PROXIMAL SENSING OF SOIL pH AND LIME REQUIREMENT BY MID INFRARED DIFFUSE REFLECTANCE SPECTROSCOPY

R.A. VISCARRA ROSSEL[1,2,*], D.J.J. WALVOORT[1,3], A.B. McBRATNEY[1], L.J. JANIK[4], J.O. SKJEMSTAD[4]

[1]*Australian Centre for Precision Agriculture, The University of Sydney, NSW 2006, Australia*
[2]*INRA Laboratoire de Science du Sol, 35042 Rennes, France*
[3]*Wageningen University & Research Centre, Wageningen, The Netherlands*
[4]*CSIRO Land and Water, Glen Osmond SA 5064, Australia*
[*]*Corresponding author -mail: r.rossel@agec.usyd.edu.au*

ABSTRACT

We describe the use of mid-infrared (MIR) diffuse reflectance spectroscopy in combination with partial least squares regression (PLS) for the prediction of soil pH and lime requirement. Delete-one-jackknife cross-validation was used to determine the optimal number of factors for the most accurate prediction of each soil property. Predictions of soil pH in $0.01M$ $CaCl_2$ (pH_{CaCl_2}), H_2O (pH_{H_2O}) and lime-requirement buffer (pH_{buffer}) were compared to those derived from laboratory extractions. Comparisons of lime requirement were made using pH_{buffer} measurements and a response surface regression. The mid-infrared with partial least squares predictions were less accurate. The nugget variance of lime requirements derived using both the conventional and the MIR showed that the latter may be a more precise analytical technique. Kriging of the MIR data produced a map suitable for variable-rate liming.

INTRODUCTION

Increasingly, spectroscopic methods are being considered as possible alternatives to enhance or replace conventional laboratory methods of soil analysis (e.g. Janik et al.,1995; Viscarra Rossel & McBratney, 1998a). The reasons for this may be that the techniques are rapid, timely, less expensive than conventional soil analysis (after initial set-up costs), and are occasionally more accurate than conventional laboratory methods (e.g. McCauley et al., 1993). The development of a proximal spectrometric soil sensing system would be an useful step towards real-time, continuous site-specific management, which is the ultimate aim of precision agriculture (PA). The aims of this work are: (i.) to determine the utility of MIR soil analysis with partial least-squares regression (MIR-PLS) for the prediction of soil pH and lime requirement, and (ii.) to assess its adaptability for use in precision agriculture.The MIR portion of the electromagnetic spectrum occurs in the 2500 to 25000 nm wavelength range, corresponding to a wave number range from 4000 to 400 cm^{-1}. Mid-infrared spectroscopy has a particular advantage over other spectroscopic techniques for soil analysis (e.g. near-infrared) as it is sensitive to the intense fundamental vibrations of both the organic and inorganic phases of the soil. A review of the technique may be found in Janik et al. (1998).

METHODS

Field Sampling and Analyses

Surface soil samples were collected from a 17.5 ha field at Kelso NSW, Australia on a 40 m grid at each of 122 grid nodes. Viscarra Rossel & McBratney (2001) described the sampling strategy and laboratory procedures for measurements of soil pH_{CaCl_2} and pH_{H_2O}. The Mehlich lime-requirement buffer was used for soil:lime requirement measurements (pH_{buffer}). The lime

required to attain a target pH of 6.5 was predicted using a response surface model derived by Viscarra Rossel & McBratney (2001). The model has the following form:

$$LR \ (Mg/ha) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 xy + \beta_5 y^2 + \varepsilon$$

where β_h are parameters of the model; x is the equilibrated pH_{buffer}, y is the selected target pH post-liming, and $\varepsilon \sim N(0, \sigma^2)$ is a normally distributed error term with zero mean and variance σ^2.

Mid Infrared Analyses

Portions of the soil samples were ground and sieved to a size fraction smaller than 80 μm for analyses as neat powders using the diffuse reflectance infrared method described by Janik et al. (1995). Spectra were recorded from 4000 to 500 cm^{-1} at 3.9 cm^{-1} intervals. The equipment consisted of a rapid scanning Fourier Transform spectrometer, with an extended range KBr beam splitter and Peltier-cooled DTGS detector with a spectral range of 8300 - 470 cm^{-1} and a resolution of 8 cm^{-1}.

Partial Least Squares and Cross Validation

Given a soil sample, we want to use its MIR spectrum to predict soil property y, in our case soil pH or lime requirement. Therefore we need a calibration model. Suppose we have taken n soil samples $i=1...n$. Each soil sample i is analysed for property y by means of a conventional laboratory method. In addition, MIR spectra $x_i = [x_1, x_2, ..., x_k]'$ consisting of absorbances x are collected at k wavenumbers. Let $y = [y_1, y_2, ..., y_n]'$ be a vector of n measurements of soil property y, and let $X = [x_1, x_2, ..., x_n]'$ be a matrix of corresponding MIR spectra. A calibration model can now be constructed by means of partial least squares regression (PLS). PLS decomposes X and y into factor scores (T) and factor loadings (P and q) according to:

$$X = TP'+E$$
$$y = Tq+f$$

X and y are usually zero-centred prior to decomposition. The decomposition is performed in such a way that the first p ($< min(n,k)$) factors explain most of the variation in X and y. The remaining factors resemble noise and can thus be ignored, hence the addition of residuals E and f. The resulting matrices and vectors generally have a much lower dimension than X and y. Given a new spectrum x, soil property y can be estimated as a linear combination of the factors scores and factor loadings of x. For more details on PLS see e.g. Martens and Næs (1989). Delete-one-jackknifing (Efron and Tibshirani, 1994) was used to estimate the number of factors to retain in our model. The Root Mean Squared prediction Error (RMSE) was used to summarise the results. Generally, the model with the lowest RMSE is selected.

Spatial Analysis

Semi-variograms fitted with exponential models were used to quantify the spatial dependence of lime requirements derived from the response surface model and those predicted by the MIR-PLS technique. The semi-variograms were used to block krige the data using 10 m blocks and 2.5 m spacing. A minimum of 50 points were used for each estimation. Kriged estimates were compared, and a map of MIR-PLS predicted lime requirements contoured.

RESULTS

Soil Data Descriptions

Soil pH_{CaCl_2}, pH_{H_2O} and lime requirements were negatively skewed. The soil pH_{buffer} data were normally distributed but for a few low outliers. Summary statistics of the data from conventional laboratory analyses are given on the left side of Table 1.

TABLE 1. Statistical description of 'known' soil data analysed using conventional laboratory methods of analyses, and that analysed using the mid infrared technique. n = 118

	Conventional laboratory analyses		MIR-PLS predictions	
Property	Mean	St. Dev.	Mean	St. Dev.
pH$_{CaCl_2}$	4.37	0.26	4.37	0.25
pH$_{H_2O}$	5.3	0.27	5.29	0.25
pH$_{buffer}$	5.85	0.1	5.85	0.1
LR(6.5) Mg/ha	5.82	1.21	5.83	1.13

Qualitative Mid Infrared Analyses

Qualitative analysis of soil composition may be made by observation of peaks in the MIR spectra of soils and their PLS loadings. A representative spectra of the soil from the experimental field showing some dominant soil components is given in Figure 1.
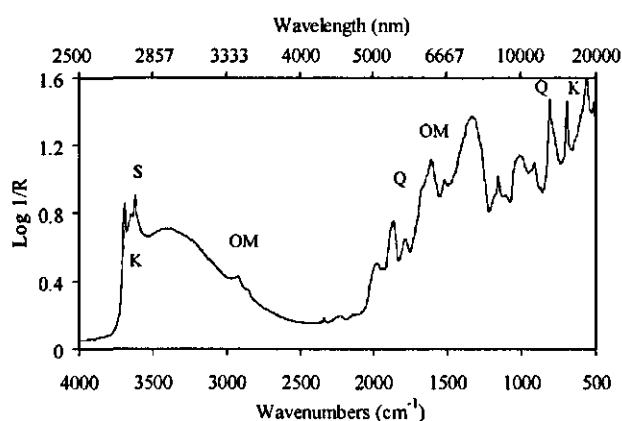

FIGURE 1 Representative mid infrared spectrum of soil from the experimental field. Dominant soil components: quartz (Q) as sand, organic matter (OM), kaolinite (K), and smectite (S)

The major spectral features associated with the first three factor loadings weights (see below) for the predictions of soil pH$_{CaCl_2}$, pH$_{buffer}$ and lime requirement are shown in Figure 2.
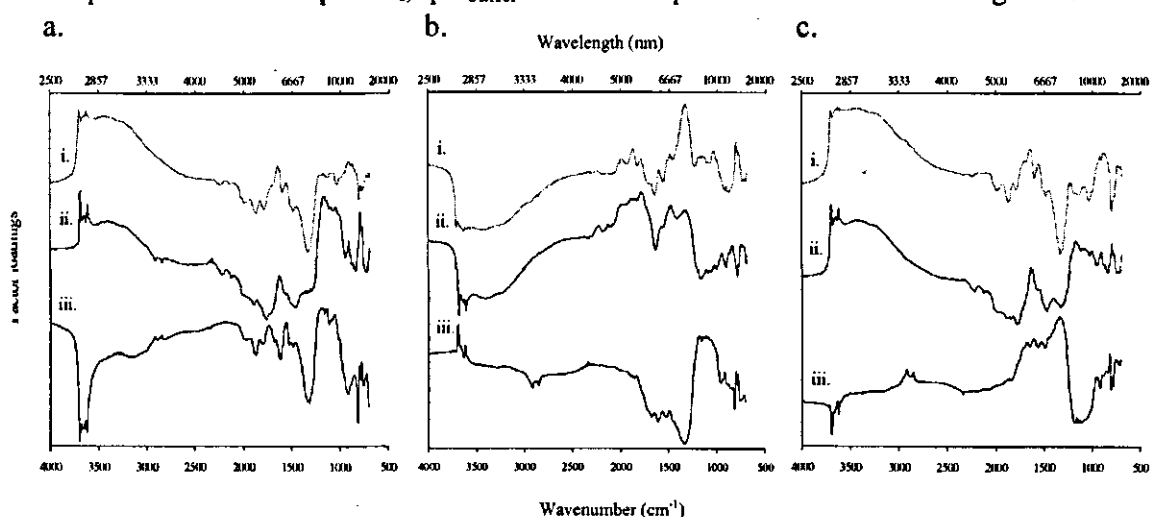

FIGURE 2. The first three factor loading weights (i., ii., iii., respectively) used for partial least squares predictions of (a.) soil pH$_{CaCl_2}$, (b.) soil Mehlich pH$_{buffer}$, and (c.) lime requirement to a target pH of 6.5.

499

The first few factor loadings contain most of the information, which indicate the correlations between the soil property in question and the MIR frequencies (Janik et al., 1995). Generally, for pH_{CaCl_2} and lime requirement the first three factor loadings show positive contributions from organic matter and smectite and negative contributions due to quartz as sand, and kaolin (Figure 2a and 2c, cf. Figure 1). All of these properties affect the buffering capacity of soil. The inverse relationship between pH_{buffer} and lime requirement is shown by their loadings spectra (Figure 2b and 2c).

Quantitative Mid Infrared Analyses with Partial Least Squares and Cross Validation

Quantitative analysis to determine the suitability of the PLS factor loadings and scores to model the spectra involved testing up to 25 factors for each soil property. Results for eighteen of these factors are shown in Figure 3 but fifteen factors were sufficient for adequate predictions of the soil properties tested.
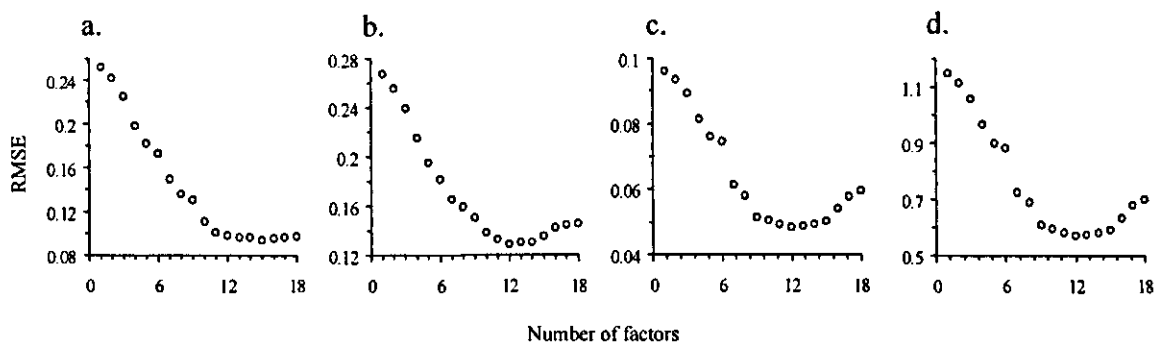
a.   b.   c.   d.

Number of factors

FIGURE 3. Root mean-square error (RMSE) of prediction for (a.) soil pH_{CaCl_2}, (b.) pH_{H_2O}, (c.) pH_{buffer} and (d.) lime requirement to a target pH of 6.5

For all soil properties, rapid reductions in the RMSE occurred for ten factors, after which only small decreases were evident (Figure 3). The optimal number of factor loadings (i.e., minimum RMSE values resulting from the cross-validation procedure) was twelve for all soil properties except pH_{CaCl_2} whose optimal number of factors for prediction and spectral reconstruction was fifteen. Without significant loss in prediction accuracy, it was possible to consider only eleven factor loadings for both pH_{CaCl_2} and pH_{H_2O} (equivalent to a RMSE of 0.1), and nine factors for both pH_{buffer} and lime requirement (equivalent to a RMSE of 0.05 and 0.6 respectively). The large number of factors detected and used in the analysis indicates the complexity of the soil matrix. These factor loadings were then used to construct the calibration matrix for the (independent) prediction of soil properties using the delete-one-jackknife procedure.

Linear regression plots between 'known' soil properties determined using conventional laboratory techniques, and those predicted from MIR-PLS cross-validation are given in Figure 4. Soil pH_{CaCl_2}, pH_{H_2O}, pH_{buffer} and lime requirements had relatively low, acceptable RMSEs of prediction (Figure 4). Best predictions were obtained for soil pH_{CaCl_2} with a R^2 value of 0.85 and a slope of 0.9 (Figure 4a). Lime requirement tended to be over-estimated at low values and under-estimated at high lime requirements (Figure 4d). This is reflected by the poorer prediction R^2 value of 0.75 and the lower slope 0.82 (Figure 4a).
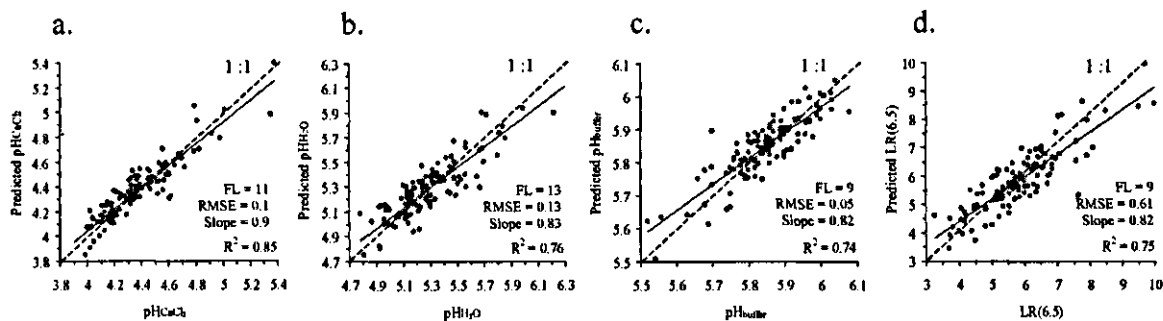
500

FIGURE 4. Regression plots for MIR-PLS predicted vs. 'known' soil data analysed using conventional laboratory methods of analyses for (a.) pH_{CaCl_2}, (b.) pH_{H_2O}, (c.) pH_{buffer} and (d.) lime requirement to a target pH of 6.5. Given for each property are the number of factor loadings (FL) used in the predictions as well as the corresponding RMSE. n = 118

Spatial Analysis

An advantage of the MIR technique is that many properties may be determined from a single spectrum. For example in this instance, maps of soil pH_{CaCl_2}, pH_{H_2O}, pH_{buffer} and lime requirement may be easily drawn from the MIR-PLS predictions (statistics of predictions given in lower half of Table 1). For brevity of presentation, only the spatial analysis of lime requirements will be presented. The exponential semi-variogram parameters for lime requirements derived using the response surface and those predicted using the MIR-PLS technique are shown in Table 2.

TABLE 2. Exponential semi-variogram parameters for lime requirement to a target pH of 6.5 (LR(6.5)) (Mg/ha) predicted using the response surface (RS) and mid infrared analysis (MIR)

	C_0	C	C_0/C_0+C	a' (m)
RS LR(6.5)	0.26	1.22	0.18	59
MIR LR(6.5)	0.15	1.15	0.12	56

Note: a' is the the effective range, i.e. the lag where $\gamma(a')$ is approximately equal to $(C_0 + 0.95C)$,

The most notable difference in the semi-variogram parameters shown in Table 2 are the nugget variances (C_0), C_0 being higher for the lime requirements derived using conventional laboratory methods and the response surface model. The proportion of nugget to sill variance for the latter is 18 %, while for the lime requirements predicted using the MIR-PLS technique is 12 %. This may suggest that the analytical precision of the MIR technique is higher than conventional analysis. Viscarra Rossel & McBratney (1998b) discussed the implications that the nugget variance has on soil sampling and analysis for precision agriculture.

The kriged predictions of lime requirements are compared in Figure 5a and the kriged map of MIR-PLS predicted lime requirements shown in Figure 5b.

Kriged lime requirements were comparable (Figure 5a). The larger nugget to sill ratio of the lime requirements derived by conventional analysis and using the response surface model resulted in a more even distribution of estimates having higher values and a slightly larger kriging variance (data not shown). Thus the lime requirement map derived from interpolation of the MIR-PLS predictions (Figure 5b) is slightly more variable and may be deemed suitable for variable-rate lime applications.

CONCLUSIONS

Mid-infrared analysis may be used to enhance existing information, with the possibility for its use in the development of a field-based proximal soil MIR sensing system. An advantage of

the technique is that from a single spectrum many properties may be determined, thus offering the possibility for considerable cost savings and increased efficiency over conventional laboratory soil extractions. Furthermore, the MIR technique is rapid, making it possible to analyse a large number of samples in a practical and timely manner. These properties make MIR-PLS analyses very attractive for spatial soil studies and precision agriculture.

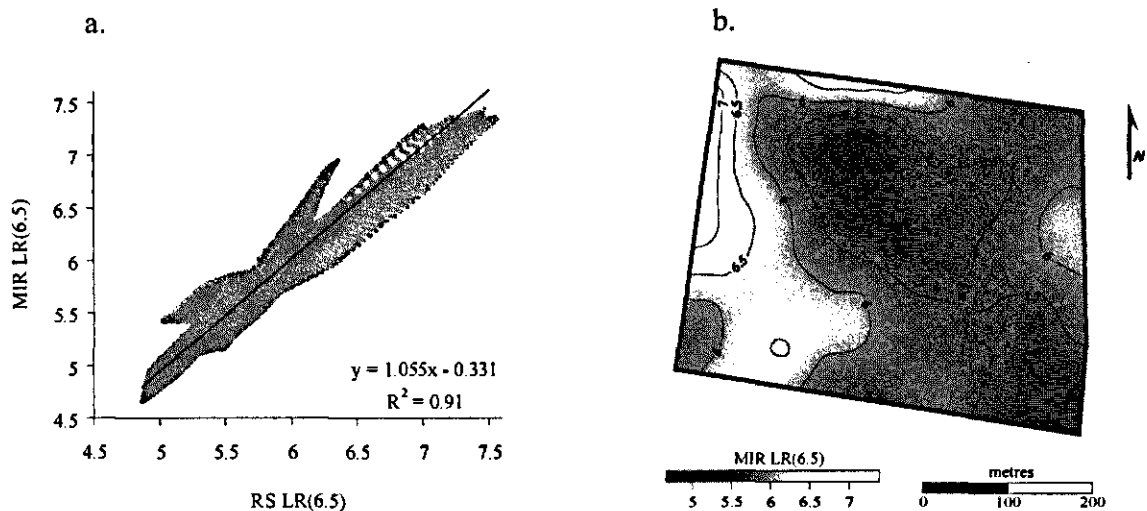a.   b.

FIGURE 5. (a.) regression plot between kriged mid infrared (MIR) and response surface (RS) lime requirements to a target pH of 6.5 LR(6.5), and (b.) MIR map of LR(6.5). All lime requirements in Mg/ha.

REFERENCES

Efron, B., Tibshirani, R.J. (1994). An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability, No 57. CRC Press, pp. 436

Janik, L.J., Skjemstad, J.O. and Raven, M.D. 1995. Characterisation and analysis of soils using mid-infrared partial least squares. I. Correlations with XRF-determined major element composition. *Australian Journal of Soil research* 33: 621-636.

Janik, L.J., Merry, R.H. and Skjemstad, J.O. 1998. Can mid infra-red diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture,* 38(7): 681-696.

Martens H., Næs, T. (1989). Multivariate Calibration. John Wiley & Sons, Chichester, p. 419.

McCauley, J.D., Engel, B.A., Scudder, C.E., Morgan, M.T. and Elliot, P.W. 1993. Assessing the spatial variability of organic matter. *American Society of Agricultural Engineers.* ASAE Paper No. 93-1555, St. Joseph, MI, USA.

Viscarra Rossel R.A. and McBratney, A.B. 2001. A response-surface calibration model for rapid and versatile site-specific lime-requirement predictions in south-eastern Australia. *Australian Journal of Soil Research* 39: 185-201.

Viscarra Rossel, R.A. and McBratney, A.B. 1998a. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of clay and water content. *Geoderma,* 85(1): 19-39.

Viscarra Rossel, R.A. and McBratney, A.B. 1998b. Soil chemical analytical accuracy and costs: implications from Precision Agriculture. *Australian Journal of Experimental Agriculture* 38: 765-775.