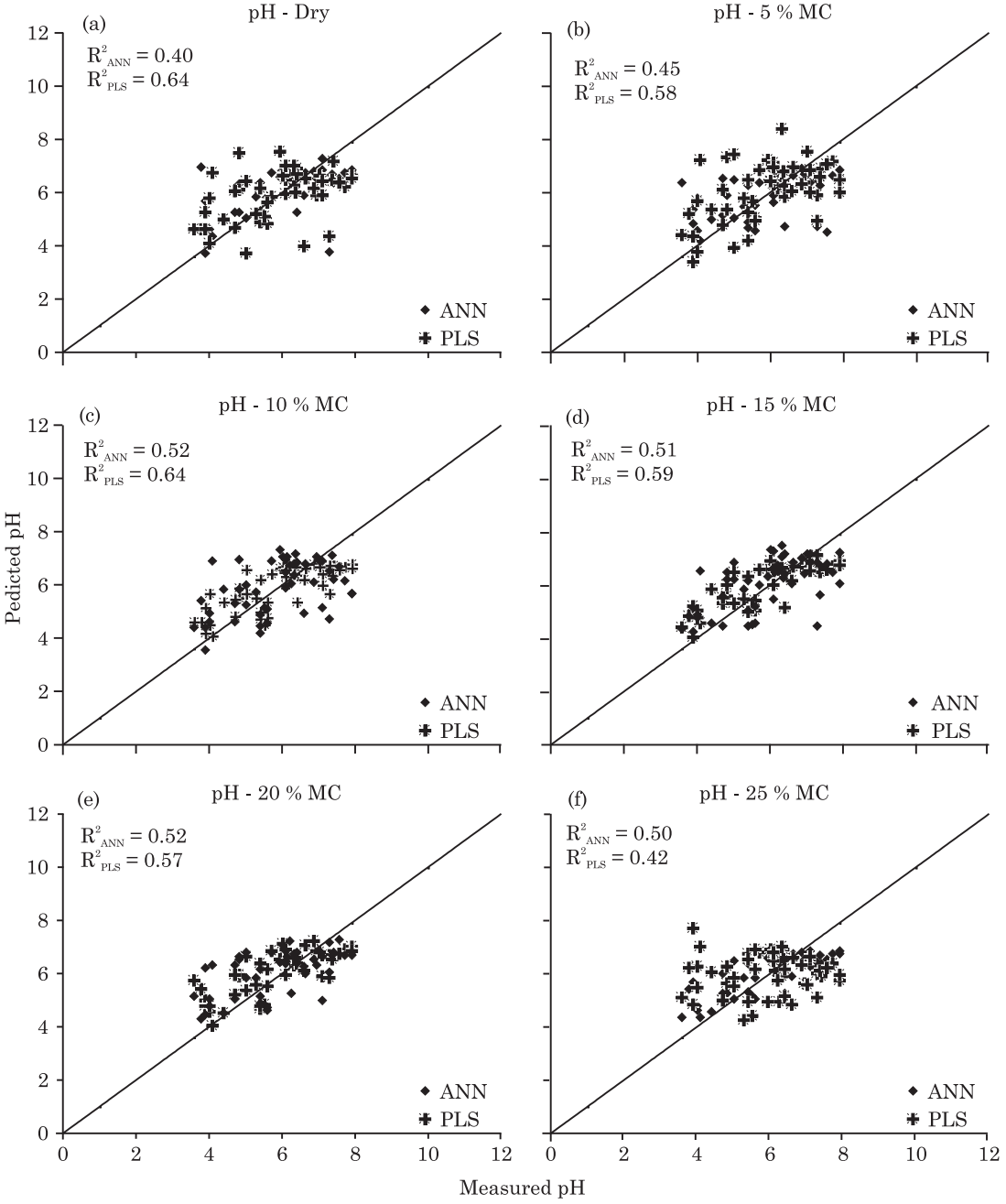

Figure 4. Effect of moisture content (MC) and calibration technique on R^2 of pH prediction with partial least squares (PLS) and the artificial neural network (ANN).

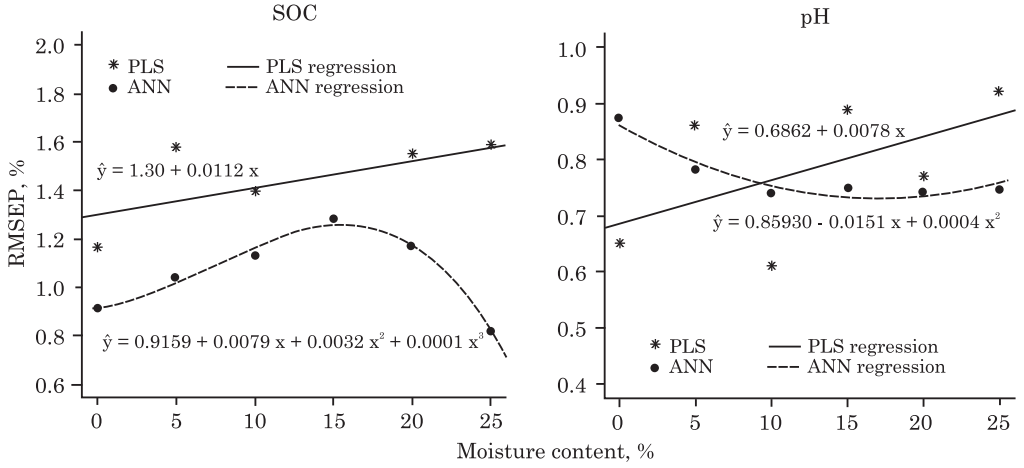

Figure 5. Root mean square error of prediction (RMSEP) for PLS and ANN models at different moisture content (MC) levels.

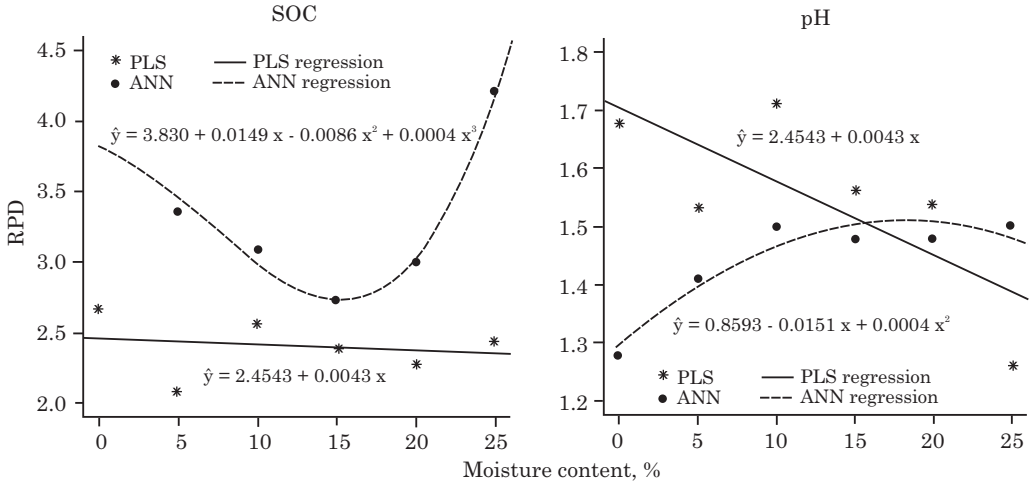

Figure 6. Variations in residual prediction deviation (RPD) with moisture content obtained from partial least squares (PLS) and artificial neural network (ANN) models.

LITERATURE CITED

BELLINASO, H.; ALEXANDRE, J.; DEMATTÊ J.A.M. & ROMEIRO, S.A. Soil spectral library and its use in soil classification. R. Bras. Ci. Solo, 34:861-870, 2010.

BOGREKCI, I. & LEE, W.S. Spectral measurement of common soil phosphate. Trans. ASAE., 48:2371-2378, 2005.

COHEN, M.; MYLAVARAPU, R.S.; BOGREKCI, I.; LEE W.S. & CLARK, M.W. Reflectance spectroscopy for routine agronomic soil analyses. Soil Sci., 172:469-485, 2007.

GHOLIZADEH, A.A.; SOOM, M.A.M.; SABERIOON, M.M. & BORÙVKA, L. Visible and near infrared reflectance spectroscopy to determine chemical properties of paddy soils. J. Food Agric. Environ., 11:859-866, 2013.

GOMEZ, C.; VISCARRA-ROSSEL R.A. & MCBRATNEY, A.B. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. Geoderma, 146:403-411, 2008.

JANIK, L.J.; FORRESTER, S.T. & RAWSON, A. The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. Chemometr. Intell. Lab., 97:179-188, 2009.

KODAIRA, M. & SHIBUSAWA, S. Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping. Geoderma, 199:64-79, 2013. (Special Issue)

LEONE, A.P.; VISCARRA-ROSSEL, R.A.; AMENTA, P. & BUONDONNO, A. Prediction of soil properties with PLSR and vis-NIR spectroscopy: Application to Mediterranean soils from Southern Italy. Curr. Anal. Chem., 8:283-299, 2012.

LI, D.; CHEN, X.; PENG, Z.; CHEN, S.; CHEN, W.; HAN, L. & LI, Y. Prediction of soil organic matter content in a litchi orchard of South China using spectral indices. Soil Till. Res., 123:78-86, 2012.

MALEKI, M.R.; van HOLM, L.; RAMON, H.; MERCKX, R.; DE BAERDEMAEKER, J. & MOUAZEN, A.M. Phosphorus sensing for fresh soils using visible and near infrared spectroscopy. Biosyst. Eng., 95:425-436. 2006.

MARIN-GONZALEZ, O.; KUANG, B.; QURAISHI, M.Z.; MUNOZ-GARCIA, M.A. & MOUAZEN, A.M. On-line measurement of soil properties without direct spectral response in near infrared spectral range. Soil Till. Res., 132:21-29, 2013.

McCARTY, G.W.; REEVES, J.B.; REEVES, V.B.; FOLLETT, R.F. & KIMBLE, J.M. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J., 66:640-646, 2002.

McLEAN, E.O. Soil pH and lime requirement. In: PAGE, A.L., ed. Methods of soil analysis. 2.ed. Madison, ASA, 1986. Part 2, p.199-224. (Agronomy Monograph, 9)

MORGAN, C.L.S.; WAISER, T.H.; BROWN, D.J. & HALLMARK, C.T. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. Geoderma, 151:249-256, 2009.

MOUAZEN, A.M.; DE BAERDEMAEKER, J. & RAMON, H. Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. Soil Till. Res., 80:171-183, 2005.

MOUAZEN, A.M.; KAROUI, R.; DE BAERDEMAEKER, J. & RAMON, H. Characterization of soil water content using measured visible and near infrared spectra. Soil Sci. Soc. Am. J., 70:1295-1302, 2006.

MOUAZEN, A.M.; MALEKI, M.R.; DE BAERDEMAEKER, J. & RAMON, H. On-line measurement of some selected soil properties using a VIS-NIR sensor. Soil Till. Res., 93:13-27, 2007.

MOUAZEN, A.M.; KUANG, B.; DE BAERDEMAEKER, J. & RAMON. H. Comparison between principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. Geoderma, 158:23-31, 2010.

NELSON, D.W. & SOMMERS, L.E. Total carbon, organic carbon, and organic matter. In: PAGE, A.L., ed. Methods of soil analysis. 2.ed. Madison, ASA, 1986. Part 2, p.539-579.

QURAISHI, M.Z. & MOUAZEN, A.M. Development of a methodology for in situ assessment of topsoil dry bulk density. Soil Till. Res., 126:229-237, 2013.

RAMADAN, Z.; HOPKE, P.K.; JOHNSON, M.J. & SCOW, K.M. Application of PLS and Back-Propagation Neural Networks for the estimation of soil properties. Chemometr. Intell. Lab., 75:23-30, 2005.

SAVITZKY, A. & GOLAY, M.J.E. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem., 36:1627-1639, 1964.

SORIANO-DISLA, J.M.; JANIK, L.J.; VISCARRA-ROSSEL, R.A.; MacDONALD, L.M. & McLAUGHLIN, M.J. The Performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl. Spectrosc. Rev., 49:139-186, 2014.

STATSOFT, I. Statistica (data analysis software system), 2011.

STENBERG, B.; JENSEN, L.S.; NORDKVIST, E.; BRELAND, T.A.; PEDERSON, A.; GUOMUNDSSON, J.; BRUUN, S.; SALO, T.; PALMASON, F.; HENRIKSEN, T.M. & KORSAETH, A. Near infrared reflectance spectroscopy for quantification of crop residue, green manure and catch crop C and N fractions governing decomposition dynamics in soil. J. Near Infrared Spec., 12:331-346, 2004.

STENBERG, B.; VISCARRA-ROSSEL, R.A.; MOUAZEN, A.M. & WETTERLIND, J. Visible and near infrared spectroscopy in soil science. Adv. Agron., 107:163-215, 2010.

SUMMERS, D.; LEWIS, M.; OSTENDORF, B. & CHITTLEBOROUGH, D. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. Ecol. Indic., 11:123-131, 2011.

TEKIN, Y.; TUMSAVA[a], Z. & MOUAZEN, A.M. Effect of moisture content on prediction of organic carbon and pH using visible and near-infrared spectroscopy. Soil Sci. Soc. Am. J., 76:188-198, 2012.

TEKIN, Y.; KUANG, B. & MOUAZEN, A.M. Potential of on-line visible and near infrared spectroscopy for measurement of pH for deriving variable rate lime recommendations. Sensors, 13:10177-10190, 2013.

VASQUEZ, G.M.; GRUNWALD, S. & SICKMAN, J.O. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma, 146:14-25, 2008.

VISCARRA-ROSSEL, R.A.; WALVOORT, D.J.J.; McBRATNEY, A.B.; JANIK, L.J. & SKJEMSTAD, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, 131:59-75, 2006.

VISCARRA-ROSSEL, R.A. & BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma, 158:46-54, 2010.

VISCARRA-ROSSEL, R.A. & CHEN, C. Digitally mapping the information content of visible-near infrared spectra of surficial Australian soils. Rem. Sens. Environ., 115:1443-1455, 2011.

WILLIAMS, P.C. & NORRIS, K. Variable affecting near-infrared spectroscopic analysis. In: WILLIAMS, P.C. & NORRIS, K., eds. Near-infrared technology in agricultural and food industries. 2.ed. St. Paul, American Association of Cereal Chemists, 2001. p.171-189.

YANG, H. & MOUAZEN, A.M. Vis/near and mid-infrared spectroscopy for predicting soil N and C at a farm scale. In: THEOPHANIDES, T., ed. Infrared spectroscopy-life and biomedical sciences. Rijeka, Croatia, Intech Press, 2012. p.185-210.

ZHAO, S.J.; ZHANG, J.; XU, Y.M. & XIONG, Z.H. Non-linear projections to latent structures method and its applications. Indian Eng. Chem. Res., 45:3843-3852, 2006.