Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties

R.A. Viscarra Rossel[a,*], D.J.J. Walvoort[b], A.B. McBratney[a], L.J. Janik[c], J.O. Skjemstad[c]

[a]Australian Centre for Precision Agriculture, Faculty of Agriculture, Food and Natural Resources, McMillan Building A05, The University of Sydney, NSW 2006, Australia
[b]Wageningen University and Research Centre, Wageningen, The Netherlands
[c]CSIRO Land and Water, Glen Osmond SA 5064, Australia

Abstract

Historically, our understanding of the soil and assessment of its quality and function has been gained through routine soil chemical and physical laboratory analysis. There is a global thrust towards the development of more time- and cost-efficient methodologies for soil analysis as there is a great demand for larger amounts of good quality, inexpensive soil data to be used in environmental monitoring, modelling and precision agriculture. Diffuse reflectance spectroscopy provides a good alternative that may be used to enhance or replace conventional methods of soil analysis, as it overcomes some of their limitations. Spectroscopy is rapid, timely, less expensive, non-destructive, straightforward and sometimes more accurate than conventional analysis. Furthermore, a single spectrum allows for simultaneous characterisation of various soil properties and the techniques are adaptable for 'on-the-go' field use. The aims of this paper are threefold: (i) determine the value of qualitative analysis in the visible (VIS) (400–700 nm), near infrared (NIR) (700–2500 nm) and mid infrared (MIR) (2500–25,000 nm); (ii) compare the simultaneous predictions of a number of different soil properties in each of these regions and the combined VIS–NIR–MIR to determine whether the combined information produces better predictions of soil properties than each of the individual regions; and (iii) deduce which of these regions may be best suited for simultaneous analysis of various soil properties. In this instance we implemented partial least-squares regression (PLSR) to construct calibration models, which were independently validated for the prediction of various soil properties from the soil spectra. The soil properties examined were soil pH_{Ca}, pH_w, lime requirement (LR), organic carbon (OC), clay, silt, sand, cation exchange capacity (CEC), exchangeable calcium (Ca), exchangeable aluminium (Al), nitrate–nitrogen (NO_3–N), available phosphorus (P_{Col}), exchangeable potassium (K) and electrical conductivity (EC). Our results demonstrated the value of qualitative soil interpretations using the loading weight vectors from the PLSR decomposition. The MIR was more suitable than the VIS or NIR for this type of analysis due to the higher incidence spectral bands in this region as well as the higher intensity and specificity of the signal. Quantitatively, the accuracy of PLSR predictions in

each of the VIS, NIR, MIR and VIS–NIR–MIR spectral regions varied considerably amongst properties. However, more accurate predictions were obtained using the MIR for pH, LR, OC, CEC, clay, silt and sand contents, P and EC. The NIR produced more accurate predictions for exchangeable Al and K than any of the ranges. There were only minor improvements in predictions of clay, silt and sand content using the combined VIS–NIR–MIR range. This work demonstrates the potential of diffuse reflectance spectroscopy using the VIS, NIR and MIR for more efficient soil analysis and the acquisition of soil information.

1. Introduction

Soil is a heterogeneous system whose processes and mechanisms are complex and difficult to fully comprehend. Many conventional soil analytical techniques are used in an attempt to establish the relationship between soil physical and chemical properties and individual soil components, often disregarding their complex, multi-component interactions. Indeed, soil chemical extractions that alter the equilibrium between the phases may further complicate the interpretation of results. Historically our understanding of the soil system and assessment of its quality and function has been gained through this type of laboratory analysis. We need to further develop our analytical techniques to better understand the soil as a complete system and a resource so that we may make more efficient use of it and simultaneously preserve it for future generations. This is more important now than ever before since the acquisition of larger amounts of accurate soil data is essential if we are to manage our base resources sensibly to meet the food and fibre demands of future populations (Viscarra Rossel and McBratney, 1998a). It is perhaps for these reasons that spectroscopic techniques (e.g. mass spectroscopy (MS), nuclear magnetic resonance (NMR), visible (VIS), near infrared (NIR) and mid infrared (MIR) spectroscopy) are being considered as possible alternatives (or surrogates) to enhance or replace conventional laboratory methods of soil analysis (Janik et al., 1998). Most of these techniques are non-destructive and therefore allow the preservation of the basic integrity of the soil system. Furthermore, and in particular with infrared spectroscopy, a single spectrum allows for simultaneous characterisation of various soil constituents. VIS and infrared reflectance spectroscopy have advantages over some of the conventional techniques of soil analysis, e.g. they are rapid, timely and less expensive, hence are more efficient when a large number of analyses and samples are required. Moreover, spectroscopic techniques do not require expensive and time-consuming sample pre-processing or the use of (environmentally harmful) chemical extractants. Visible and infrared spectroscopy may, on instances, be more straightforward than conventional soil analysis and on occasions also more accurate. For example, McCauley et al. (1993) suggested that VIS spectroscopy may be more accurate than dichromate digestions for analysis of soil organic carbon and Viscarra Rossel et al. (2001) suggested that the precision of the MIR–PLS technique for predictions of soil pH and lime requirement is higher than conventional analysis). One other advantage is the potential adaptability of the techniques for 'on-the-go' field use (Viscarra Rossel and McBratney, 1998b). These are particularly important advantages now that there is an increasing global need for larger amounts of good quality inexpensive spatial soil data to be used in environmental monitoring, modelling and precision agriculture.

Infrared spectroscopic techniques are highly sensitive to both organic and inorganic phases of the soil, making their use in the agricultural and environmental sciences particularly relevant. Intense fundamental molecular frequencies related to soil components occur in the MIR between wavelengths 2500 and 25,000 nm. The visible and infrared portions of the electromagnetic spectrum are highlighted in Fig. 1.

Weak overtones and combinations of these fundamental vibrations due to the stretching and bending of NH, OH and CH groups dominate the NIR (700–2500 nm) and electronic transitions the VIS (400–700 nm) portions of the electromagnetic (EM) spectrum. Over-
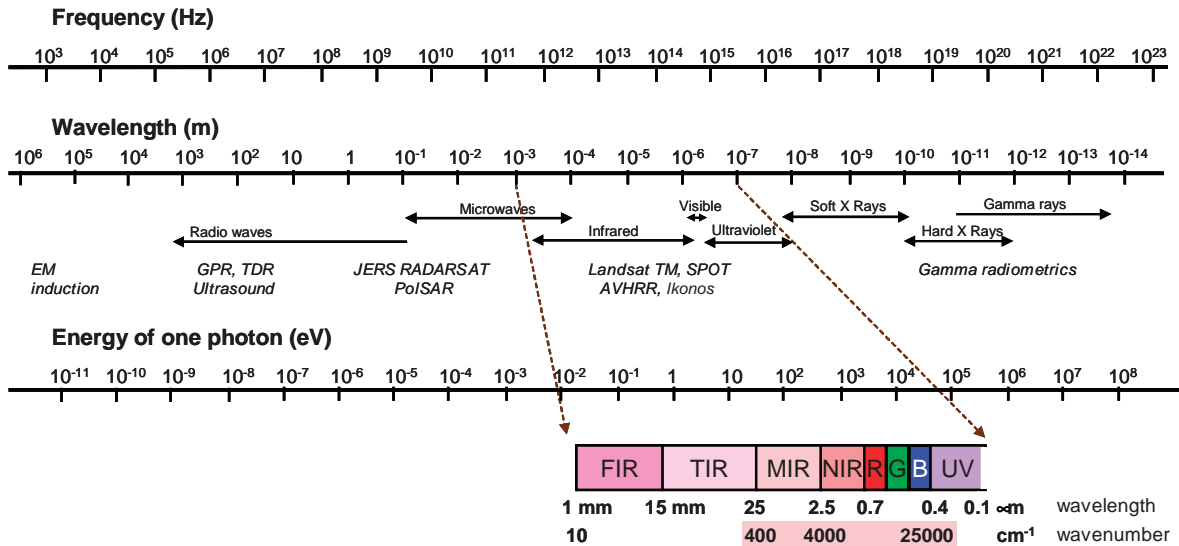
Fig. 1. The electromagnetic (EM) spectrum highlighting the visible and infrared portions (after McBratney et al., 2003).

tones and combination modes, which can also occur in the MIR, overlap making qualitative and quantitative interpretations in the visible and NIR more difficult. Janik et al. (1998) provide a good review of soil analysis using infrared techniques with particular attention to MIR.

Quantitative spectral analysis of soil using visible and infrared reflectance spectroscopy requires sophisticated statistical techniques to discern the response of soil attributes from spectral characteristics. Various methods have been used to relate soil spectra to soil attributes. For example, Ben-Dor and Banin (1995) used multiple regression analysis (MRA) to relate specific bands in the NIR to a number of soil properties. Shibusawa et al. (2001) used stepwise multiple linear regression (SMLR) for the estimation of various soil properties from the NIR spectra of soil acquired by a field-deployed on-the-go soil sensing system. Shepherd and Walsh (2002) used multivariate adaptive regression splines (MARS) for the estimation of soil properties from soil spectral libraries. Fidêncio et al. (2002) employed radial basis function networks (RBFN) to relate soil organic matter to soil spectra in the NIR region. Daniel et al. (2003) implemented artificial neural networks to estimate soil organic matter, phosphorus and potassium from the VIS–NIR spectrum. However, principal components regression (PCR) (e.g. Chang et al., 2001) and partial least-squares regression (PLSR) (e.g. McCarty et al., 2002) are the most common techniques for spectral calibration and prediction. PLSR is performed in a slightly different manner to PCR. Rather than first decomposing the spectra into a set of eigenvectors and scores and performing the regression with soil attributes in a separate step, PLSR actually uses the soil information during the decomposition process. PLSR takes advantage of the correlation that exists between the spectra and the soil, thus the resulting spectral vectors are directly related to the soil attribute (Geladi and Kowalski, 1986). The advantages of PLSR are that it handles multicollinearity, it is robust in terms of data noise and missing values, and unlike PCR it balances the two objectives of explaining response and predictor variation (thus calibrations and predictions are more robust) and it performs the decomposition and regression in a single step. Table 1 provides a review of the literature comparing quantitative predictions of various primary and secondary soil attributes, from aluminium to zinc, using various multivariate techniques and spectral response in the ultra violet (UV), VIS, NIR and MIR regions of the spectrum.

Table 1 demonstrates the potential of quantitative visible and infrared soil analysis, suggesting the regions of the spectrum that may be most suitable for predictions of various soil properties. For example,

Table 1

A review of the literature comparing quantitative predictions of various soil attributes using a multivariate statistical technique and spectral response in the ultra violet (UV), visible (VIS), near infrared (NIR) and mid infrared (MIR) regions of the electromagnetic spectrum

Soil attribute	Spectral region	Spectral range (nm)	Multivariate method[a]	n_{calib} | n_{valid}[b]	RMSE	R^2	Authors
Acid (exch.); cmol/kg	VIS–NIR	400–2498	PCR (11)	30|119	24.4	0.65	Chang et al. (2001)
Al (exch.); cmol/kg	MIR	2500–25,000	PLSR	183		0.64	Janik et al. (1998)
Biomass (N); mg/kg	NIR	1100–2300	PLSR (8)	180 x-val		0.71	Reeves and McCarty (2001)
Biomass (N); mg/kg	NIR	1100–2498	PLSR (6)	120|59		0.79	Reeves et al. (1999)
Biomass; g/kg	MIR	2500–25,000	PLSR	23		0.69	Janik et al. (1998)
Biomass; mg/kg	VIS–NIR	400–2498	PCR (9)	30|119	389.71	0.60	Chang et al. (2001)
C (inorg.); g/kg	MIR	2500–25,000	PLSR (16)	177|60		0.98	McCarty et al. (2002)
C (inorg.) g/kg	NIR	1100–2498	PLSR (19)	177|60		0.87	McCarty et al. (2002)
C (inorg.); g/kg	VIS–NIR	400–2498	PLSR (6)	76|32	0.15	0.96	Chang and Laird (2002)
C (total); g/kg	MIR	2500–25,000	PLSR (17)	177|60		0.95	McCarty et al. (2002)
C (total); g/kg	NIR	1100–2498	PLSR (16)	177|60		0.86	McCarty et al. (2002)
C (total); g/kg	NIR	1100–2498	PLSR (7)	120|59		0.96	Reeves et al. (1999)
C (total); g/kg	VIS–NIR	400–2498	PLSR (5)	76|32	0.65	0.91	Chang and Laird (2002)
C (total); g/kg	VIS–NIR	400–2498	PCR (7)	30|119	0.79	0.87	Chang et al. (2001)
C; %	UV–VIS–NIR	250–2450	PLSR (6)	59 x-val	0.06		Walvoort and McBratney (2001)
C:N ratio	VIS–NIR	400–2498	PLSR (6)	76|32	0.21	0.88	Chang and Laird (2002)
CEC; cmol(+)/kg	MIR	2500–25,000	PLSR	183		0.88	Janik et al. (1998)
CEC; cmol(+)/kg	NIR	1000–2500	MRA (63 bands)	35|56		0.64	Ben-Dor and Banin (1995)
CEC; mmol(+)/kg	NIR	700–2500	PCR	121|40		0.67	Islam et al. (2003)
CEC; cmol(+)/kg	VIS–NIR	400–2498	PCR (8)	30|119	38.2	0.81	Chang et al. (2001)
CEC; cmol(+)/kg	VIS–NIR	350–2500	MARS	493|247	38	0.88	Shepherd and Walsh (2002)
CEC; mmol(+)/kg	UV–VIS–NIR	250–2500	PCR	121|40		0.64	Islam et al. (2003)
Σ exch. cations; cmol(+)/kg	MIR	2500–20,000	PLSR			0.84	Janik and Skjemstad (1995)
Ca; cmol/kg	MIR	2500–25,000	PLSR	183		0.89	Janik et al. (1998)
Ca; mmol(+)/kg	NIR	700–2500	PCR	121|40		0.72	Islam et al. (2003)
Ca; g/kg	VIS–NIR	400–2500	modified PLSR	309		0.90	Cozzolino and Moron (2003)
Ca (exch.); cmol(+)/kg	VIS–NIR	350–2500	MARS	493|247	28	0.88	Shepherd and Walsh (2002)
Ca (exch.); cmol(+)/kg	VIS–NIR	400–2498	PCR (12)	30|119	40	0.75	Chang et al. (2001)
Ca; mmol(+)/kg	UV–VIS–NIR	250–2500	PCR	121|40		0.67	Islam et al. (2003)
Carbonate; %	MIR	2500–20,000	PLSR			0.95	Janik and Skjemstad (1995)
Carbonate; %	NIR	1000–2500	MRA (25–3113 bands)	22|47		0.69	Ben-Dor and Banin (1995)
Cu; mg/kg	VIS–NIR	400–2500	modified PLSR	310		0.82	Cozzolino and Moron (2003)
Cu (Mehlich III); mg/kg	VIS–NIR	400–2498	PCR (14)	30|119	5.37	0.25	Chang et al. (2001)
EC; dS/m	MIR	2500–25,000	PLSR	88		0.23	Janik et al. (1998)
EC; µS/cm	VIS–NIR	400–2400	SMLR (456, 984, 1014)	15|10		0.65	Shibusawa et al. (2001)
EC; mS/cm	UV–VIS–NIR	250–2500	PCR	121|40		0.10	Islam et al. (2003)
Fe (DTPA); mg/kg	MIR	2500–25,000	PLSR	183		0.55	Janik et al. (1998)
Fe (free); %	NIR	700–2500	PCR	121|40		0.49	Islam et al. (2003)
Fe; mg/kg	VIS–NIR	400–2500	modified PLSR	311		0.90	Cozzolino and Moron (2003)
Fe (Mehlich III); mg/kg	VIS–NIR	400–2498	PCR (9)	30|119	61.4	0.64	Chang et al. (2001)
Fe (free); %	UV–VIS–NIR	250–2500	PCR	121|40		0.52	Islam et al. (2003)
K; g/kg	VIS–NIR	400–2500	modified PLSR	317		0.72	Cozzolino and Moron (2003)
K; mmol(+)/kg	UV–VIS–NIR	250–2500	PCR	121|40		0.00	Islam et al. (2003)
K (exch.); mg/kg	MIR	2500–25,000	PLSR	183		0.33	Janik et al. (1998)
K (avail.); mg/kg	VIS–NIR	400–1100	NN	41		0.80	Daniel et al. (2003)
K (exch.); cmol/kg	VIS–NIR	400–2498	PCR (13)	30|119	4.2	0.55	Chang et al. (2001)
LR; t/ha	MIR	2500–25,000	PLSR	188		0.86	Janik et al. (1998)
LR; t/ha	NIR	700–2500	PLSR	188		0.73	Janik et al. (1998)
Mg (exch.); cmol/kg	MIR	2500–25,000	PLSR	183		0.76	Janik et al. (1998)
Mg; mmol(+)/kg	NIR	700–2500	PCR	121|40		0.59	Islam et al. (2003)

Table 1 (*continued*)

Soil attribute	Spectral region	Spectral range (nm)	Multivariate method[a]	n_{calib} | n_{valid}[b]	RMSE	R^2	Authors
Mg; g/kg	VIS–NIR	400–2500	modified PLSR	315		0.90	Cozzolino and Moron (2003)
Mg (exch.); cmol(+)/kg	VIS–NIR	350–2500	MARS	493|246	11	0.81	Shepherd and Walsh (2002)
Mg (exch.); mg/kg	VIS–NIR	400–2498	PCR (9)	30|119	12.8	0.68	Chang et al. (2001)
Mg; mmol(+)/kg	UV–VIS–NIR	250–2500	PCR	121|40		0.63	Islam et al. (2003)
Mn (DTPA); mg/kg	MIR	2500–25,000	PLSR	183		0.57	Janik et al. (1998)
Mn (exch.); cmol/kg	MIR	2500–25,000	PLSR	183		0.66	Janik et al. (1998)
Mn (Mehlich III); mg/kg	VIS–NIR	400–2498	PCR (12)	30|119	56.4	0.70	Chang et al. (2001)
N; %	MIR	2500–20,000	PLSR			0.88	Janik and Skjemstad (1995)
N; %	UV–VIS–NIR	250–2450	PLSR (11)	59 x-val	0.007		Walvoort and McBratney (2001)
N (NO_3); mg/100g	VIS–NIR	400–2400	SMLR (589, 1014)	15|10		0.54	Shibusawa et al. (2001)
N (miner.); mg/kg	NIR	1100–2498	PLSR (1)	179 x-val		0.08	Reeves et al. (1999)
N (pot. min); mg N/kg	VIS–NIR	400–2498	PCR (8)	30|119	26.05	0.72	Chang et al. (2001)
N (active); mg/kg	NIR	1100–2300	PLSR (8)	180 x-val		0.84	Reeves and McCarty (2001)
N (active); mg/kg	NIR	1100–2498	PLSR (9)	120|59		0.92	Reeves et al. (1999)
N organic (total); %	MIR	2500–25,000	PLSR	188		0.86	Janik et al. (1998)
N (total); %	NIR	1100–2500	MLR (1702, 1870, 2052)	72|48		0.92	Dalal and Henry (1986)
N (total); mg/kg	NIR	1100–2300	PLSR (10)	180 x-val		0.94	Reeves and McCarty (2001)
N (total); mg/kg	NIR	1100–2498	PLSR (8)	120|59		0.95	Reeves et al. (1999)
N (total); g/kg	VIS–NIR	400–2498	PLSR (7)	76|32	0.04	0.86	Chang and Laird (2002)
N (total); g/kg	VIS–NIR	400–2498	PCR (7)	30|119	0.06	0.85	Chang et al. (2001)
Na (exch.); cmol/kg	MIR	2500–25,000	PLSR	183		0.33	Janik et al. (1998)
Na (exch.); cmol/kg	VIS–NIR	400–2498	PCR (7)	30|119	1.3	0.09	Chang et al. (2001)
Na; mmol(+)/kg	UV–VIS–NIR	250–2500	PCR	121|40		0.34	Islam et al. (2003)
OC; %	MIR	2500–20,000	PLSR			0.92	Janik and Skjemstad (1995)
OC; %	MIR	2500–25,000	PLSR	188		0.93	Janik et al. (1998)
OC; g/kg	MIR	2500–25,000	PLSR (17)	177|60		0.94	McCarty et al. (2002)
OC; (acidified soil) g/kg	MIR	2500–25,000	PLSR (19)	177|60		0.97	McCarty et al. (2002)
OC; %	NIR	1100–2500	MLR (1744, 1870, 2052)	72|48		0.93	Dalal and Henry (1986)
OC; %	NIR	1100–2500	RBFN	140|60	0.32	0.96	Fidêncio et al. (2002)
OC; %	NIR	700–2500	PCR	121|40		0.68	Islam et al. (2003)
OC; g/kg	NIR	1100–2498	PLSR (18)	177|60		0.82	McCarty et al. (2002)
OC; mg/kg	NIR	1100–2300	PLSR (8)	180 x-val		0.94	Reeves and McCarty (2001)
OC (acidified soil); g/kg	NIR	1100–2498	PLSR (17)	177|60		0.80	McCarty et al. (2002)
OC; g/kg	VIS–NIR	400–2498	PLSR (6)	76|32	0.62	0.89	Chang and Laird (2002)
OC; g/kg	VIS–NIR	350–2500	MARS	449|225	0.31	0.80	Shepherd and Walsh (2002)
OC; dag/kg	VIS–NIR	350–1050	PLSR (5)	43|25	0.36		Viscarra Rossel et al. (2003)
OC; %	UV–VIS–NIR	250–2500	PCR	121|40		0.76	Islam et al. (2003)
OM; %	MIR	2500–25,000	PLSR (4)	31 x-val	0.72	0.98	Masserschmidt et al. (1999)
OM; %	NIR	1000–2500	MRA (30 bands)	39|52		0.55	Ben–Dor and Banin (1995)
OM; %	VIS–NIR	400–1100	NN	41		0.86	Daniel et al. (2003)
OM; %	VIS–NIR	400–2400	SMLR (606, 1311, 1238)	15|10		0.65	Shibusawa et al. (2001)
P (avail.); mg/kg	MIR	2500–25,000	PLSR	186		0.07	Janik et al. (1998)
P (avail.); mg/kg	VIS–NIR	400–1100	NN	41		0.81	Daniel et al. (2003)
pH	MIR	2500–20,000	PLSR			0.72	Janik and Skjemstad (1995)
pH	NIR	1100–2300	PLSR (8)	180 x-val		0.74	Reeves and McCarty (2001)
pH	NIR	1100–2498	PLSR (11)	120|59		0.73	Reeves et al. (1999)
pH	VIS–NIR	350–2500	MARS	505|253	0.43	0.70	Shepherd and Walsh (2002)
pH_{Ca}	MIR	2500–25,000	PLSR	183		0.67	Janik et al. (1998)

Table 1 (*continued*)

Soil attribute	Spectral region	Spectral range (nm)	Multivariate method[a]	n_{calib} | n_{valid}[b]	RMSE	R^2	Authors
pH$_{Ca}$	VIS–NIR	400–2498	PCR (13)	30|119	0.56	0.56	Chang et al. (2001)
pH$_w$	MIR	2500–25,000	PLSR	183		0.56	Janik et al. (1998)
pH$_w$	NIR	700–2500	PCR	121|40		0.70	Islam et al. (2003)
pH$_w$	VIS–NIR	400–2400	SMLR (959, 1214)	15|10		0.54	Shibusawa et al. (2001)
pH$_w$	VIS–NIR	400–2498	PCR (13)	30|119	0.55	0.57	Chang et al. (2001)
pH$_w$	UV–VIS–NIR	250–2500	PCR	121|40		0.71	Islam et al. (2003)
Clay; %	MIR	2500–20,000	PLSR			0.87	Janik and Skjemstad (1995)
Clay; %	MIR	2500–25,000	PLSR	88		0.79	Janik et al. (1998)
Clay; %	NIR	1000–2500	MRA (63 bands)	35|56		0.56	Ben-Dor and Banin (1995)
Clay; %	NIR	700–2500	PCR	121|40		0.75	Islam et al. (2003)
Clay; %	VIS–NIR	400–2498	PCR (12)	30|119	4.06	0.67	Chang et al. (2001)
Clay; g/kg	VIS–NIR	350–2500	MARS	305|152	7.5	0.78	Shepherd and Walsh (2002)
Clay; %	VIS–NIR	400–2500	modified PLSR	321		0.86	Cozzolino and Moron (2003)
Clay; %	UV–VIS–NIR	250–2450	PLSR (5)	59 x-val	2.9		Walvoort and McBratney (2001)
Clay; %	UV–VIS–NIR	250–2500	PCR	121|40		0.72	Islam et al. (2003)
Sand; %	MIR	2500–25,000	PLSR	88		0.94	Janik et al. (1998)
Sand; %	VIS–NIR	400–2498	PCR (8)	30|119	11.93	0.82	Chang et al. (2001)
Sand; %	VIS–NIR	400–2500	modified PLSR	319		0.70	Cozzolino and Moron (2003)
Sand; g/kg	VIS–NIR	350–2500	MARS	305|152	10.8	0.76	Shepherd and Walsh (2002)
Sand; %	UV–VIS–NIR	250–2500	PCR	121|40		0.53	Islam et al. (2003)
Silt; %	MIR	2500–25,000	PLSR	88		0.84	Janik et al. (1998)
Silt; %	VIS–NIR	400–2498	PCR (8)	30|119	9.51	0.84	Chang et al. (2001)
Silt; %	VIS–NIR	400–2500	modified PLSR	317		0.80	Cozzolino and Moron (2003)
Silt; g/kg	VIS–NIR	350–2500	MARS	305|152	4.9	0.67	Shepherd and Walsh (2002)
Silt; %	UV–VIS–NIR	250–2500	PCR	121|40		0.05	Islam et al. (2003)
Resp. rate; CO_2–C/kg/day	VIS–NIR	400–2498	PCR (9)	30|119	205.37	0.66	Chang et al. (2001)
Specific surface area; m^2/g	NIR	1000–2500	MRA (63 bands)	35|56		0.70	Ben-Dor and Banin (1995)
w 10 kPa; %	MIR	2500–25,000	PLSR	23		0.83	Janik et al. (1998)
w 30 kPa; %	MIR	2500–25,000	PLSR	23		0.90	Janik et al. (1998)
w (air dry); %	MIR	2500–25,000	PLSR	303		0.70	Janik et al. (1998)
w (air dry); g/g	NIR	700–2500	PCR	121|40		0.80	Islam et al. (2003)
w (air dry); %	NIR	1000–2500	MRA (25 bands)	35|56		0.62	Ben-Dor and Banin (1995)
w (air dry); %	NIR	1100–2500	MLR (1926, 1954, 2150)	72|48		0.97	Dalal and Henry (1986)
w (oven dry); %	VIS–NIR	400–2400	SMLR (606, 1329, 1499)	15|10		0.66	Shibusawa et al. (2001)
w; kg/kg	VIS–NIR	400–2498	PCR (8)	30|119	0.005	0.84	Chang et al. (2001)
w (air dry); g/g	UV–VIS–NIR	250–2500	PCR	121|40		0.85	Islam et al. (2003)
Zn (Mehlich III); mg/kg	VIS–NIR	400–2498	PCR (13)	30|119	15.28	0.44	Chang et al. (2001)

[a] Multivariate techniques include multiple regression analysis (MRA), stepwise multiple linear regression (SMLR), multivariate adaptive regression splines (MARS), radial basis function networks (RBFN), principle components regression (PCR), partial least-squares regression (PLSR). Shown in brackets are the spectral bands used or the number of bands or number of PCR components or number of PLSR factors used in the predictions.

[b] n_{calib}\|n_{valid} show the number of samples used in the spectral calibration and the number of factors use in the validation. X-val suggests that the validation was conducted independently using a statistical cross-validation technique.

in the MIR the average R^2 value for prediction of soil organic carbon and organic matter has been reported to be 0.96, in the NIR 0.81 and in the VIS 0.78 (Table 1). For predictions of clay content, the MIR has been shown to perform better with an average R^2 of 0.82 compared with 0.71 for the combined VIS–NIR. Cation exchange capacity (CEC) shows better response in the MIR with an average R^2 of prediction of 0.88 compared to 0.73 in the VIS–NIR. For other properties see Table 1. Although MIR spectra

generally produced more accurate results, the technology is more complex and more expensive than that used for VIS and NIR measurements. Thus it may be important to note that in a lot of cases there may be a trade-off between the relative accuracy of results and the cost of the instrumentation.

The aims of this paper are to: (i) determine the value of qualitative analysis in the visible (VIS) (400–700 nm), near infrared (NIR) (700–2500 nm) and mid infrared (MIR) (2500–25,000 nm); (ii) compare the simultaneous predictions of a number of different soil properties in each of these regions and the combined VIS–NIR–MIR to determine whether the combined information produces better predictions of soil properties than each of the individual regions; and (iii) deduce which of these regions may be best suited for simultaneous analysis of various soil properties.

2. Materials and methods

2.1. Soil sampling and laboratory analyses

Surface soil samples (0–20 cm) were collected from a 17.5 ha agricultural field in Kelso NSW, Australia, −33°24′04″ S and 149°38′51″ E, with a maximum elevation of 741 m above sea level. The soil in the field consisted of mainly red chromosols and brown kurosols (Isbell, 2002). The 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 summarised in Table 2.

2.2. Spectral analyses

VIS–NIR analysis involved sub-sampling portions of the ground <2 mm soil (~20 g) and placing them in a sample holder with a quartz window for spectral reflectance measurements. MIR analysis involved grinding these sub-samples 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). The spectral reflectance of soil samples was measured using two spectrometers: (i) an ultraviolet–visible–near infrared (UV–VIS–NIR) spectrometer (Varian Cary 500) equipped with a diffuse reflectance accessory (Labsphere DRA-CA-50D), with a spectral range of 350–2500 nm (28,570–4000 cm^{-1}); and (ii) a BioRad FTS 175 rapid scanning Fourier transform (FT) mid infrared (MIR) spectrometer, with an extended range KBr beam splitter and Peltier-cooled DTGS detector with a spectral range of 1200–20,000 nm (8300–470 cm^{-1} and 16 cm^{-1} resolution).

2.3. Spectral data management

Parts of the spectra were removed prior to performing the statistical analysis as these were thought to be either insensitive or influenced by artefacts produced by the spectrometers (Table 3).

The VIS–NIR and MIR reflectance data were combined into a single data table and transformed percent diffuse spectral reflectance (R%) values into optical density (i.e. absorbance) units—$\log_{10}(R)$ as

Table 2
Laboratory methods of soil analyses

Soil property	*n*	Technique	Reference
pH_{Ca}	118	1:5 soil:0.01 M $CaCl_2$	White (1969)
pH_W	118	1:5 soil:H_2O extract	White (1969)
Organic carbon (OC); dag/kg	118	Dichromate oxidation	Walkley and Black (1934)
Lime requirement[a] (LR); Mg/ha	118	Mehlich buffer (pH_b)	Mehlich (1976)
Sand, silt, clay; dag/kg	116	Hydrometer	Gee and Bauder (1986)
Cation exchange capacity (CEC); mmol(+)/kg	49	0.01 M $AgTU^+$	Rayment and Higginson (1992)
Exchangeable calcium (Ca); mmol(+)/kg	49	0.01 M $AgTU^+$	Rayment and Higginson (1992)
Exchangeable aluminium (Al); mmol(+)/kg	49	1 M KCl	Rayment and Higginson (1992)
Nitrate–nitrogen (NO_3–N); mg/kg	49	Colorimetrically	Rayment and Higginson (1992)
Available phosphorus Colwell (P_{Col}); mg/kg	49	Colorimetrically	Rayment and Higginson (1992)
Exchangeable potassium (K); mmol(+)/kg	49	Colorimetrically	Rayment and Higginson (1992)
Electrical conductivity (EC); dS/m	49	1:5 soil:H_2O extract	Rayment and Higginson (1992)

[a] The calibration model used to estimate lime requirement was that of Viscarra Rossel and McBratney (2001).

Table 3
Spectral ranges excluded from the analysis

Instrument	Wavelength; nm	Wave number; cm^{-1}	Reasons for exclusion
Varian Cary 500	795–810	12,579–12,346	Change of grating and filter
Varian Cary 500	2400–2499	4167–4002	Scan insensitive
FT–MIR	>14,286 and <2500	<700 and >4000	Scan insensitive

the latter shows better correlation to sample concentrations. We considered four spectral ranges in subsequent analysis (Table 4).

2.4. Chemometric analyses

Partial least-squares regression (PLSR) with leave-one-out cross-validation was used to calibrate the spectral data with the reference (laboratory) soil data.

2.4.1. Partial least-squares regression (PLSR)

PLSR is a popular modelling technique used in chemometrics and is commonly used for quantitative spectral analysis. It is used to construct predictive models when there are many predictor variables that are highly collinear. The technique is closely related to principal components regression (PCR). However, unlike PCR, the PLSR algorithm selects successive orthogonal factors that maximise the covariance between predictor (X spectra) and response variables (y laboratory data). By fitting a PLSR model, one hopes to find a few PLSR factors that explain most of the variation in both predictors and responses. Briefly, PLSR 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 mean-centred prior to decomposition. The decomposition is performed simultaneously and in such a way that the first few 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. Generally, the resulting matrices and vectors have a much lower dimension than X and y. Therefore, given a new spectrum x, the

soil property y can be estimated as a (bi)linear combination of the factors scores and factor loadings of x. For more details on the PLSR1 algorithm used see e.g. Martens and Næs (1989).

2.4.2. Selection of the optimal PLSR calibration model

Leave-one-out cross-validation (Efron and Tibshirani, 1994) was used to determine the number of factors to retain in the calibration models. In this instance, thirty bilinear factors were tested. To select the optimal cross-validated calibration model, we computed the root mean squared error (RMSE) of predictions. Generally, the model with the lowest RMSE is selected; however, we also tested the appropriateness of selecting a more parsimonious model than the one corresponding to the minimum RMSE. The model that gave the minimum RMSE was used as the point of reference (RMSE$_{ref}$) and all models with fewer factors (RMSE>RMSE$_{ref}$) were compared with the point of reference. The aim was to find a model with the fewest number of factors (m-factor), such that the RMSE for this model is not significantly greater than RMSE$_{ref}$. Hence we applied a randomisation t-test with $\alpha = 0.1$ (H_0: RMSE=RMSE$_{ref}$) (Van der Voet, 1994).

2.4.3. PLSR model predictions—validation

The calibration models were independently validated against the soil data. The procedure employed for the quantification of prediction biases and errors was also that of leave-one-out cross-validation (Efron and Tibshirani, 1994). The validation predictions involved computing each m-factor PLSR model using $n - 1$ soil samples and predicting the soil property of the sample removed. Explicitly, the procedure entailed removing a soil sample from the prediction set, then computing the m-factor PLSR model using the $n - 1$ samples and predicting the soil property of the sample removed. Then this sample is replaced into the

Table 4
Spectral data ranges considered in the analysis

Spectral range	Wavelength; nm	Wave number; cm^{-1}
VIS	400–795	25,000–12,579
NIR	810–2400	12,346–4167
MIR	2500–14,286	4000–700
VIS–NIR–MIR	400–14,286	25,000–700

prediction set and the subsequent sample removed. The m-Factor PLSR model is recomputed using the latter $n-1$ soil samples and prediction of the soil property is made for this sample removed.. and so on until all predictions are made. This procedure was repeated for all samples and accordingly, all soil properties. The mean error (ME) was used to quantify bias and the root mean square error (RMSE) the accuracy of predictions (their precision may then be easily inferred). We also calculated adjusted coefficients of determination (R_{adj}^2) values to compare our estimates to those found in other studies.

3. Results and discussion

3.1. Laboratory analysis

The surface soil used in the study is a typical light sandy clay loam found in the southern wheat-belt of New South Wales. Soil pH is acidic, exhibiting low levels of organic carbon, cation exchange capacity and nutrients and requiring high levels of lime for neutralisation. A statistical description of the soil data is given in Table 5.

In turn, these data together with the reflectance spectra of soil samples were used to develop PLSR calibration models. These models were validated using leave-one-out cross-validation.

3.2. Qualitative description of the spectral data

The spectra of all soils were similar, with minor features apparent in the visible and near infrared portions of the spectrum. Higher optical density and greater number of absorption peaks were apparent in mid infrared region (Fig. 2).

Fig. 2 shows a typical soil spectrum in each of the EM regions, indicating the red (R), green (G) and blue (B) bands in the visible range, water absorption bands in the near infrared and some dominant soil components, such as the type of clay mineral, quartz and organic compounds, in the mid infrared portion. Sets of spectra can be characterised qualitatively by observing the positive and negative peaks, which occur at specific wavelengths, of the first few loadings weight vectors from the PLSR decomposition. The first of these represent a first-order approximation to

Table 5
Statistical description of the observed soil data analysed using conventional laboratory methods of analyses and their cross-validated PLSR predictions in each of the visible (VIS), near infrared (NIR), mid infrared (MIR) and combined visible–infrared (VIS–NIR–MIR) regions of the EM spectrum

| Soil attribute	Observed				Predicted (VIS)			Predicted (NIR)			Predicted (MIR)			Predicted (VIS–NIR–MIR)		
	n	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
pH$_{Ca}$	118	4.37	0.26	3.90–5.38	4.37	0.17	3.94–4.94	4.37	0.24	3.91–5.07	4.37	0.25	3.83–5.44	4.37	0.25	3.67–5.42
pH$_w$	118	5.30	0.27	4.78–6.30	5.30	0.18	4.83–5.88	5.30	0.25	4.81–6.18	5.29	0.25	4.58–6.40	5.29	0.26	4.44–6.38
pH$_b$	118	5.85	0.10	5.52–6.08	5.85	0.06	5.66–5.96	5.85	0.10	5.59–6.18	5.85	0.09	5.50–6.06	5.85	0.09	5.44–6.07
LR	118	5.82	1.21	3.27–9.98	5.82	0.69	4.54–8.15	5.84	1.09	2.13–8.83	5.83	1.11	3.36–10.00	5.83	1.13	3.27–10.73
OC	118	1.34	0.28	0.81–1.98	1.34	0.23	0.94–2.03	1.34	0.23	0.80–2.14	1.34	0.25	0.85–2.14	1.34	0.25	0.92–2.14
Clay	116	14.23	3.04	8.00–24.14	14.23	2.10	9.91–22.49	14.23	2.46	7.66–21.22	14.21	2.69	4.75–20.69	14.24	2.63	5.26–22.47
Silt	116	14.54	3.07	6.02–20.18	14.55	2.04	6.81–19.91	14.58	2.17	7.86–18.54	14.55	2.33	6.41–17.72	14.56	2.42	5.89–18.23
Sand	116	70.58	5.17	58.02–83.98	70.58	4.00	62.41–83.66	70.54	4.19	59.73–82.11	70.61	4.77	61.10–87.32	70.57	4.62	61.18–87.51
CEC	49	39.27	11.38	24.00–72.00	39.12	5.71	26.39–54.30	39.11	5.21	29.67–56.01	38.77	9.15	19.80–64.75	39.00	5.46	27.64–53.90
Ca	49	24.24	8.13	11.00–46.00	24.14	5.69	13.15–37.53	24.15	3.17	18.21–34.52	23.98	6.58	10.90–42.97	23.90	7.02	5.40–44.79
Al	49	1.79	1.34	0.20–7.70	1.81	0.44	1.26–3.12	1.78	1.37	−1.18–7.33	1.81	1.04	0.33–5.42	1.80	1.05	−0.03–4.30
NO$_3$–N	49	3.30	3.62	0.90–15.00	3.32	1.25	1.14–8.57	3.35	0.90	2.31–8.30	3.33	0.71	1.17–5.20	3.35	1.05	1.71–7.96
P$_{Col}$	49	14.96	4.83	6.00–29.00	14.95	0.50	13.36–15.94	14.95	0.58	12.32–15.90	14.92	0.87	11.58–17.16	14.84	1.08	10.20–16.46
K	49	5.50	2.44	1.50–13.00	5.49	1.55	1.49–9.82	5.44	2.27	0.02–11.42	5.48	1.87	0.70–10.27	5.51	2.22	1.31–12.64
EC	49	0.043	0.012	0.020–0.070	0.043	0.0045	0.03–0.06	0.043	0.0042	0.04–0.06	0.043	0.01	0.021–0.067	0.042	0.0087	0.02–0.06
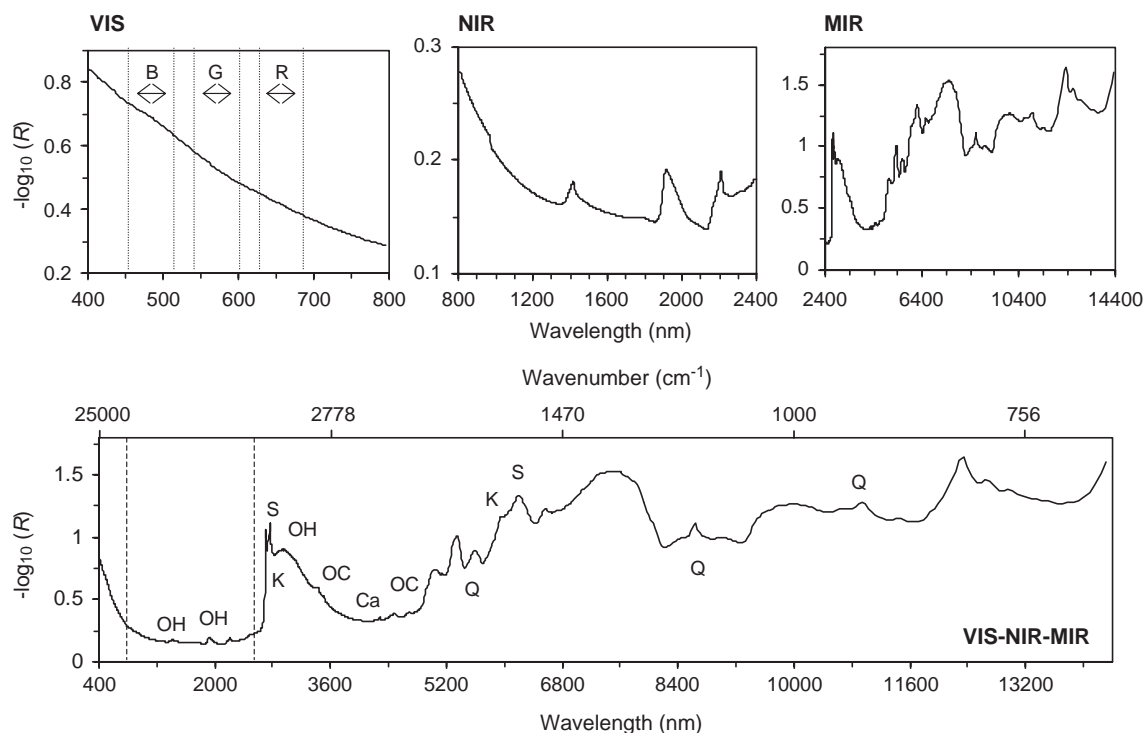
Fig. 2. Representative soil spectrum showing the visible (VIS), near infrared (NIR), mid infrared and the combined (VIS–NIR–MIR) regions. Red (R), green (G) and blue (B) bands are shown in the VIS spectrum. The VIS–NIR–MIR spectrum is annotated with some dominant soil components and absorption peaks: quartz (Q) as sand, organic compounds (OC), calcite (Ca), kaolinite (K), smectite (S), and the (OH) features of free water and lattice minerals.

the 'pure' component spectrum and can be useful for making assignments of spectral bands that may be important in the analysis (Haaland and Thomas, 1988). Positive peaks are due to the component of interest, while negative peaks correspond to interfering components (Haaland and Thomas, 1988). Thus the loading weight vectors can be considered as indicators of the correlation between the infrared frequencies and the soil constituents of interest. For example, the major features associated with the first three PLSR factor loadings weights of OC and pH_{Ca}, for each of the four spectral regions, are shown in Fig. 3, and for LR and sand content in Fig. 4.

NIR spectra are dominated by weak overtones and combinations of fundamental vibrational bands for H–C, H–N and O–H bonds that occur in the MIR region, while VIS spectra by electronic transitions. Although the usefulness of qualitative analysis is lower when there is a low intensity of spectral bands, as in the VIS

range (Fig. 2), moderate positive peaks that are important for the prediction of soil OC are evident at 410, 570 and 660 nm (Fig. 3a). This is consistent with other research findings (e.g. Shonk et al. 1991; Viscarra Rossel et al., 2003). For pH_{Ca}, the third factor loading weights show negative contributions for bands between 470 and 670 nm, with a pronounced negative peak at 600 nm, while positive contributions for bands <470 and >670 nm (Fig. 3b). For LR (Fig. 4a) there is a slight positive peak at 470 nm and there is an increasing importance of bands between 540 and 720 nm, peaking at around 590 nm. Negative contributions exist at 430 nm and at bands >720 nm. Positive peaks at 460, 540 and 650 nm, appear to be important for predictions of sand content there while interfering peaks occur in bands <450 and >670 nm (Fig. 4b).

In the NIR range the prominent features in both Figs. 3 and 4 are the absorption peaks associated
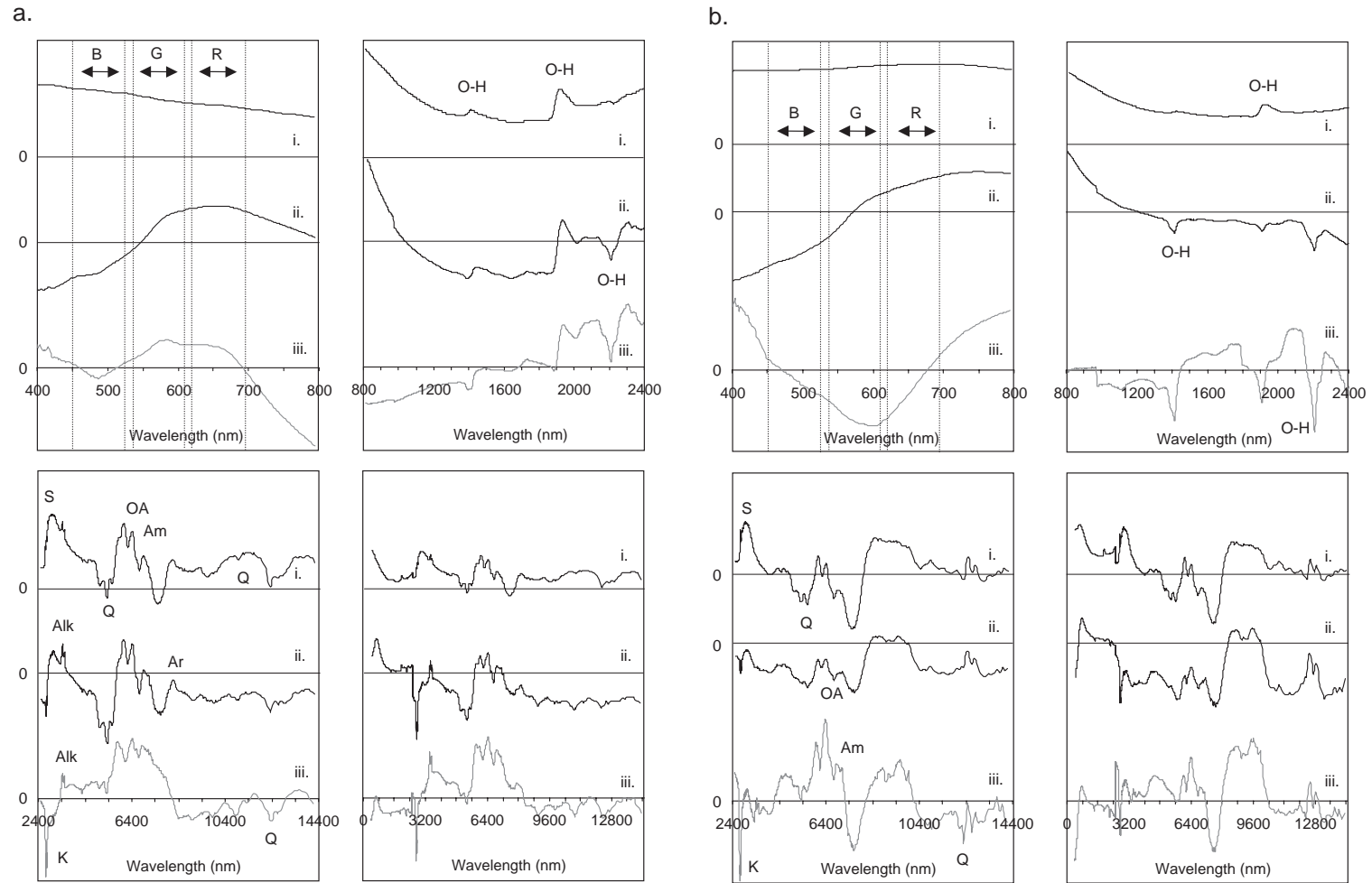
Fig. 3. The first three PLSR factor loading weight vectors (i, ii and iii) centred on zero for (a) soil organic carbon (OC) and (b) pH$_{Ca}$ in the visible (VIS), near infrared (NIR), mid infrared (MIR) and the combined VIS–NIR–MIR regions. The following features are indicated: in the VIS range red (R), green (G) and blue (B) bands, in the NIR O–H features of water and clay minerals and in the MIR soil components are identified as organic acids (OA), amides (Am), alkyls (Alk), and aromatic (Ar) groups, smectite (S), quartz (Q) and kaolin (K).
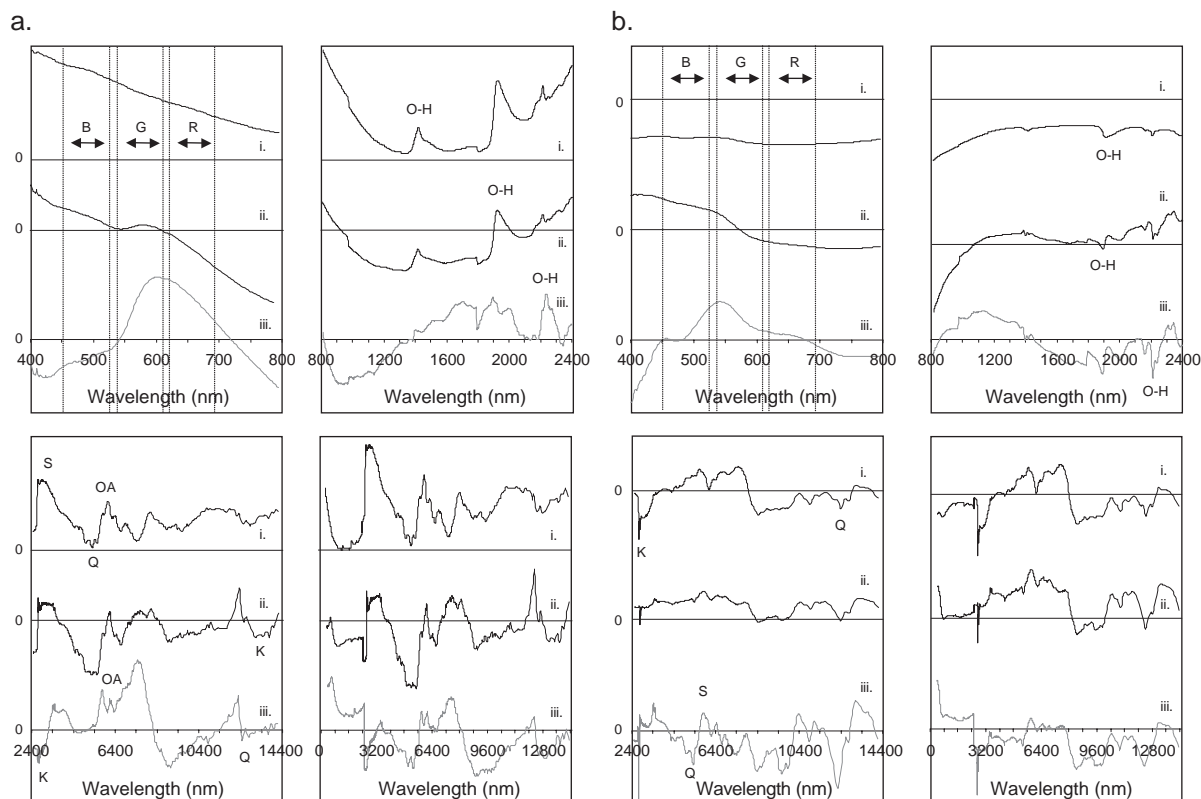
Fig. 4. The first three PLSR factor loading weight vectors (i, ii and iii) centred on zero for (a) lime requirement (LR) and (b) sand content in the visible (VIS), near infrared (NIR), mid infrared (MIR) and the combined VIS–NIR–MIR regions. The following features are indicated: in the VIS range red (R), green (G) and blue (B) bands, in the NIR O–H features of water and clay minerals and in the MIR soil components are identified as organic acids (OA), smectite (S), quartz (Q) and kaolin (K).

with the bending and stretching of the O–H bonds of free water around 1400, 1950 and in lattice minerals at 2200 nm. In the MIR region, organic compounds such as organic acids (e.g. COOH), alkyl, amide and aromatic groups show up as positive peaks (to varying degrees) in all three loading weight vectors of OC, while quartz and kaolin show up as interfering components (Fig. 3a). These findings are similar to those of Janik and Skjemstad (1995) and Janik et al. (1998). Generally, for pH_{Ca} (Fig. 3b), LR (Fig. 4a) and sand content (Fig. 4b) the first three factor loadings weights in the MIR show positive contributions from smectite and organic acids and negative contributions due to quartz as sand, and kaolin. The combined VIS–NIR–MIR loading weight vectors show similar features as those of the three independent regions of the spectrum (Figs. 3 and 4).

3.3. Multivariate analysis using partial least-squares regression (PLSR)

Multivariate calibration techniques such as PCR and PLSR have been used to extract soil attribute information from the reflectance spectra of soils in the VIS, NIR and MIR (e.g. see Table 1). To demonstrate how quantitative predictions and interpretations were made in this study, we will continue with the LR example. For the VIS LR calibration model the number of factors with the lowest RMSE was six (Fig. 5a VIS). However, by testing the appropriateness of selecting a more parsimonious model than the one with the lowest RMSE (Fig. 5b VIS) we found that a four-factor model caused an insignificant decrease in accuracy. Thus the more parsimonious four-factor calibration model was selected for the spectral reconstruction and statistically independent (cross-
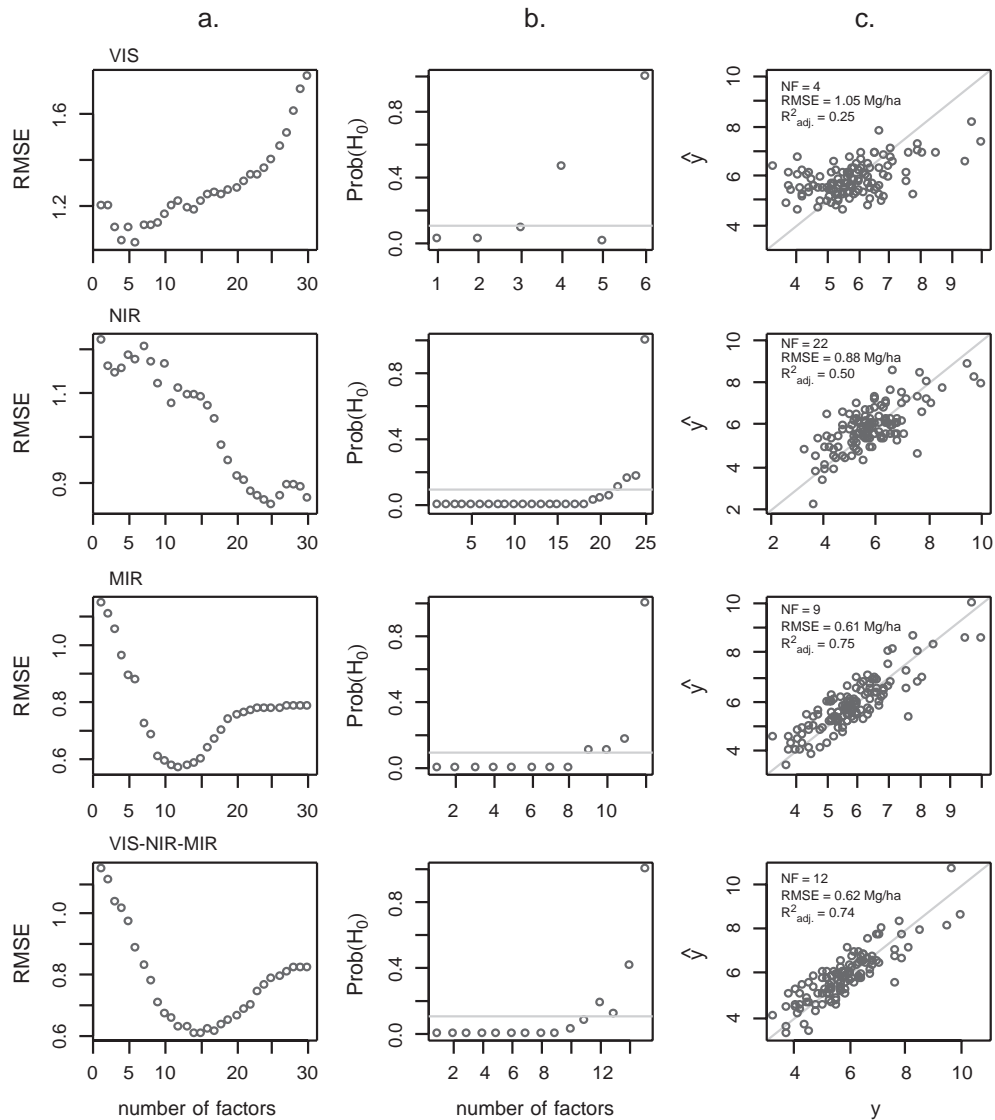
Fig. 5. Partial least-squares regression modelling and prediction output for soil lime requirement (LR) in each of the VIS, NIR, MIR and VIS–NIR–MIR. Columns: (a) shows the cross-validated root mean square error (RMSE) of prediction against the number of factors (NF); (b) shows selection of the model with the fewest number of factors, such that the RMSE for this model is equal to, or not significantly greater than RMSE$_{ref}$. The level of significance used was $\alpha = 0.1$; and (c) shows the observed (y) against the cross-validated PLSR predictions ($\hat{y}$) of soil LR with the validation statistics.

validated) predictions of LR (Fig. 5c VIS). The accuracy (RMSE) of predictions was 1.04 Mg/ha (Table 6), with predictions ranging from 4.54 to 8.15 Mg/ha (Table 5). In the NIR range, the number of factors for the LR calibration model with the lowest RMSE was twenty-five (Fig. 5a NIR) but the more parsimonious model was one with twenty-two factors

(Fig. 5b NIR) with an RMSE of 0.88 Mg/ha (Table 6). The cross-validated predictions are shown in Fig. 5c (NIR), ranging from 2.13 to 8.83 Mg/ha (Table 5). The predictions reported here have a lower R^2 value than that found in the literature (Table 1). (However, remember that by definition $R^2_{adj} < R^2$.) In the MIR the number of factors for the LR calibration model with

Table 6
Validation statistics for the cross-validated predictions of various soil attributes using partial least-squares regression (PLSR) and spectral response in the visible (VIS), near infrared (NIR), mid infrared (MIR) and combined VIS–NIR–MIR

Soil attribute	VIS					NIR					MIR					VIS–NIR–MIR				
	NF	ME	RMSE	R^2_{adj}	p-value	NF	ME	RMSE	R^2_{adj}	p-value	NF	ME	RMSE	R^2_{adj}	p-value	NF	ME	RMSE	R^2_{adj}	p-value
pH_{Ca}	5	0.00	0.21	0.36	0.000	20	0.00	0.17	0.57	0.000	11	0.00	0.10	0.86	0.000	14	0.00	0.10	0.85	0.000
pH_w	5	0.00	0.21	0.36	0.000	20	0.00	0.19	0.54	0.000	11	−0.01	0.13	0.75	0.000	14	−0.01	0.14	0.73	0.000
pH_b	4	0.00	0.09	0.24	0.000	23	0.00	0.07	0.53	0.000	9	0.00	0.05	0.74	0.000	12	0.00	0.05	0.73	0.000
LR	4	−0.01	1.05	0.25	0.000	22	0.01	0.88	0.50	0.000	9	0.00	0.61	0.75	0.000	12	0.01	0.62	0.74	0.000
OC	5	0.00	0.18	0.60	0.000	5	0.00	0.18	0.60	0.000	6	0.00	0.15	0.73	0.000	6	0.00	0.15	0.72	0.000
Clay	3	0.00	2.28	0.43	0.000	6	0.00	1.91	0.60	0.000	7	−0.02	1.74	0.67	0.000	6	0.01	1.73	0.67	0.000
Silt	6	0.01	2.55	0.31	0.000	6	0.03	2.35	0.41	0.000	4	0.00	2.17	0.49	0.000	6	0.02	2.11	0.52	0.000
Sand	7	0.00	3.77	0.47	0.000	6	−0.04	3.30	0.59	0.000	7	0.03	2.61	0.74	0.000	8	−0.01	2.57	0.75	0.000
CEC	1	−0.15	10.23	0.16	0.002	1	−0.15	10.43	0.13	0.006	5	−0.49	9.36	0.34	0.000	1	−0.26	10.77	0.09	0.021
Ca	5	−0.11	6.70	0.31	0.000	1	−0.09	7.72	0.07	0.039	4	−0.27	6.48	0.38	0.000	8	−0.35	6.60	0.37	0.000
Al	1	0.01	1.31	0.01	0.200	22	−0.02	0.88	0.61	0.000	5	0.02	1.00	0.43	0.000	1	0.00	1.06	0.37	0.000
NO_3–N	1	0.02	3.60	0.00	0.293	1	0.04	3.72	−0.02	0.795	1	0.03	3.59	−0.01	0.538	1	0.05	3.68	−0.02	0.734
P_{Col}	1	−0.01	4.94	0.06	0.047	1	−0.01	4.91	0.01	0.238	1	−0.04	5.24	0.20	0.001	1	−0.12	5.21	0.07	0.033
K	3	−0.02	2.02	0.29	0.000	16	−0.06	1.84	0.47	0.000	7	−0.02	1.92	0.38	0.000	10	0.01	1.83	0.46	0.000
EC	1	0.00	0.01	0.05	0.065	1	0.00	0.01	0.04	0.087	6	0.00	0.01	0.38	0.000	7	0.00	0.01	0.29	0.000

the lowest RMSE was twelve (Fig. 5a MIR), however, the more parsimonious nine-factor model (Fig. 5b MIR) was selected for predictions (Fig. 5c MIR). The accuracy of predictions was 0.61 Mg/ha (Table 6) and predictions ranged from 3.36 to 10.00 Mg/ha (Table 5). Using the entire VIS–NIR–MIR range parsimony was achieved using a twelve-factor PLSR calibration model (Fig. 5b VIS–NIR–MIR). Compared to predictions using MIR spectra, however, there was no significant improvement in accuracy (Fig. 5c VIS–NIR–MIR and Table 5).

A summary of the cross-validation prediction statistics of the remaining soil properties is given in Table 6. For example, in both the VIS and NIR ranges, predictions of soil OC produced a RMSE equal to 0.18 dag/kg and (Table 6). In the NIR the RMSE and R^2 values for cross-validation predictions of exchangeable Al were 0.88 mmol(+)/kg and 0.61, respectively (Table 6). This result is comparable to the R^2 of 0.65 reported in the review (Table 1). In the MIR cross-validation predictions of soil pH_{Ca} produced a RMSE of prediction of 0.1 pH units and a R^2_{adj} of 0.86 (Table 6). The assessment statistics for predictions of soil pH found in the literature review (Table 1) were less accurate.

3.4. Comparison of cross-validated predictions using VIS, NIR, MIR and combined VIS–NIR–MIR spectra

To compare the predictive ability of each of the VIS, NIR, MIR and VIS–NIR–MIR for different soil properties, we normalised the RMSE data by dividing it by the mean RMSE for each property in each spectral region and plotted the results. Predictions for each of the fifteen soil attributes using the three spectral regions and the combined spectra were compared to assess their usefulness for predictions of specific attributes (Fig. 6).

Predictions using VIS spectra were generally less accurate and had lower R^2_{adj} values than predictions using NIR, MIR or VIS–NIR–MIR range spectra (Fig. 6). The NIR region looks to be most useful for predictions of exchangeable Al and exchangeable K (Fig. 6). In the MIR, the ability to discriminate quartz, clay minerals and various organic constituents (Figs. 3 and 4) allowed good characterisation of various soil chemical and physical properties (Table 6). Thus, the MIR produced more accurate results (and higher R^2_{adj}
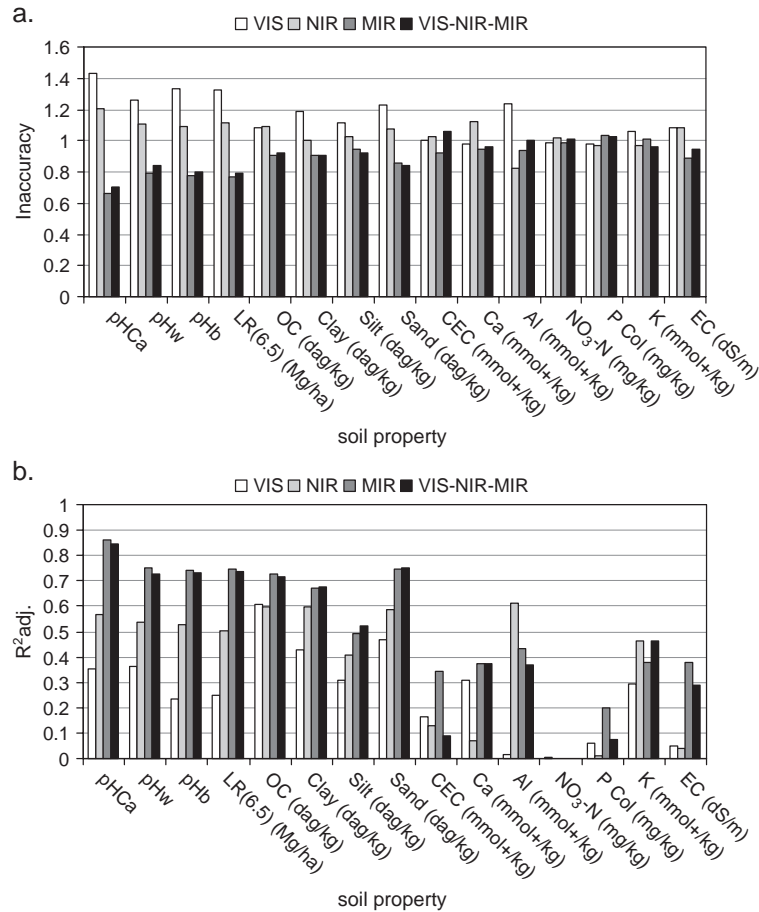
Fig. 6. Comparison between partial least-squares regression predictions of various soil attributes in the visible (VIS), near infrared (NIR), mid infrared (MIR) and combined visible-infrared (VIS–NIR–MIR) regions of the EM spectrum. (a) The inaccuracy of cross-validated predictions (normalised RMSE) and (b) R^2_{adj} values.

values) than VIS or NIR predictions for pH, LR, OC, CEC, clay, silt and sand contents, P and EC (Fig. 6). There were some slight improvements in the predictions of clay, silt and sand using information from the combined VIS–NIR–MIR spectra. By and large, however, there were no improvements in accuracy from the combined spectra. Nitrate–N did not show response in any of the spectral regions (Table 6 and Fig. 6).

Our results show that diffuse reflectance spectroscopy in each of the VIS, NIR and MIR can be used for the simultaneous assessment of various soil properties. The choice of which spectral region to use will depend on: (i) the accuracy of the predictions, (ii) the cost of the technology and (iii)

the amount of sample preparation required. For example, take the predictions of soil OC; if the aim is to characterise the spatial variability of soil OC, the 0.03% gain in prediction accuracy from using an MIR instead of an VIS spectrometer may not be significant as MIR spectrometers are considerably dearer and the technique requires more sample preparation. In this instance using a VIS spectrometer may suffice. For predictions of clay content, a 0.17% gain in prediction accuracy from using an MIR instead of an NIR spectrometer may not be significant because of the greater cost of the MIR technology. These trade-offs may be more pronounced when considering field implementation of spectroscopic techniques, e.g. as on-the-go proximal

soil sensing systems (Viscarra Rossel and McBratney, 1998b; Shibusawa et al., 2001). Presently, it is possible to purchase portable VIS and NIR fibre optic spectrometers that may be developed for on-the-go field use, however the cost and availability of robust portable MIR instruments is prohibitive. Clearly as time progresses and technologies develop, the availability and cost of MIR will decrease making it possible to use MIR in the field. For the time being, MIR spectroscopy will remain in the laboratory and has the potential to replace many of our conventional techniques of soil analysis.

4. Conclusion

Our work demonstrated the potential of diffuse reflectance spectroscopy, using the VIS, NIR, MIR and combined spectra. Diffuse reflectance spectroscopy may be used to more efficiently acquire soil information. We demonstrated how qualitative soil interpretations may aid with the identification and assignment of spectral bands to soil constituents. For example, in the VIS range bands at 410, 570 and 660 nm showed good correlations with soil OC. Various organic compounds (e.g. amides) throughout the MIR showed good correlations to OC, while quartz and kaolin showed up as interfering co-variant components. Quantitatively, the accuracy of PLSR predictions in each of the spectral regions varied considerably amongst properties; although predictions using the MIR were generally better. The following RMSEs were obtained for the simultaneous MIR–PLSR cross-validated predictions of: $pH_{Ca}=0.10$ pH units; $pHw=0.13$ pH units; $LR=0.61$ Mg/ha; $OC=0.15$ dag/kg; $CEC=9.36$ mmol(+)/kg; clay$=1.74$ dag/kg; silt$=2.17$ dag/kg; sand$=2.61$ dag/kg; $P_{Col}=5.24$ mg/kg and $EC=0.01$ dS/m. The NIR produced more accurate predictions for exchangeable Al and K (RMSE$=0.88$ and 1.84 mmol(+)/kg, respectively). There were only minor improvements in predictions of clay, silt and sand content (RMSEs$=1.73$, 2.11 and 2.57, respectively) using the combined VIS–NIR–MIR spectra. 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 analyse a large number of samples in a practical and timely manner. These properties make spectroscopic analyses combined with PLSR very attractive for environmental monitoring, modelling and precision agriculture.

References

Ben-Dor, E., Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of America Journal 59, 364–372.

Chang, C.-W., Laird, D.A., 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Science 167 (2), 110–116.

Chang, C.-W., Laird, D.A., Mausbach, M.J., Hurburgh Jr., C.R., 2001. Near-infrared reflectance spectroscopy—principal components regression analysis of soil properties. Soil Science Society of America Journal 65, 480–490.

Cozzolino, D., Moron, A., 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. Journal of Agricultural Sciences 140, 65–71.

Dalal, R.C., Henry, R.J., 1986. Simultaneous determination of moisture, organic carbon and total nitrogen by near infrared reflectance spectrophotometry. Soil Science Society of America Journal 50, 120–123.

Daniel, K.W., Tripathi, N.K., Honda, K., 2003. Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). Australian Journal of Soil Research 41, 47–59.

Efron, B., Tibshirani, R.J., 1994. An introduction to the Bootstrap. Monographs on Statistics and Applied Probability, vol. 57. CRC Press, p. 436.

Fidêncio, P.H., Poppi, R.J., De Andrade, J.C., 2002. Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy. Analytica Chimica Acta 453, 125–134.

Gee, G.W., Bauder, J.W., 1986. Particle size analysis. In: Klute, A. (Ed.), Methods of Soil Analysis: Part I. Second edition. Agronomy Monograph, vol. 9. ASA and SSSA, Madison, WI, USA, pp. 383–411.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta 185, 1–17.

Haaland, D.M., Thomas, E.V., 1988. Partial least-squares methods for spectral analyses: 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Analytical Chemistry 60, 1193–1202.

Isbell, R.F., 2002. The Australian Soil Classification. CSIRO Australia, Collingwood, VIC.

Islam, K., Singh, B., McBratney, A.B., 2003. Simultaneous estimation of various soil properties by ultra-violet, visible and near-infrared reflectance spectroscopy. Australian Journal of Soil Research 41, 1101–1114.

Janik, L.J., Skjemstad, J.O., 1995. Characterisation and analysis of soils using mid-infrared partial least squares: II. Correlations

with some laboratory data. Australian Journal of Soil Research 33, 637–650.

Janik, L.J., Skjemstad, J.O., Raven, M.D., 1995. Characterization 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., 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 and Sons, Chichester, p. 419.

Masserschmidt, I., Cuelbas, C.J., Poppi, R.J., De Andrade, J.C., De Abreu, C.A., Davanzo, C.U., 1999. Determination of organic matter in soils by FTIR/diffuse reflectance and multivariate calibration. Journal of Chemometrics 13, 265–273.

McBratney, A.B., Mendonca Santos, M.L, Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurements. Soil Science Society of America Journal 66, 640–646.

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

Mehlich, A., 1976. New buffer pH method for rapid estimation of exchangeable acidity and lime requirement of soils. Communications in Soil Science and Plant Analysis 7 (7), 637–652.

Rayment, G.E., Higginson, F.R., 1992. Australian Laboratory Handbook of Soil and Water Chemical Methods. Inkata Press, Australia.

Reeves III, J.B., McCarty, G.W., 2001. Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and fibre-optic probe. Journal of Near Infrared Spectroscopy 9, 25–34.

Reeves III, J.B., McCarty, G.W., Meisinger, J.J., 1999. Near infrared reflectance spectroscopy for the analysis of agricultural soils. Journal of Near Infrared Spectroscopy 7, 179–193.

Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal 66, 988–998.

Shibusawa, S., Imade Anom, S.W., Sato, S., Sasao, A., Hirako, S., 2001. Soil mapping using the real-time soil spectrophotometer.

In: Grenier, G., Blackmore, S. (Eds.), ECPA 2001, Third European Conference on Precision Agriculture, vol. 1. Agro Montpellier, pp. 497–508.

Shonk, G.A., Gaultney, L.D., Schulze, D.G., Van Scoyoc, G.E., 1991. Spectroscopic sensing of soil organic matter content. Trans ASAE 34 (5), 1978–1984.

Van der Voet, H., 1994. Comparing the predictive accuracy of models using a simple randomisation test. Chemometrics and Intelligent Laboratory Systems 25, 313–323.

Viscarra Rossel, R.A., McBratney, A.B., 1998a. Soil chemical analytical accuracy and costs: implications from precision agriculture. Australian Journal of Experimental Agriculture 38, 765–775.

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

Viscarra Rossel, R.A., 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., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2001. Proximal sensing of soil pH and lime requirement by mid infrared diffuse reflectance spectroscopy. In: Grenier, G., Blackmore, S. (Eds.), ECPA 2001, Third European Conference on Precision Agriculture, vol. 1. Agro Montpellier, pp. 497–508.

Viscarra Rossel, R.A., Walter, C., Fouad, Y., 2003. Assessment of two reflectance techniques for the quantification of field soil organic carbon. In: Stafford, J., Werner, A. (Eds.), Precision Agriculture. Fourth European Conference on Precision Agriculture. Wageningen Academic Publishers, Berlin, pp. 697–703.

Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Science 37, 29–38.

Walvoort, D.J.J., McBratney, A.B., 2001. Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. In: Grenier, G., Blackmore, S. (Eds.), ECPA 2001, Third European Conference on Precision Agriculture, vol. 1. Agro Montpellier, pp. 503–507.

White, R.E., 1969. On the measurement of soil pH. Journal of the Australian Institute of Agricultural Science 35, 3–14.