Mapping homogeneous spectral response zones in a soil profile

Edward J. Jones[1], Alex. B. McBratney[1]

[1] Sydney Institute of Agriculture, Faculty of Science, The University of Sydney.

5   *Correspondence to*: Edward J. Jones (edward.jones@sydney.edu.au)

Abstract.

Homogeneous spectral response zones represent relatively uniform regions of soil that may be useful for identifying soil horizons or delineating soil units spatially. External parameter orthogonalisation (EPO) and direct standardisation (DS) were assessed for their ability to conserve intrinsic soil information of spectra under variable moisture conditions, as experienced

10   when taking measurements *in situ*. A 1 m x 1 m section of a soil profile was intensively sampled using visible near-infrared diffuse reflectance spectroscopy at 2.5 cm vertical intervals and 10 cm horizontal intervals. Further samples were taken on a 10 cm grid and scanned in a laboratory under field moist and air-dry conditions. A principal component space was constructed based on the *in situ* scans following either EPO transformation, DS transformation or following pre-processing only (PP). Scores from the first four principal components – which accounted for more than 0.97 of the total variance in each

15   case – were subject to k-means clustering to identify homogeneous spectral response zones. Laboratory-based scans were then projected onto the same principal component space and fitted to the pre-existing cluster centroids. Both EPO and DS were found to have potential in reconciling differences observed between *in situ* and laboratory-based measurements compared to pre-processing only (PP). EPO outperformed DS in terms of conserving the relationship between PC scores (LCCC: EPO = 0.84, DS = 0.58, PPO = 0.44; RMSE: EPO = 11.8, DS = 15.4, PPO = 38.5) and also in identifying

20   homogeneous spectral response zones that corresponded to field observed horizons.

1. Introduction

Horizons are characteristic features of soils, which represent regions of relative uniformity in a highly heterogeneous medium. Historically, horizons have offered an efficient way of characterising a profile by capturing the maximum variation within a soil profile using a minimum number of investigation sites. Horizons form through many factors including the

25   accumulation of OM, deposition of æolian or alluvial material, surface weathering, or translocation of clays or Fe/Al chelates (Isbell, 2002). They are identified in the field by observing changes of soil properties with depth. Common diagnostic criteria include colour, texture, mineral composition, structure, redoximorphic features and the presence of inclusions.

Many horizon diagnostic criteria such as colour, texture and mineral composition can be estimated using visible near-

30   infrared diffuse reflectance spectroscopy (VisNIR) (e.g. Viscarra Rossel *et al*., 2009). Previous studies have utilised this relationship to characterise horizons with VisNIR. Galvao *et al*. (1997) investigated VisNIR spectra of 35 air-dry and ground, horizon-based samples from six profiles in Brazil. The authors identified that the principal components of VisNIR spectra held intrinsic information that showed a characteristic decrease with depth. Viscarra Rossel and Webster (2011) analysed VisNIR spectra from 13,654 air-dried and ground samples from Australia. Horizon centroids in canonical space

35   were identified and by reallocating samples to the nearest centroid, it was possible to distinguish topsoil and subsoil horizons. Meanwhile, Fajardo *et al*., (2016) intensively sampled 59 air-dry soil cores at 2 cm increments up to 130 cm depth with a VisNIR contact probe. Principal components (PCs) of the spectra were subject to fuzzy clustering and a digital gradient was applied to identify spectrally derived horizon boundaries that exhibited similarity to traditional horizons.

Common themes identified from these studies are: i) the need to apply a dimension reduction technique before extracting useful information in the VisNIR spectra; and ii) the relative ease at discriminating between topsoil and subsoil groupings. Topsoil and subsoil discrimination was attributed to characteristic decreases in OM content, and corresponding increases in the influence of the mineral phase and clay content with depth, which were reflected in the spectra.

These three studies were all conducted on air-dry cores or ground samples. Limited attention has been given to identifying horizons *in situ* using VisNIR.

Collection of useful soil spectra *in situ* presents a number of challenging environmental factors compared to laboratory-based scanning. Incomplete control of soil moisture, temperature, surface condition and small scale heterogeneity all add complexity to the task of collecting useful spectra in the field. Variable moisture is of particular concern as it can modify a spectrum to such an extent that the variation between moisture contents can exceed variation between samples (Wijewardane *et al.*, 2016a). Moisture in a sample scatters and absorbs illumination radiation resulting in a general decrease in reflectance (Bowers and Hanks, 1965). Correcting for moisture effects is challenging, as observed decreases in reflectance are nonlinear (Lobell and Asner, 2002).

A number of methods have been put forward to deal with moisture effects when calibrating models to estimate soil properties from spectra collected in field condition:

1. **Spiking** a calibration set of ground samples with some field condition spectra, so that calibration algorithms are not over trained on moisture susceptible regions (Viscarra Rossel *et al.*, 2009; Guerrero *et al.*, 2010).
2. **Global moisture modelling** proceeds by first assigning samples into moisture content classes, and then applying individual calibration models for each moisture class (Mouazen *et al.*, 2006; Nocita *et al.*, 2013).
3. **Direct standardisation** (DS) uses a transfer matrix to convert spectra scanned under moist condition to the equivalent spectra scanned under air-dry conditions (Ji *et al.*, 2015).
4. **External parameter orthogonalisation** (EPO) projects both moist and air-dry spectra into a new space, orthogonal to the influence of soil moisture (Minasny *et al.*, 2011).

Comparative analysis has confirmed that the four methods are viable approaches to reduce the confounding influence of moisture when applying calibrated models to field condition spectra (Ji *et al.*, 2015; Wijewardane *et al.*, 2016b). However, soil spectra also hold intrinsic information that may not be quantifiable in terms of a response variable, such as the identification of homogeneous spectral response zones for identification of soil horizons or mapping the distribution of soil units spatially. Of the four methods mentioned above, spiking and global moisture modelling are unsuitable for investigating intrinsic spectral information as they only manage, but do not remove, the moisture effect. On the other hand, EPO and DS show potential as they remove the effect of soil moisture, while conserving information held within a spectrum.

External parameter orthogonalisation was developed to separate the effects of an unwanted external parameter from spectra containing useful information. Roger *et al.* (2003) first applied EPO to reduce the effect of temperature variation on VisNIR spectra to improve estimates of the sugar content of intact apples. Applications to removing the effect of moisture on soil samples were first demonstrated by Minasny *et al.* (2011), who successfully estimated SOC levels from rewetted samples. It has since been used successfully to improve estimates of inorganic carbon, total carbon, sand and clay content (Ge *et al.*, 2014; Ackerson *et al.*, 2015; Wijewardane *et al.*, 2016a).

Direct standardisation was developed to allow transfer of calibrated models developed on one spectrometer to be used on another spectrometer (Wang *et al*., 1991). The approach establishes a relationship between the spectra obtained by the 'master' spectrometer and the corresponding spectra obtained by the 'slave' spectrometer; the relationship is then used to transform the slave spectra to correspond with the master spectra. It has been adapted to removing the effects of soil

85   moisture where the moist spectra act as the slave set and are converted to the air-dry master set (Ji *et al*., 2015). Both spectra are collected with the same spectrometer in this case.

It remains unclear if moisture corrected spectra either by EPO and DS can conserve sufficient intrinsic soil information for the identification of homogeneous spectral response zones under field conditions. This study aims to evaluate EPO and DS

90   in terms of their ability to conserve relationships between VisNIR spectra obtained *in situ*, and those obtained under laboratory conditions for the identification of homogeneous spectral response zones.

2. Methods

2.1 Site description

The study site was located on Westwood Farm, an experimental property 3 km northwest of the township of Cobbitty, NSW,

95   Australia (33°59'44.9"S 150°39'11.9"E) (Fig. 1). The parent material of the site is Ashfield Shale, a Triassic sedimentary rock comprising black mudstones and grey shales (Howard, 1969). Soils developing from this parent material are known to have a well-developed texture profile and the marine nature of the parent material commonly results in expression of sodicity in the subsoil. The mineralogy of the clay fraction of this soil is dominated by kaolinite, producing soils of low to moderate fertility (Davey *et al*., 1975). The site has been extensively cleared for agricultural purposes and is currently used for

100   intensive grazing on naturalised kikuyu (*Pennisetum clandestinum*) and paspalum (*Paspalum dilatatum*) grasses.

2.2 Profile preparation

A pit was excavated 1.5 m wide, 5 m long and reaching a depth of 1.5 m at the centre. Four horizons were identified and the soil was classified as a Brown Kurosol in the Australian Soil Classification (Isbell, 2002), or an Abruptic Lixisol using the World Reference Base. Notable features of the soil include an abrupt textural contrast from sandy clay loam in the E horizon

105   to medium heavy clay in the Bt1 horizon (Table 1). The Bt1 horizon was also found to be strongly acidic, pH (1:5 H2O) < 5.5 (Hazelton and Murphy, 2016). A small quantity of magnetic gravel (~2-4 mm diameter) was found in the A and E horizons, and heavy mottling occurs in the Bt2. Horizon based sampling and laboratory analysis was conducted to further characterise the soil, including a surface sample taken at 0-2 cm depth (Table 2)

110   A 1 m x 1 m sampling region was delineated on the pit wall and sheared to a smooth surface (Fig. 2). The final shearing was conducted in a horizontal direction, progressing from the soil surface to the bottom of the sampling region to limit surface contamination from falling debris. Galvanised nails were inserted on a 10 cm grid to guide sampling.

2.3 *In situ* scanning and sample collection

Visible near-infrared (VisNIR) spectra were obtained with an AgriSpecTM device, connected via fibre-optic cable to a

115   contact probe attachment (Analytical Spectral Devices, Boulder, Colorado, USA). A Spectralon® tile (Labsphere Inc., North Sutton, New Hampshire, USA) was used to take a baseline reading every 15-20 measurements. Indico® Pro software was used to interface with the spectrometer; spectra were exported at 1 nm resolution covering the 350 - 2,500 nm wavelength range. Visible near-infrared readings were taken in 2.5 cm increments to give 41 readings over each 1 m transect. Eleven

vertical transects were taken at 10 cm lateral spacing, as well as three horizontal transects at 0, 50 and 100 cm depth (Fig. 3).

120   Bulk density cores were extracted on a 10 cm grid for further scanning under laboratory conditions (Fig. 3). Bulk density cores at 0 cm depth were taken perpendicular to the soil surface, i.e. driven into the soil surface. Those taken at depth were taken parallel to the soil surface, i.e. driven into the pit wall. Bulk density cores were immediately placed in aluminium tins and sealed with vinyl tape to conserve field moist condition.

2.4 Constructing the projection matrices

125   The DS transfer matrix was constructed as per Wang *et al*., (1991) and the EPO projection matrix following Minasny *et al*., (2012). A single library was used to construct both EPO and DS transformation matrices. Structural differences in the EPO and DS matrices are immediately evident (Fig. 4). However, features around the 1,400 and 1,900 nm water absorption bands can be identified in both.

2.5 Data processing

130   All data processing and analysis was performed in the R environment for statistical computing (R Core Team, 2016).

2.5.1 Spectral pre-processing

Some discontinuities were observed at the site of the VisNIR detector junctions, i.e. 1,000 and 1,800 nm. The *spliceCorrection()* function from the "prospector" package was employed to remove these artefacts (Stevens and Ramirez-Lopez, 2013). This process corrects for the offset of VNIR and SWIR2 and applies linear interpolation at the edges to create

135   a smooth junction with the SWIR1 range. Spectra were then trimmed to remove areas at the end of the detector range with low signal to noise ratios, leaving the 500 – 2,450 nm wavelength range. Reflectance readings were converted to absorbance using, $A = \log(1/R)$. Data was compressed by a factor of two through the dropping of alternate wavelengths. Compressing data reduces calculation time, without affecting model performance, as much of the data is highly correlated. Spectra were then smoothed using a Savitzky-Golay filter with a window size of 11 and a second order polynomial (Savitzky and Golay,

140   1964).

2.5.2 Principal component analysis

Principal component analysis (PCA) is a statistical procedure commonly utilised as a dimensionality reducing technique when processing VisNIR spectra. Data are subjected to a number of orthogonal projections, each accounting for the maximum variability remaining in the dataset. The effectiveness of PCA attributed to autocorrelation between wavelengths

145   in VisNIR spectra, which can be reduced so that a small number of variables explain the vast majority of observed variance. The *in situ* VisNIR dataset was used to build the principal component (PC) space. The PP, DS and EPO spectral datasets were individually centred and scaled to a mean of zero and unit variance and PCA performed. The centring and scaling parameters, as well as the loadings of the PCs, were then used to project laboratory-based VisNIR scans under field moist and air-dry condition onto the same PC space for comparison.

150   #### 2.5.3 *k*-means clustering

k-means clustering is an iterative process which partitions observations into clusters based on minimum distance from cluster centroids. Following partitioning new cluster centroids are calculated and observations are repartitioned to the new centroids. The algorithm proceeds until an error function is minimised (Eq. 1), so as to minimise the within cluster variance (MacQueen, 1967).

155

Equation 1

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2$$

Where:

160 Given a set of observations $(x_1, x_2, \ldots, x_n)$ where each observation is a d-dimensional vector, k-means clustering partitions the n observations into k sets $S = \{S_1, S_2, \ldots, S_k\}$ so as to minimise the within-cluster sum of squares.

The PCs of *in situ* scans were subject to k-means clustering to identify zones of homogeneous spectral response. To standardise the analysis the number of clusters was set equal to four, i.e. the same number of soil horizons observed.

165 Methods are available to automate the selection of cluster number if the number of required clusters is unknown, e.g. cubic clustering criterion (Sarle, 1983). The PCs of moist and air-dry laboratory scans were fit to the cluster centroids established from *in situ* scans.

2.5.4 Evaluation

170 Differences between PP, DS and EPO spectra under field moist and air-dry condition were assessed by calculating Lin's concordance correlation coefficient (LCCC) (Eq. 2) and root mean square error (RMSE) (Eq. 3) of PC scores projected into the PC space of *in situ* scans.

Equation 2

175
$$LCCC = \frac{2s_{xy}}{s_x^2 + s_y^2 + (\bar{x} - \bar{y})}$$

Where:

s_{xy} is the covariance between the two variables

s_x^2, s_y^2 are the variance of each variable

$\bar{x}, \bar{y}$ are the mean of each variable

180

Equation 3

$$RMSE = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}}$$

Where:

y_i is the ith observed value

185 $\hat{y}_i$ is the ith predicted value

n is the number of observations

Qualitative assessment of homogeneous spectral response zones was provided by comparison of the distribution of classes on the profile with field observed horizons.

190

3 Results and discussion

3.1 Moisture variability of the soil profile

Soil moisture content varied vertically and horizontally within the profile (Fig. 5). A local maximum was seen at the soil surface before decreasing to 20 cm depth then increasing again in the clayey subsoil. A maximum of 0.25 was observed at 50 cm depth and a minimum of 0.10 was observed at 20 cm depth. This large vertical distribution of moisture variability in the vertical sense was also met with significant variability laterally. At 100 cm depth the moisture content ranged from 0.17 to 0.24. The observed vertical and lateral variability in moisture content reinforces the need to remove the effects of moisture to gain useful insights from the spectra.

3.2 Moisture and treatment effects on spectra

Spectra taken on field condition cores had a lower reflectance than those taken on air-dry or ground samples, as noted previously (Bowers and Hanks, 1965) (Fig. 6a). The effect was nonlinear, an increased reduction was generally observed with increasing wavelengths, and two broad absorption bands were observed at 1400 nm and 1900 nm representing overtones of the fundamental vibrations of water molecules (Stoner and Baumgardner, 1980). Pre-processing only had little effect on removing the influence of variable soil moisture (Fig. 6b). Although spectra did converge around the mineral peak at 2200 nm, large differences were still observed, specifically at the broad 1400 nm and 1900 nm water absorbance peaks. Direct standardisation reduced the influence of soil moisture (Fig. 6c). For topsoil samples, DS resulted in near perfect alignment of moist samples. However, for subsoil samples DS worked best between 800 – 1850 nm with divergence observed in the visible section and also at wavelengths greater than 1850 nm. External parameter orthogonalisation produced a high degree of convergence between air-dry and field moist samples in the orthogonal space (Fig. 6d). Differences between the topsoil and subsoil samples also appeared to be accentuated.

3.3 Transfer to principal component space

The first PC explained 0.69 of the variance for PP, 0.63 for DS and 0.79 for EPO respectively. The cumulative proportion of variance explained by the first four PCs was greater than 0.97 in each instance. All treatments showed a high correlation between scans taken *in situ* and those taken in the laboratory under field moist condition, indicating that soil moisture was effectively conserved and that field moist results can be extrapolated to *in situ* readings (Fig. 7).

The effect of moisture can be seen by comparing PC scores of samples scanned in the laboratory under moist and air-dry condition. Deviations between moist and air-dry PP spectra occurred for all of PC1 and in the subsoil and topsoil for PC2 and PC3 respectively. Direct standardisation exhibited a strong coherence for PC1 and PC2; however, deviations occurred for PC3 and PC4 in the subsoil. Following EPO, there was a strong coherence throughout all four PCs and the PC scores also exhibited stronger vertical differentiation than was seen under PP and DS. Comparison of the first four PC scores for moist and air-dry scans shows that EPO (LCCC = 0.84, RMSE = 9.6) conserved more intrinsic information than DS (LCCC = 0.58, RMSE = 22.3) and PP (LCCC = 0.44, RMSE = 37.0) (Fig. 8).

3.4 Comparison of clusters to observed horizons

The four classes identified from clustering the PP PC scores were only able to effectively identify the A horizon *in situ*, as other horizons showed no continuous vertical disaggregation (Fig. 9). Under moist conditions, PP effectively isolated the A and E horizon from the B horizons. However, under air-dry conditions only one horizon was identified, indicating that the

spectra of air-dry subsoil samples are more similar to moist topsoil samples. This reaffirms the notion that variation in moisture can greatly exceed variation between samples (Wijewardane *et al*., 2016a).

230

Direct standardisation effectively identified three horizons *in situ*, despite the A and E horizons being combined. Under field moist condition in the laboratory, the separation of the two B horizons is less clear and is completely removed under ground condition; where DS could only effectively identify two horizons, with the E horizon split in half.

235 External parameter orthogonalisation was the most effective approach for identify horizons *in situ*, and for conserving class allocations under variable moisture conditions. Continuous horizontal bands, that resembled the field observed horizons were identified under all scanning environments.

The success of horizon identification by *k*-means clustering of VisNIR spectra is attributable to horizon delineation being
240 derived by strong changes in colour and clay content in this soil, not properties less spectrally active, such as structure. Organic carbon ranged from 3.39 g dag^{-1} for between the soil surface and 0.37 g dag^{-1} in the Bt2 horizon, while clay ranged from 16 to 67 g dag^{-1} (Table 2).

Assigned classes often did not translate to contiguous, mutually exclusive zones on the soil profile. Associations of classes
245 were observed, especially in the heavily mottled Bt2 horizon. Within horizon variation is expected, as horizons are never uniform. Horizons may represent gradational zones between two more clearly identifiable horizons, as distinguished with transitional AB and BC horizons. Alternatively, discrete parts of one horizon may be present in another, as represented by broken horizons A/B and B/C. In addition, VisNIR is capable of identifying horizons not identified through field observations (Fajardo *et al*., 2016).
250

The preservation of the spatial variability of horizons when captured in this way will no doubt provide insight into the development and functioning of soils. As opposed to the homogenisation that occurs when soils are ground and sieved prior to analysis. The benefits of this spatial disaggregation warrant further investigation.

3.5 Evaluation of DS and EPO

255 Direct standardisation produced variable results for the profile wall under the observed moisture contents. Slight improvements in the prediction accuracy of models calibrated following DS have been found when the moisture content of the training set is similar to the moisture content of the unknown sample (Wijewardane *et al*., 2016b). This moisture-explicit DS adds complexity to the moisture correction process. To apply the correct DS transfer set a priori knowledge of the sample's moisture content is required. Any method to ascertain soil moisture that requires drying a sample fundamentally
260 renders the correction processes redundant, as the dried sample could instead be scanned, and it is also impractical *in situ*. One approach is to predict the soil moisture content directly from the VisNIR spectra. Haubock *et al*., (2008) found that soil moisture could be predicted, $R^2 = 0.71$, with a normalised soil moisture index utilising just the 1,800 and 2,119 nm wavelength channels. However, using this approach could lead to compounding errors when a sample is placed in the wrong moisture class. If creation of moisture classes were to be applied to this soil profile, three different calibration models would
265 be required in total, and two would be required within the majority of lateral transects. It remains unclear if underlying homogeneous spectral response zones would be retained or if they would become a reflection of predicted moisture content and the subsequent transfer matrix applied.

270 As both moist and air-dry spectra are projected into the same space when applying EPO, a priori knowledge of soil moisture content is not required. EPO was more effective under the variable soil moisture levels seen in this soil profile and as expected when surveying a larger area for delineation of soil map units. It is thus seen as a more effective approach.

4 Conclusions

Both EPO and DS were able to reduce the negative effects of soil moisture on VisNIR spectra, whilst retaining useful spectral information. More intrinsic soil information was retained following EPO, as opposed to DS, and k-means classes
275 consistent with field observed horizons were better expressed under field moist and air-dry condition. The approach can be easily upscaled to mapping soil units spatially.

280 **References**

Ackerson, J. P., Demattê, J. A. M., Morgan, C. L. S. (2015). Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. Geoderma 259-260, 196-204.

Bowers, S.A. & Hanks, R.J. (1965). Reflection of radiant energy from soils. Soil Science, 100, 130–138.

Davey, B. G., Russell, J. D., Wilson, M. J. (1975). Iron oxide and clay minerals and their relation to colours of red and
285 yellow podzolic soils near Sydney, Australia. Geoderma 14(2), 125-138.

Fajardo, M., McBratney, A., Whelan, B. (2016). Fuzzy clustering of Vis–NIR spectra for the objective recognition of soil morphological horizons in soil profiles. Geoderma 263, 244-253.

Galvao, L. S., Vitorello, Í., Formaggio, A. R. (1997). Relationships of spectral reflectance and color among surface and subsurface horizons of tropical soil profiles. Remote Sensing of Environment 61(1), 24-33.

290 Ge, Y., Morgan, C. L., Ackerson, J. P. (2014). VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. Geoderma 221, 61-69.

Guerrero, C., Zornoza, R., Gómez, I., Mataix-Beneyto, J. (2010). Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. Geoderma 158(1), 66-77.

Hazelton, P., Murphy, B. (2016). Interpreting soil test results: What do all the numbers mean? CSIRO publishing, Clayton
295 South, Australia.

Howard, G. (November 1969). The Geology of New South Wales. Sydney: Geological Society of Australia. pp. 417–421. Packham (Ed.).

Isbell, R., 2002, The Australian Soil Classification, Revised Edition. Australian Soil and Land Survey Handbooks Series 4, CSIRO Publishing, Clayton.

300 Ji, W., Viscarra Rossel, R., Shi, Z. (2015). Accounting for the effects of water and the environment on proximally sensed vis–NIR soil spectra and their calibrations. European Journal of Soil Science 66(3), 555-565.

Lobell, D. B., Asner, G. P. (2002). Moisture effects on soil reflectance. Soil Science Society of America Journal 66(3), 722-727.

Mouazen, A. M., Karoui, R., De Baerdemaeker, J., Ramon, H. (2006). Characterization of soil water content using measured
305 visible and near infrared spectra. Soil Science Society of America Journal 70(4), 1295-1302.

Nocita, M., Stevens, A., Noon, C., van Wesemael, B. (2013). Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. Geoderma 199, 37-42.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

310 Roger, J.-M., Chauchard, F., Bellon-Maurel, V. (2003). EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. Chemometrics and Intelligent Laboratory Systems 66(2), 191-204.

Sarle, W.S. Cubic clustering criterion SAS Instit. Inc., Cary, NC, SAS Tech. Rep. A-108, 1983.

Savitzky, A., Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. Analytical

315 Chemistry 36(8), 1627–1639.

Stevens, A., Ramirez-Lopez, L. (2013). An introduction to the prospectr package. R package Vignette R package version 0.1.3.

Viscarra Rossel, R. A., Cattle, S. R., Ortega, A., Fouad, Y. (2009). In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma 150(3), 253-266.

320 Viscarra Rossel, R. A., Webster, R. (2011). Discrimination of Australian soil horizons and classes from their visible–near infrared spectra. European Journal of Soil Science 62(4), 637-647.

Wang, Y., Veltkamp, D. J., Kowalski, B. R. (1991). Multivariate instrument standardization. Analytical chemistry 63(23), 2750-2756.

Wijewardane, N. K., Ge, Y., Morgan, C. L. S. (2016a). Moisture insensitive prediction of soil properties from VNIR

325 reflectance spectra based on external parameter orthogonalization. Geoderma 267, 92-101.

Wijewardane, N. K., Ge, Y., Morgan, C. L. S. (2016b). Prediction of soil organic and inorganic carbon at different moisture contents with air-dry ground VNIR: a comparative study of different approaches. European Journal of Soil Science 67(5), 605-615.

**Table 1 Horizon-based field observations of a 1 m x 1 m surface of the soil profile.**

Horizon	Depth (cm)	Transition	Boundary shape	Munsell (moist)	Texture[†]	Structure	Mottling and inclusions
A	0-16	abrupt	even	10YR 3/4	L	Moderate Granular	Magnetite/Maghemite
E	16-33	clear	wavy	7.5YR 4/4	SCL	Strong subangular blocky	Magnetite/Maghemite
Bt1	33-47	gradual	wavy	7.5YR 4/6	MHC	Strong angular blocky	-
Bt2	47-100+	-	-	10YR 4/6	HC	Strong angular blocky	20% red, 20% grey mottles

[†] L – loam; SCL – sandy clay loam; MHC – medium-heavy clay; HC – heavy clay.

Table 2 Laboratory analysis of horizon-based bulk soil properties, including a 0-2 cm topsoil sample.

Horizon	Depth (cm)	OC	TN	Clay	Silt	Sand	pHw	pHc	EC	Exch. Cations (cmol(+) kg^{-1}) Ca	K	Mg	Na	Al	CEC	ESP
-	0-2	3.39	0.24	16	20	64	5.5	4.8	0.21	3.1	1.0	2.9	0.1	0.1	7.2	1.2
A	0-16	0.96	0.04	18	14	68	6.0	4.8	0.06	1.2	0.5	1.1	0.1	0.2	3.1	3.0
E	16-33	0.67	0.03	24	12	64	5.7	4.6	0.06	0.9	0.3	2.0	0.2	0.5	3.9	4.3
Bt1	33-47	0.81	0.04	57	10	32	5.4	4.4	0.05	1.5	0.4	5.2	0.4	1.3	8.9	4.6
Bt2	47-100+	0.37	0.00	67	11	22	5.3	4.0	0.07	0.2	0.2	5.1	0.8	7.2	13.5	5.8

OC – organic carbon (g dag); TN – total nitrogen (g dag); pHw – pH (1:5 H2O); pHc – pH (1:5 CaCl2); EC – electrical conductivity (dS m^{-1}); CEC – cation exchange capacity (cmol(+) kg^{-1}); ESP – exchangeable sodium percentage (%).
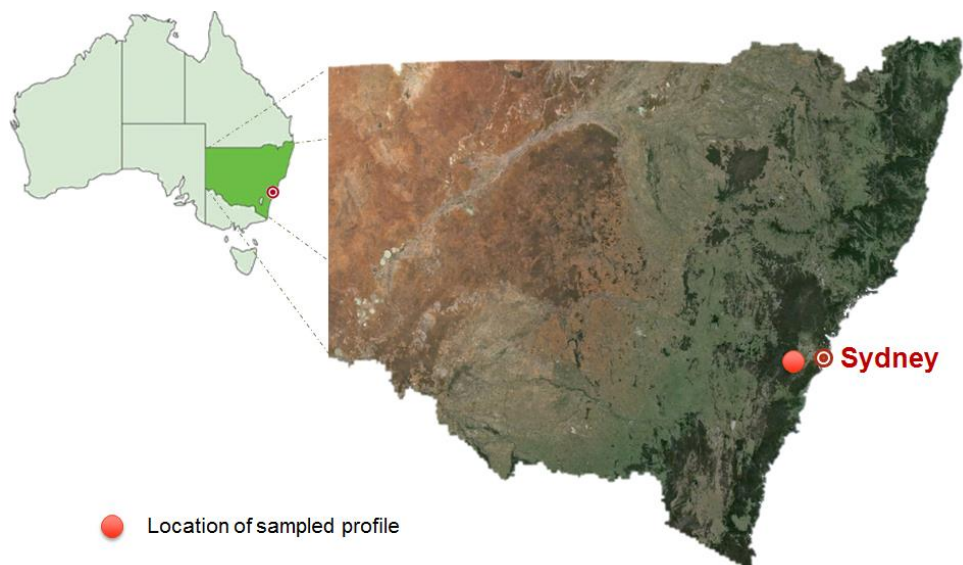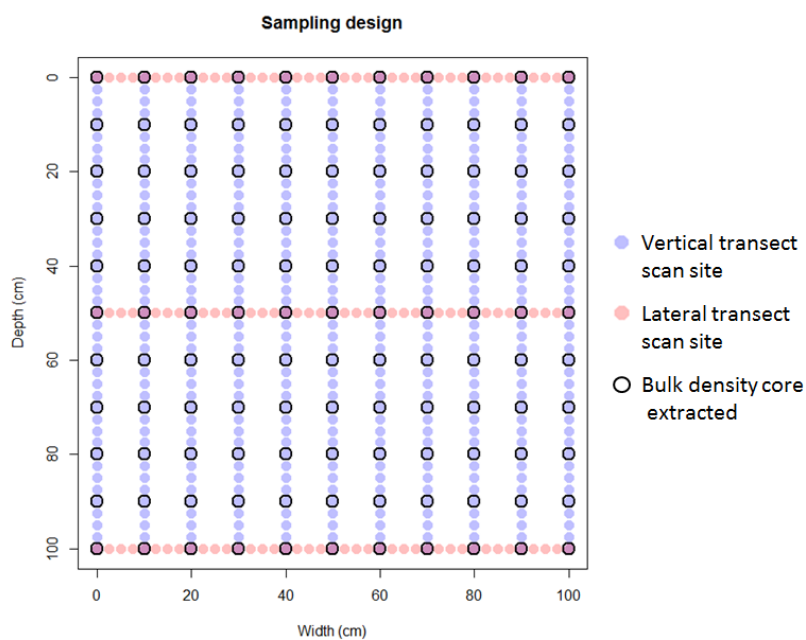
Figure 1: Location of the sampled profile in relation to Sydney within the state of New South Wales, Australia.
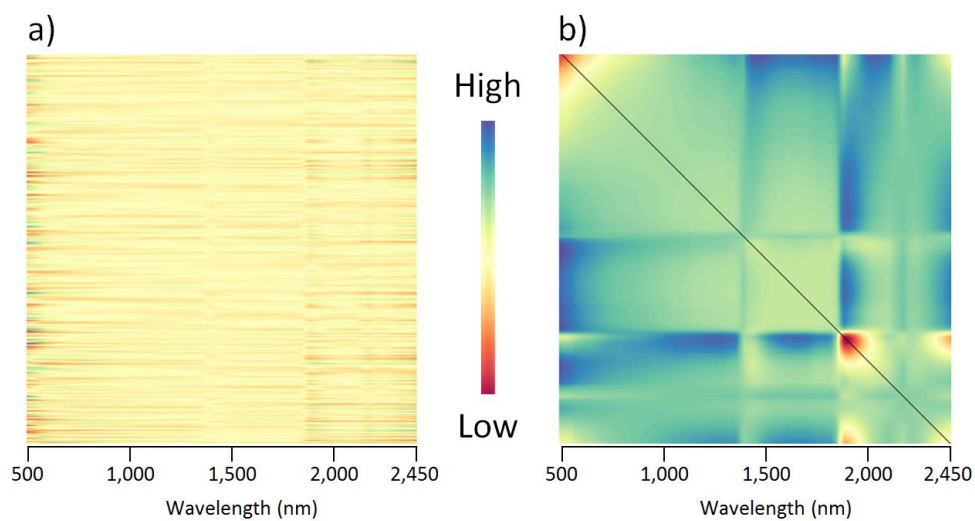
5

Figure 2: Photograph of the prepared soil profile displaying a natural face section (left) and the prepared 1 m x 1 m sampling area. Galvanised nails were inserted on a 10 cm grid to guide sampling locations.

Figure 3: Schematic representation of sampling design of the soil profile.

Figure 4: Comparison of matrix structure: a) direct standardisation transfer matrix; and b) external parameter orthogonalisation projection matrix.
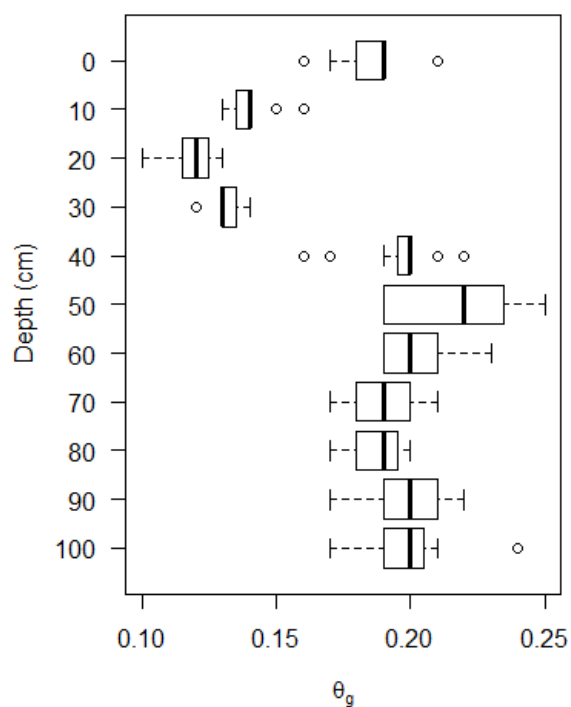
**Figure 5: Box plots displaying the distribution of gravimetric moisture content by depth.**
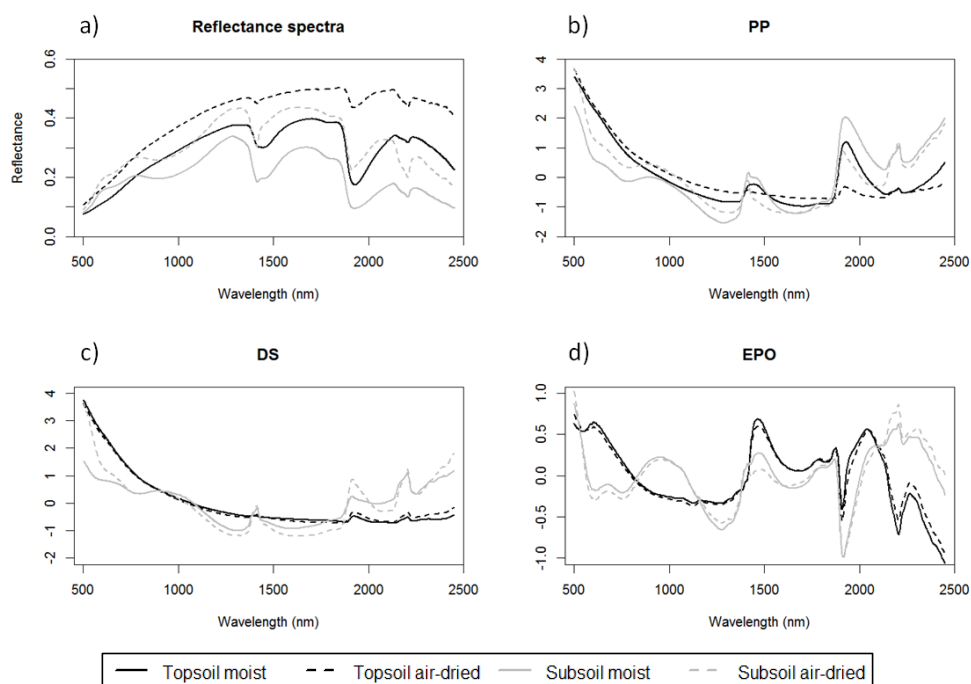
Figure 6: Comparison of a representative topsoil and subsoil sample scanned in field moist and air-dried condition: a) trimmed and splice corrected reflectance spectra (500-2450 nm); b) pre-processed (PP) spectra; c) direct standardisation (DS) approach whereby the moist sample is corrected to resemble the air-dry sample; d) external parameter orthogonalisation (EPO) spectra whereby both moist and air-dry spectra are projected into a new space orthogonal to moisture effects.
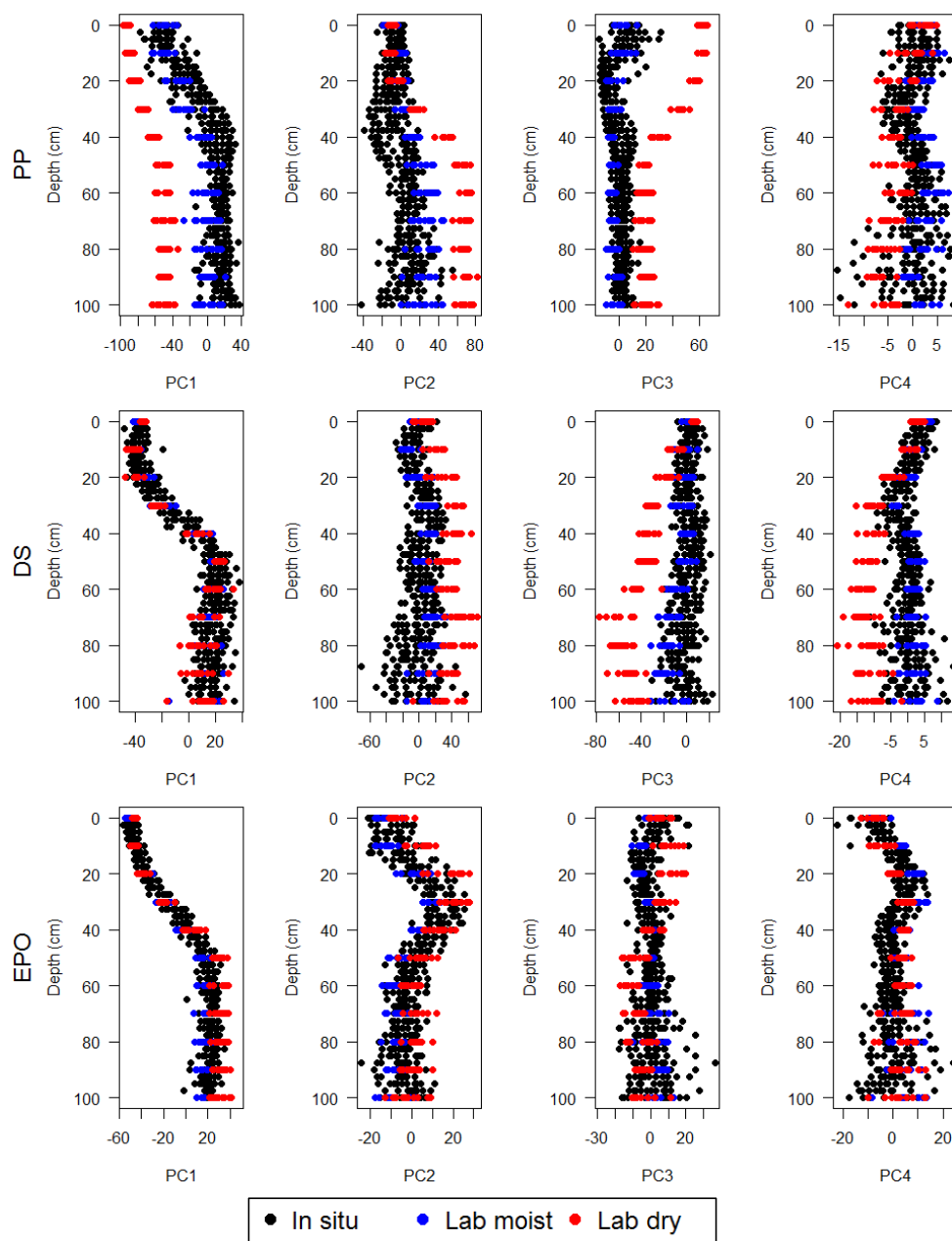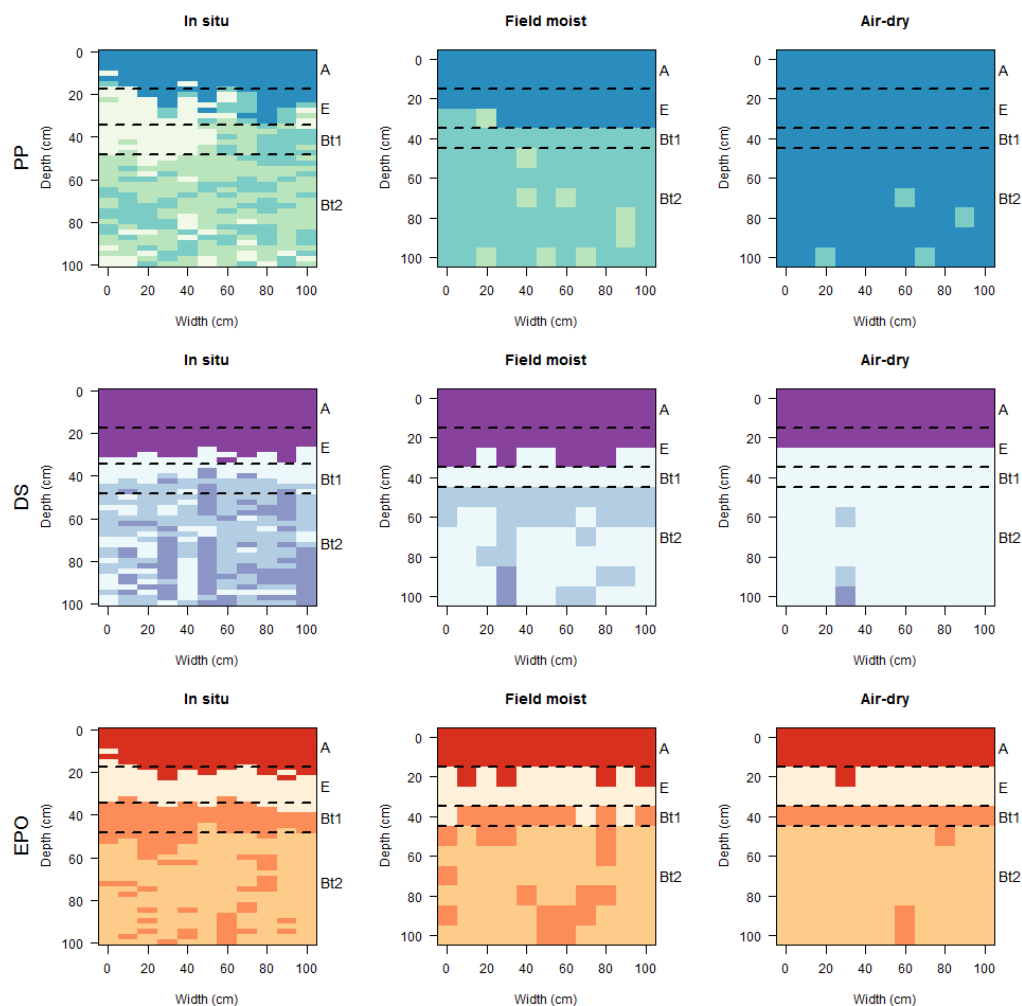
Figure 7: Principal component scores for VisNIR spectra obtained *in situ* (black), field moist in the laboratory (blue) and air-dry in the laboratory (red).

5
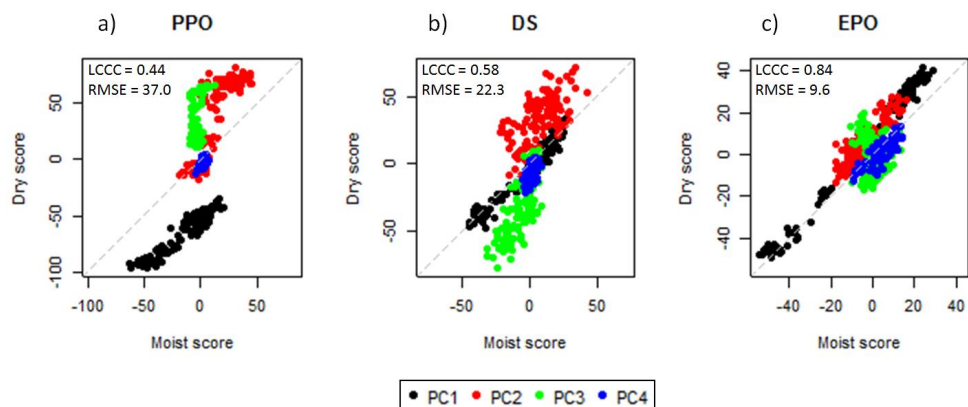
Figure 8: The first four PCs of VisNIR spectra under moist and air-dry condition: a) pre-processing only; b) direct standardisation; c) external parameter orthogonalisation.

5

Figure 9: The distribution of classes identified by k-means clustering of the first four PC scores of spectra following: i) PP – pro-processed only spectra; ii) DS – direct standardisation; and EPO – external parameter orthogonalisation. Field observed horizon boundaries are indicated by dashed horizontal lines. Horizon designations are indicated.