Mapping the information content of Australian visible-near infrared soil spectra

R.A. Viscarra Rossel[1*], A. Chappell[1], P. de Caritat[2], and N.J. McKenzie[1]

[1] CSIRO Land and Water, Soil and Landscape Science, GPO Box 1666 Canberra ACT 2601, Australia
[2] Geoscience Australia, GPO Box 378, Canberra ACT 2601, Australia

Abstract
We describe the information content of soil visible–near infrared (vis–NIR) reflectance spectra and map their spatial distribution across Australia. The spectra of 4030 surface soil sample from across the country were compressed using a principal component analysis (PCA) and the resulting scores were mapped by ordinary point kriging. The largely dominant and common feature in the maps was the difference between the more expansive, older and more weathered landscapes in the centre and west of Australia and the generally younger, more complex landscapes in the east. A surface soil class map derived from the clustering of the principal components was similar to an existing soil classification map. Visible–NIR reflectance spectra provide an integrative measure to rapidly and efficiently measure the constituents of the soil. It can replace the use of traditional soil properties to describe the soil and make geomorphological interpretations of its spatial distribution and therefore it can be used to classify soil objectively.

Key Words
Reflectance spectra, visible–near infrared, kriging, Australian soil.

Introduction
Recent technological advances in computing and measurement technologies are providing soil scientists with tools that can be used to improve the efficiency of soil measurements. Over approximately the last three decades and across diverse fields of research from remote sensing to chemometrics, interest has developed in the rapid, non-destructive measurements of the intrinsic optical properties of the soil in the visible–near infrared (vis–NIR) and mid-infrared (mid-IR) ranges of the electromagnetic spectrum between 400 and 2500 nm and 2500 nm and 25,000 nm, respectively (e.g. Ben-Dor *et al*. 1999). Although mid-IR soil spectra has a stronger signal, the advantages of the vis-NIR are: the analysis requires less sample preparation, the instrumentation is portable and can be easily used in the field, direct links can be made with hyperspectral remote sensing (e.g. Gomez *et al*. 2008) and iron oxides, clay minerals and soil colour can be easily measured directly from the spectra (Viscarra Rossel *et al*. 2009).

Field and laboratory experiments have shown that wavelength-specific absorptions of electromagnetic radiation in the vis–NIR provide diagnostic reflectance spectra for the chemical, physical and mineralogical composition of the soil. For example, broad and shallow absorption bands in wavelengths smaller than 1000 nm can be due to iron oxides; narrow, well-defined absorptions near 1400 nm and 1900 nm can be related to hydroxyl and water; absorptions near 2200 nm to clay minerals; and various absorptions throughout the spectrum to organic matter.

We propose that spectra: (a) provide an integrative measure that can be used to rapidly and efficiently measure the constitution of the soil, and (b) spectra can replace the use of traditional soil properties for describing the soil and making geomorphological interpretations of its spatial distribution. To support our proposal, we describe the information content of Australian soil vis–NIR spectra and map its spatial distribution across Australia.

Methods
Soil samples
The soil samples used here originated from different sources, comprising continental, regional and field scale soil surveys. The samples were from: (i) the Australian Commonwealth Scientific and Industrial Research Organisation's (CSIRO) National Soil Archive, (ii) The National Geochemical Survey of Australia (NGSA), (iii) state department and regional surveys and (iv) a number smaller surveys conducted for research into precision agriculture. Their spatial distribution is shown in Figure 1. The samples cover the range of Australian soil orders (Isbell 1996) over the varying climatic regimes across Australia (Figure 1). All 4030 samples were air dry and approximately 100 g subsamples were passed through a 2 mm sieve for the spectroscopic analysis.

1 – 6 August 2010, Brisbane, Australia. Published on DVD.

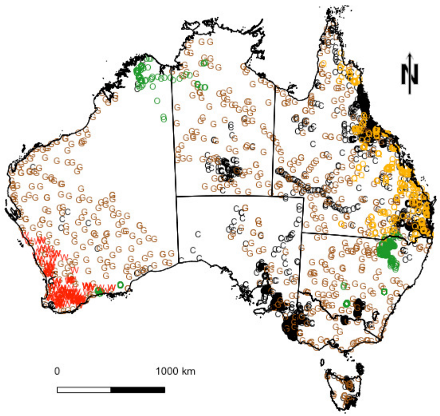

Figure 1. Spatial distribution of the data. The different symbols indicate their origin: CSIRO National Soil Archive (C), The National Geochemical Survey of Australia (NGSA) (G), state department surveys from Queensland (Q) and Wester Australia (WA) and other regional and field scale surveys (O).

Vis-NIR spectroscopy and spectroscopic analyses

The diffuse reflectance spectra of the 4030 soil samples were measured using the Labspec® vis-NIR spectrometer (Analytical Spectral Devices, Bolder, Colorado, USA) with a spectral range of 350–2500 nm. The soils were measured using a contact probe (Analytical Spectral Devices, Bolder, Colorado, USA) and a Spectralon® panel was used to obtain a white reference once every 10 measurements. For each soil measurement, 30 spectra were averaged to improve the signal-to-noise ratio. Spectra were collected with a sampling resolution of 1 nm so that each spectrum was made up of 2151 wavelengths.

A continuum removal technique (Clark and Roush 1984) was used to isolate absorption features in the spectra. We calculated the continuum-removed (CR) spectrum by dividing the original reflectance values by the corresponding values of the continuum line.

Multivariate statistical analysis

The continuum-removed (CR) spectra were mean-centred and then analysed using principal component analysis (PCA). We used both the scores and loadings to interpret our data. To provide a more general description of the 4030 soil samples, the first four principal component scores (which accounted for 98% of the variance in the data) were clustered using the k-means algorithm. A scree plot of the percent variance explained by cluster was used to select the number of clusters. The spatial distribution of the clusters was plotted and the absorption features of the average spectrum of each of cluster were used to interpret the soil samples.

Geostatistical analyses and mapping

We computed experimental variograms for the scores of each of the first three principal components. We fitted, using weighted least squares, a number of different models to our data (spherical, exponential, Matern), including models that describe a nested combination of two individual variograms (Webster & Oliver 2001). The rationale for fitting nested models was that soil variability can often be described by a combination of different processes, one nested within another and acting at different spatial scales. We used the Akaike Information Criterion (AIC) (Akaike 1973) to select the best, most parsimonious model. The model with the best least squares fit to the experimental semi-variances of the scores, for the first three principal components, was a nested exponential model with a nugget effect. We used ordinary point kriging (Webster & Oliver 2001) to map the spatial distribution of each of the first three principal component scores across Australia on a 5 km square grid. The kriged principal component scores surfaces were clustered using the k-means technique. This resulted in a classed map of surface soil characteristics across Australia.

Results

Mapping the information content of soil vis-NIR spectra

The variograms of the principal components are shown in Figure 2. The semi-variances in the first principal component (PC) are larger than those in the other PCs and explain a larger proportion of the spatially dependent variation. However, the variograms of PC two and three also demonstrate good spatial dependence, the nested exponential models provide a good fit (Figure 2) and the loadings spectra show strong and clearly identifiable features (Figure 3) that contribute to the interpretation of Australian soil. Therefore we retained them in the analysis.

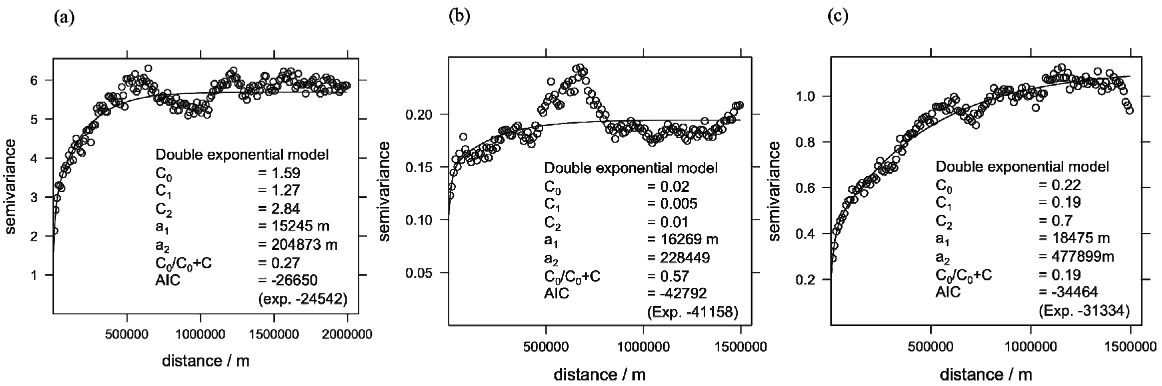

Figure 2. Experimental variograms and fitted double exponential models for (a) principal component 1, (b) principal component 2 and (c) principal component 3. Model parameters are given together with the Akaike Information Criterion (AIC).

The nested variation, represented by the different variogram ranges (Figure 2), indicates that the soil and processes that formed it vary over different spatial scales The maps produced by kriging the first three PCs and their kriging variance are shown in Figure 4.

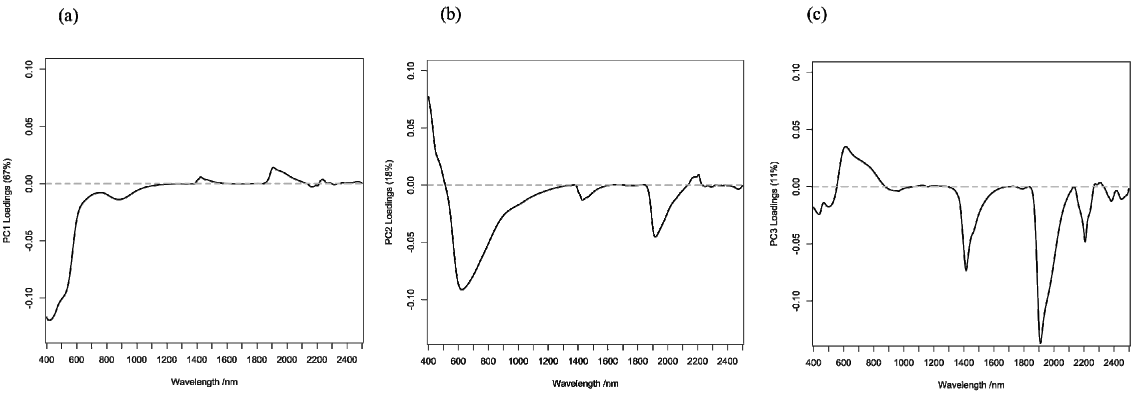

Figure 3. Loadings spectra of the first three principal components, which together account for 96% of the variability in the spectra

The largely dominant and common feature in the three PC maps is the difference between the more expansive, older and more weathered landscapes in the centre and west of the country and the generally younger, more complex landscapes in the east (Figure 4). The spatial distribution of the first PC (Figure 4a), whose (negative) loadings are dominated by haematite (Figure 3a), depict the highly weathered landscapes of the west, extending through to South Australia (SA), southern Northern Territory (NT), and parts of Queensland (Qld) and south western New South Wales (NSW). The short-range variation of the first PC is restricted to eastern Australia and south-western WA and appears to be associated with the clay mineralogy. The more dominant, longer-range component is mostly found in the centre and west of the country and is associated with the distribution of haematite in the more weathered landscapes (Figure 4a). The second and third PCs depict smaller and more subtle differences in the soil. The spatial distribution of the second PC (Figure 4b), whose negative loadings are dominated by absorptions that can be related to soil organic carbon and smectite (Figure 3b), depict the younger depositional landscapes of eastern Australia, where the shorter-range variation predominates. The negative loadings of the third PC map (Figure 4c) depict the spatial distribution of soils with abundant amounts of smectite. The positive loadings depict coastal soils with higher amounts of soil organic matter (Figure 4c).

Figure 5 shows the surface soil class map derived from the clustering of the three PC maps in Figure 4. Class one represents highly weathered soils that contain abundant amounts of haematite and kaolinite (Figure 5a). Class two represents coastal soils and the agricultural soils of the Australian wheat-belt. Class three soils are found largely inland of the previous class and extend throughout south-western Qld, north western NSW and Victoria (Vic) and north eastern SA. Class four soils are found across the northern part of the Northern Territory and Queensland and south-eastern NSW and south-western WA and appear to coincide with the wetter areas of Australia. The classed surface soil map (Figure 9a) shows that the information acquired from the spectra is similar to that depicted by soil classifications available from soil surveys (Figure 9b).

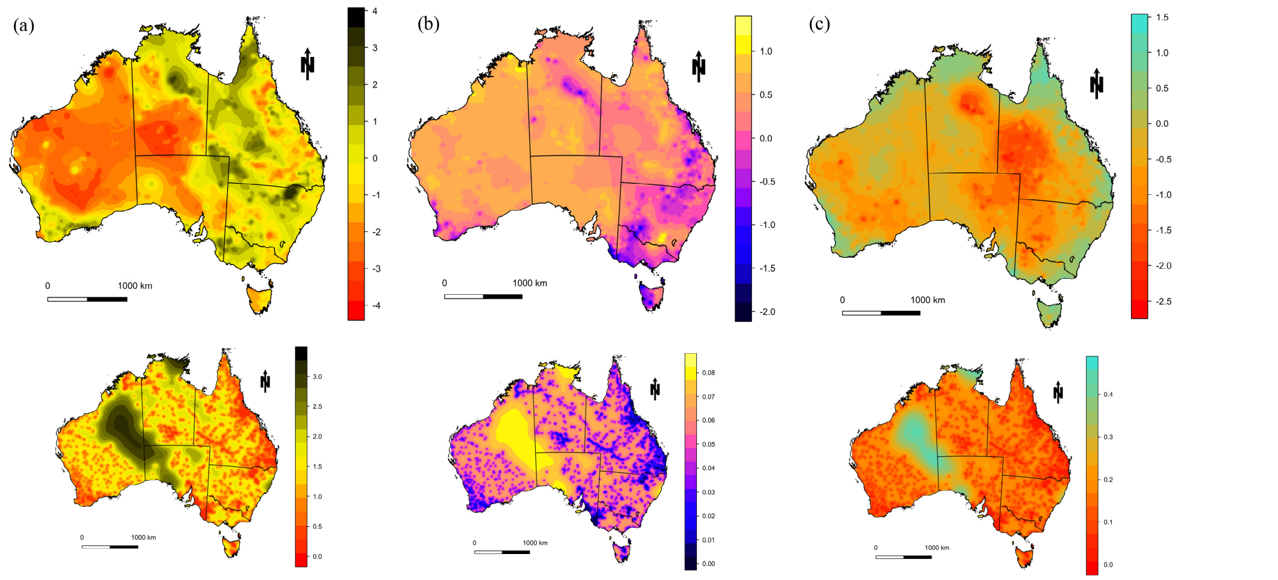

Figure 4. Mapping the information content of Australian soil vis–NIR spectra. Kriged maps (top row) and their variance (bottom row) for (a) principal component 1, (b) principal component 2 and (c) principal component 3.

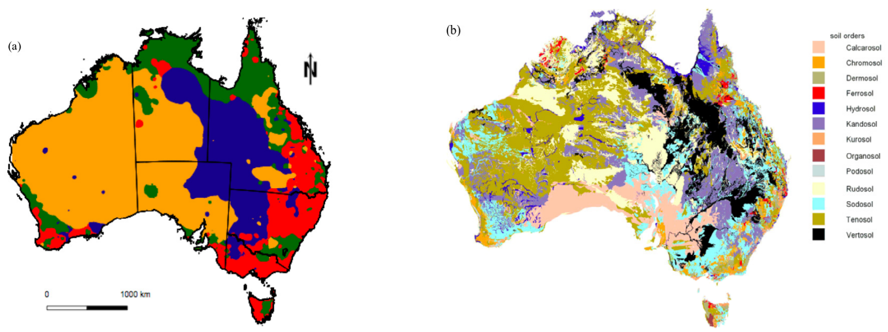

Figure 5. (a) Surface soil class map derived from the clustering of the principal component maps in Figure 4. Class 1 is represented by the orange, class 2 by the red, class 3 by the blue, and class 4 by the green. (b) The Australian Soil Classification map of soil orders (Isbell 1996).

Conclusions

In this paper we have shown that: (i) a vis–NIR diffuse reflectance spectrum is an integrative, multi-parameter measure of the soil that contains information on its fundamental constituents: mineral and organic matter and water, (ii) that principal component analysis combined with geostatistics can be used to describe the information content of the spectra and map soil and (iii) therefore vis–NIR spectra can be used for soil classification. We demonstrated this by mapping Australian soils.

References

Akaike H (1973) Information theory and an extension of maximum likelihood principle. In 'Second International Symposium on Information Theory'. (Eds BN Petrov, F Csáki, Akadémia Kiadó) pp. 267–281. (Budapest, Hungary).

Ben-Dor E, Irons JR, Epema G (1999) Soil reflectance. In 'Remote sensing for the earth sciences'. (Ed AN Renzc).pp. 111–1888. (Wiley, New York)

Clark RN, Roush TL(1984) Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. *Journal of Geophysical Research* **89**, 6329–6340.

Gomez C, Viscarra Rossel RA, McBratney AB (2008) Soil organic carbon prediction by hyperspectral remote sensing and field vis–NIR spectroscopy: an Australian case study. *Geoderma* **146**, 403–411.

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

Webster R, Oliver MA (2001) 'Geostatistics for Environmental Scientists'. (John Wiley, Chichester, U.K).