

# Determination and Prediction of Some Soil Properties using Partial Least Square (PLS) Calibration and Mid-Infra Red (MIR) Spectroscopy Analysis

Irnanda Aiko Fifi Djuuna<sup>1</sup>, Lynette Abbott<sup>2</sup> and Craig Russell<sup>2</sup>

<sup>1</sup>*Department of Soil Sciences, Faculty of Agriculture and Agriculture Technology, The State University of Papua, Gunung Salju St. Amban Manokwari 98314, West Papua, Indonesia. Tel. +62-0986- 211830; fax: +62-0986-213513 E-mail address: irnanda\_d@yahoo.com.au*

<sup>2</sup>*School of Earth and Geographical Science, The University of Western Australia, Crawley 6009*

Received 18 June 2010 / accepted 25 April 2011

## ABSTRACT

Soil chemical, physical and biological analyses are a crucial but often expensive and time-consuming step in the characterization of soils. Rapid and accurate predictions and relatively simple methods are ideally needed for soil analysis. The objective of this study was to predict some soil properties (*e.g.* pH, EC, total C, total N, C/N, NH<sub>4</sub>-N, NO<sub>3</sub>-N, P, K, clay, silt, and sand and soil microbial biomass carbon) across the Wickepin farm during summer season using a Mid-Infra Red - Partial Least Square (MIR-PLS) method. The 291 soil samples were analyzed both with soil extraction procedure and MIR Spectrometer. Calibrations were developed between MIR spectral data and the results of soil extraction procedures. Results using the PLS-MIR showed that MIR-predicted values were almost as highly correlated to the measured value obtained by the soil extraction method of total carbon, total nitrogen and soil pH. Values for EC, NH<sub>4</sub>-N, NO<sub>3</sub>-N, C/N, P, K, clay, silt, sand, and soil microbial biomass carbon were not successfully predicted by the MIR – PLS technique. There was a tendency for these factors to correlate with the MIR predicted value, but the correlation values were very low. This study has confirmed that the MIR-PLS method can be used to predict some soil properties based on calibrations of MIR values.

**Keywords:** MIR-Partial Least Square, MIR-Spectroscopy, soil properties

## INTRODUCTION

Soils are rarely homogeneous and the variability occurs both laterally and with depth. It can result from changes in the chemical balance of the soil associated with agricultural practices including nutrient uptake, crop rotations, fertilizer use, lime application (Viscarra Rossel and McBratney 1998) and leaching. To understand more about the variation in soil properties, spatially dense soil analyses are often required. Soil analytical procedures also need to be able to cope with a large number of samples.

An alternative method for assessing soil properties and their variability across a landscape is mid infrared (MIR) spectroscopy. MIR spectroscopy has considerable advantages and offers a possible alternative to conventional methods through increased speed and sensitivity (Janik *et al.* 1995). The fundamental molecular

frequencies of this technique are generally in the MIR 2,500-25,000 nm wavelengths or 4,000-250 cm<sup>-1</sup> range (Janik *et al.* 1998). This method has been used to determine some macro and micro elements of soil including soil carbon (McCarty *et al.* 2002), organic matter composition (Cheshire *et al.* 1993; Spaccini *et al.* 2001) and many other soil properties (Janik *et al.* 1998) as well as soil minerals (Janik *et al.* 1995; Janik and Skjemstad 1995). However, none of the above studies were conducted under agricultural systems which had different land use histories or at different occasions.

Mid infrared-partial least square (MIR-PLS) has not been widely used for soil property prediction in agricultural systems. However, some studies have been conducted for (i) prediction of organic carbon, nitrogen and other properties of peat soil (Holmgren and Norden 1988), (ii) the contamination of hydrocarbon in wet soil (Hazel *et al.* 1997), (iii) a range of soil properties including mineralogical analysis, organic carbon, nitrogen, carbonate, air-dry moisture and cation exchange capacity (Janik and Skjemstad 1995; Janik *et al.* 1995; Murphy and Milton 2004) and (iv) prediction of total carbon of

soil, clay content and soil N supply (Murphy and Milton 2004).

In relation to soil biological characteristics, the MIR spectral contribution of in-situ soil microorganisms is expected to be minor because only about 5% of the total organic carbon in soil is microbial carbon (Janik *et al.* 1998). The reason for this is that the signatures of microorganisms are usually difficult to detect *in situ* in most soils due to masking of soil mineral and other organic peaks. However, a study by Pankhurst *et al.* (1997) showed that there was a correlation between GC-FAME spectra and MIR spectra of the same soils using PLS regression analysis. The correlation values ranged between 0.5 to 0.75, showing that the MIR spectra could explain 75% of the variability in the FAME data. This study suggested that MIR spectra can be obtained much more rapidly compared with the conventional methods for measuring bacteria and fungi. Therefore, there is an opportunity to use this predictive capacity of MIR-PLS models as a simple bio-monitor of soil health and for soil biological mapping.

Another study (Grube *et al.* 1999) showed that the principal component concentration changes in microbial cells can be studied by quantitative estimation of the biomass MIR spectra. In addition, MIR absorption spectra had also been applied to determine bacterial biomass (*Zymomonas mobilis*) which was grown in sucrose or glucose medium (Grube *et al.* 2002).

The objective of this study was to predict some soil properties across the farm during summer season using MIR –PLS method.

## MATERIALS AND METHODS

### Location and Soil Sampling

Two hundred ninety one (291) soil samples were collected from a farm (Fairlawn) in the Wickepin Shire in south-western Australia, about 300 km south east of Perth, 32° 47' 4" S, 117° 30' 9" E during summer season in 2004. The farm covers an area of 647 ha and was divided into 14 paddocks. The region had a Mediterranean climate with hot, dry summers and cold winters, and elevations ranged from about 250 to 350m. The area was dominated by flats with some rolling hills. The major soil textures in this area were medium loam, sandy loam, clayey loam and clay. The major land use was principally sheep farming and wheat production, with annual pasture dominated by grass species.

### Soil Analysis

The soil samples were oven dried, ground and sieved to a size fraction smaller than 2 mm. The number of samples collected and the soil analyses performed are summarized in Table 1.

### MIR Analysis

Two hundred ninety one (291) air dried soil samples were ground in an agate mill at 750 rpm to

Table 1. Laboratory methods of soil analyses.

Soil property	n	Technique	References
pH H <sub>2</sub> O	291	1:5 soil: H <sub>2</sub> O extract	White (1969)
Sand, silt, clay (%)	291	Pipette method	Rayment and Higginson (1992)
Nitrate–nitrogen (NO <sub>3</sub> -N) and ammonium–nitrogen (NH <sub>4</sub> -N) (mg kg <sup>-1</sup> )	291	Skalar method from Skalar SAN <sup>plus</sup> Segmented Flow Autoanalyzer	Rayment and Higginson (1992)
Available phosphorus Colwell (PCol) (mg kg <sup>-1</sup> )	291	Bicarbonate extractable P (Spectrophotometer)	Olsen et al (1954), Rayment and Higginson (1992)
Exchangeable potassium (K) (mg kg <sup>-1</sup> )	291	Bicarbonate extractable K (AAS)	Rayment and Higginson (1992)
Electrical conductivity (EC) (dS m <sup>-1</sup> )	291	1:5 soil :H <sub>2</sub> O extract	Rayment and Higginson (1992)
Microbial biomass C (mg kg <sup>-1</sup> )	291	Fumigation-Extraction (FE) method	Vance <i>et al.</i> (1987) and Sparling (1990).
Total C and total N (%)	291	CHN-1000 Elemental Analyzer version 1.1 (1991) LECO	Rayment and Higginson (1992)

pass through a 0.2 mm mesh. A small amount of the samples was poured into 10 mm diameter stainless steel cups and the top of the powders were leveled without applying pressure. Samples were scanned using MIR from 7,000 to 500  $\text{cm}^{-1}$  at 2  $\text{cm}^{-1}$  resolution on a Spectrum One Fourier-Transform mid Infrared (FT-IR) Spectrometer Vers. B Model L120000B. Each sample was scanned in duplicate and the spectra was averaged. The infrared spectra were recorded by the diffuse reflectance (DRIFT) sampling accessory. Peak areas from spectra were used for statistical analysis.

### Data Analysis

The data files containing the spectral data points were imported as GRAMS format for the PLS calculation. The MIR calibrations were performed by Partial Least Squares (PLS) regression using the computer program for multivariate modeling 'Unscrambler' Version 7.5 (Camo A/S, Trondheim, Norway). The PLS was used to determine the best correlation between chemical reference data and spectral data to develop a calibration. The PLS regression model was built using a calibration set with 2/3 of the spectra of 213 and 291 for the overall data randomly selected observations. Cross validation using partial least regression was performed for each soil variable.

The predictive model was performed using validation of 1/3 of the unused data set of independent 78 and 70 of the whole observations. The calibration and predicted model were evaluated from their correlation coefficients ( $r$ ),  $r^2$ , the root

mean standard error of prediction (RMSEP), and the standard error of prediction (SEP). The values in the prediction set were recorded and this procedure was repeated two times with different samples in the validation set until all samples were predicted.

The best calibration and prediction model were then judged against the one with the lowest standard error of prediction and the highest correlation coefficients. Two other statistical analysis were used to evaluate the calibration: (i) RPD, the ratio of the standard deviation (SD) of the measured or references value in the prediction set to root mean square of error prediction (RMSEP), and (ii) RER, the ratio of the range of measured/references value in the prediction set to the RMSEP (Malley *et al.* 1999).

## RESULTS AND DISCUSSION

### DRIFT Spectra for Whole Soil

The DRIFT spectra for a typical soil from the Wickepin farm are shown in Fig. 1. The high peak around 900-500  $\text{cm}^{-1}$  reflected other clay mineral characteristics (Janik and Skjemstad 1995; Nguyen *et al.* 1991). This showed that mid infrared spectra contain definable peaks which could be used in spectral interpretation and differentiate between two or more samples.

### Soil Characteristics

The summary of soil characteristics across the farm is presented in Table 2.



Figure 1. Mid-infrared spectra of all soil samples from the farm at Wickepin in southwestern Australia.

Table 2. Summary statistics of 291 soil samples for 13 soil characteristics at a farm scale at the Wickepin farm (soils sampled in summer).

Soil Variables	Mean	Standard Deviation	Min	Max
NH <sub>4</sub> -N (mg kg <sup>-1</sup> )	2.11	1.03	0.24	7.19
NO <sub>3</sub> -N (mg kg <sup>-1</sup> )	6.54	4.08	0.00	25.87
pH H <sub>2</sub> O	5.80	0.33	5.21	7.51
EC (μS cm <sup>-1</sup> )	148.22	319.99	20.40	5021.00
Total C (%)	1.48	0.53	0.41	5.27
Total N (%)	0.11	0.04	0.00	0.42
C/N	13.79	7.62	8.50	136.33
P (mg kg <sup>-1</sup> )	72.39	69.78	11.04	569.01
K (mg kg <sup>-1</sup> )	45.12	17.25	8.20	118.00
Microbial Biomass-C (mg kg <sup>-1</sup> )	443.47	272.58	8.88	1672.51
Clay (%)	5.707	2.811	0.00	18.37
Silt (%)	7.272	9.004	0.00	55.01
Sand (%)	87.038	9.884	39.57	99.50

### Calibration of Results

The results obtained from the cross validation procedure are presented in Table 3 and confirmed the feasibility of predicting some soil variables. Total C, total N, EC and soil pH were successfully predicted compared to other soil properties for which the correlation values were low. The best correlation for total C, total N and soil pH were

also followed by acceptable RPD and RER values (Table 3); RPD>3 and RER>10 are considered acceptable. This indicated that the infrared spectra could explain 76% of EC and soil pH, 86% of total nitrogen and 90% of total carbon of the variability in the laboratory extraction analysis data. Other soil properties were not successfully predicted using this technique because the correlation values were lower than 50%.

Table 3. Independent validation set of 71 soil samples with the 213 samples used to develop a calibration of 2002 soil samples from Wickepin farm (MBC=microbial biomass carbon; RMSEP= root mean square of error prediction; RPD= ratio of standard deviation to standard error of prediction (SD/SEP); RER= ratio of range to SEP).

Soil variables	R <sup>2</sup>	RMSEP	RPD	RER
pH	0.76	0.22	0.98	5.37
EC	0.76	3.05	1.09	13.60
Total C	0.90	0.23	2.00	15.00
Total N	0.86	0.02	1.96	16.66
C/N	0.01	8.81	0.12	0.73
N-NH <sub>4</sub>	0.43	0.93	0.60	2.89
N-NO <sub>3</sub>	0.42	3.65	0.55	2.11
P	0.48	64.04	0.50	2.01
K	0.03	17.33	0.14	0.61
Clay	0.50	6.33	0.54	2.83
Silt	0.26	39.12	0.18	0.95
Sand	0.01	15.99	0.18	0.95
Microbial Biomass-C	0.26	251.28	0.58	2.65

Among the soil variables measured in this study, only total carbon, total nitrogen and soil pH were best predicted with a PLS-MIR calibration. Total soil carbon is the most common soil property used in NIR and MIR studies and was reported to have very high correlations with conventional soil analytical data compared to other soil properties (e.g. McCarty et al. 2002). However, the study by Brown *et al.* (2005) to predict soil C using the same technique failed at some sites due to “pseudo-independent” validation (random selection of non-independent test samples) that can overestimate predictive accuracy relative to independent validation.

A major advantage of diffuse reflectance spectroscopy for soil analysis is that from a single spectrum many properties may be (accurately) determined, thus offering the possibility for considerable cost savings and increased efficiency over conventional laboratory analysis. Furthermore, the technique is rapid, making it possible to analyze a large number of samples in a practical and timely manner. These properties make spectroscopic analyses combined with PLSR very attractive for environmental monitoring, modeling and precision agriculture (Viscara Rossel *et al.* 2006).

Other studies have also shown that not all conventional soil extractions can be predicted using the MIR-PLS technique (Janik *et al.* 1998). They showed that soil properties such as EC, K-available, P-available, P buffer, DTPA-extractable Zn and Cu, CaCl<sub>2</sub>, exchangeable Na and K had very low correlations ( $r^2 = <0.50$ ) between conventional and infrared assays. Reeves *et al.* (2001) found that soil pH ( $r^2 = 0.94$ ) was significantly better than other soil properties such as total carbon and nitrogen, and some biological properties of soil. Other soil properties measured in this study were less well correlated using the MIR-PLS technique. Although the prediction was less successful for these variables, MIR may be useful for screening purposes.

## CONCLUSIONS

The calibration values can be developed using mid infrared spectra and conventional assays for several soil parameters, particularly total carbon, total nitrogen and soil pH. This study also provides evidence that total carbon, total nitrogen and soil pH can be simultaneously measured by rapid and non-destructive mid-infrared spectroscopy. The calibrations developed in this study have not been validated beyond this location. There is an indication that the calibrations might not be widely

applicable because only the calibration for soil pH data corresponded well to the other data, but this needs to be further investigated.

## REFERENCES

- Brown DJ, RS Bricklemeyer and PR Miller. 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma* 129 (3-4): 251-267.
- Cheshire MV, JD Russell and AR Fraser. 1993. The decomposition of straw in soil: use of multiple internal reflectance infrared spectroscopy. *J Soil Sci* 44: 693-702.
- Grube M, J Zagreba, E Gromozova and M Fomina. 1999. Comparative investigation of the macromolecular composition of mycelia forms *Thielavia terrestris* by infrared spectroscopy. *Vib Spectrosc* 19: 301-306.
- Grube M, M Bekers, D Upite, and E Kaminska. 2002. IR-spectroscopic studies of *Zymomonas mobilis* and levan precipitate. *Vib Spectrosc* 28: 277-285.
- Hazel G, F Bucholtz, GN Anngarwal and KJ Ewing. 1997. Multivariate analysis of Mid-IR spectra of hydrocarbon-contaminated wet soils. *Appl Spectrosc* 51: 984-989.
- Holmgren A and B Norden. 1988. Characterization of peat samples by diffuse reflectance FT-IR spectroscopy. *Appl Spectrosc* 42: 225-262.
- Janik LJ and JO Skjemstad. 1995. Characterization and analysis of soils using Mid infrared Partial Least-Squares 2. Correlations with Some Laboratory Data. *Aust J Soil Res* 33: 637-650.
- Janik LJ, JO Skjemstad and MD Raven. 1995. Characterization and analysis of soils using mid-infrared partial least squares. I. Correlations with XRF-determined major element composition. *Aust J Soil Res* 33: 621-636.
- Janik LJ, RH Merry and JO Skjemstad. 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? *Aust J Exp Agr* 38: 681-696.
- Malley DF, L Yesmin, D Wray and S Edwards. 1999. Application of near-infrared spectroscopy in analysis of soil mineral nutrients. *Comm Soil Sci Plant Anal* 30: 999-1012.
- McCarty GW, JB Reeves, VB Reeves, RF Follett and JM Kimble. 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci Soc Am J* 66: 640-646.
- Murphy DV and N Milton. 2004. Mapping biological soil nitrogen supply using mid infrared technology. In: Lines-Kelly (ed) *Soil Biology in Agriculture*. Proceeding of a workshop on current research into soil biology in agriculture. Tamworth Sustainable Farming Training Centre 11-12 August 2004. NSW Department of Primary Industries, Orange 2800, p.136
- Nguyen TT, LJ Janik and M Raupah. 1991. Diffuse reflectance infrared fourier transform (DRIFT) spectroscopy in soil studies. *Aust J Soil Res* 29: 49-67.

- Olsen SR, CV Cole, FS Watanabe and LA Dean. 1954. Estimation of available phosphorus in soils by the extraction with sodium bicarbonate; Circ. 939; U.S. Dep. of Agric.
- Pankhurst CE, LJ Janik, CA Kirkby and BG Hawke. 1997. Application of GC-FAME and mid-infrared analysis of soil as measures of soil health. RIRDC Project CSO-10A, Final Report.
- Rayment GE and FR Higginson. 1992. Australian Laboratory Handbook of Soil and Water Chemical Methods. (Inkata Press: Melbourne). Australia, 330 pp.
- Reeves JB and GW McCarty. 2001. Quantitative analysis of agricultural soil using near infrared reflectance spectroscopy and a fiber-optic probe. *J Near Infrared Spectrosc* 9 (1): 25-34.
- Spaccini R, A Piccolo, G Haberhauer, M Stemmer and MH Gezarbek. 2001. Decomposition of maize straw in three European soils as revealed by DRIFT spectra of soil particle fractions. *Geoderma* 99: 245-260.
- Sparling GP. 1990. Soil biomass evaluation. In: PACIFICLAND Workshop the Establishment of Soil Management Experiments on Sloping Lands. IBSRAM Technical Notes No. 4 International Board for Soil Research and Management, Bangkok, pp.163-184.
- Vance ED, PC Brookes, DS Jenkinson. 1987. An extraction method for measuring soil microbial biomass C. *Soil Biol Biochem* 19: 703-707.
- Viscarra Rossel RA and AB McBratney. 1998. Soil chemical analytical accuracy and costs: implication from precision agriculture. *Aust J Exp Agr* 38: 765-775.
- Viscarra Rossel RA, DJJ Walvoort, AB McBratney, LJ Janik and JO Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131: 59-75.
- White, RE. 1969. Measurement of soil pH. *J Aust Inst Agric Sci* 35: 3-15.