Proximal sensing in soil profiles

Edward James Jones

BScAgr (Hons and Medal) - The University of Sydney

A thesis submitted to fulfil requirements for

the degree of Doctor of Philosophy

Faculty of Science

The University of Sydney

New South Wales

Australia

MMXVII

STATEMENT OF ORIGINALITY

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Edward James Jones

In loving memory of Samuel Francis (Frank) Hulme,

for all that you taught me about agriculture, hard-work and perseverance.

SUMMARY

Objective and quantitative soil information is crucial for pedological investigations and to inform diverse decision-making processes. A wide range of proximal sensing devices are available to provide quantitative soil information *in situ*. The cost-effectiveness and timeliness of the devices allows a much greater representation of the spatial distribution of soil properties to be achieved and supports soil investigation at finer spatial and temporal resolution. To use these new devices effectively, new techniques are required. Focus must also be given to ensuring that the information received from devices is made available in the field, so that this soil information may be utilised to do something useful in the field.

Currently, no sensing device can provide information on all the soil properties of interest. Use of multiple sensors provides the only viable option for now. This thesis investigated the conjoint use of visible near-infrared diffuse reflectance spectroscopy (VisNIR) and portable X-ray fluorescence spectroscopy (pXRF) for the *in situ* investigation of soil properties and profile variability. These two devices operate on distinct principles, which are thought to be complementary. Portable X-ray fluorescence spectroscopy is able to provide elemental concentration in the sample ($Z \geq 12$). Meanwhile, VisNIR is able to provide information on colour and some molecular characterisation.

Calibrating models using VisNIR and pXRF is well established, although commonly models are developed for a limited number of properties and a limited geographic range. Therefore, this thesis was conducted with three aspirations: i) construct models that are functional for a diverse range of soil profiles; ii) construct models that work *in situ*; and iii) provide results in near real-time.

Chapter 1 provides a review of the development of quantitative, field portable soil sensors, and the new field of *Digital Soil Morphometrics* that has developed around these devices. The review raises many questions surrounding quantitative soil description and highlights the fact that new techniques are required to efficiently describe soils with these devices.

Chapter 2 describes the devices, methods and sites used in this thesis. A detailed description of the functionality of VisNIR and pXRF is provided, and current applications in soil science are summarised. Methods that are common to multiple chapters are described to avoid repetition in the following chapters. Site descriptions of the fifteen soil profiles analysed are provided, including field observations, laboratory analysis and sensor information.

Chapter 3 explores the intensive use of VisNIR on a soil profile for the identification of homogeneous spectral response zones. Effects of soil moisture on spectral responses are examined and methodologies to reduce these effects, while conserving intrinsic soil information, are explored.

Chapter 4 explores the development of an improved sampling methodology to efficiently capture soil profile variability. Given the accuracy for space trade-off associated with these devices, they cannot be used in a similar method to traditional horizon description. Vertical and lateral variation of soil profiles was characterised, and this information was used to inform a methodology for sampling profiles with proximal sensors.

Chapter 5 presents a data-fusion approach to characterise the mineral composition of soils, including phyllosilicate speciation, Fe-oxides, gypsum, carbonate, quartz and feldspars. This approach combines a pattern matching algorithm to predict phyllosilicates and Fe-oxides speciation from VisNIR spectra, and an elemental mass balance based on pXRF reported elemental concentrations.

Chapter 6 investigates the use of a spectral soil inference system (SPEC-SINFERS) to augment the number of predicted properties. As not all properties of interest have detectable spectral activity by either VisNIR or pXRF. This system involved the propagation of sensor and model uncertainties through one hundred independent simulations for each pedotransfer function and allowed the integration of both regression and machine learning models.

Chapter 7 summarises results from this thesis, discusses limitations and improvements, and suggests future research directions.

Published chapters of this thesis

Chapter 1

Jones, E.J., McBratney, A.B., 2016. What is digital soil morphometrics and where might it be going? In: Hartemink, A.E., Minasny, B. (Eds.), Digital Soil Morphometrics. Springer, Dordrecht.

Chapter 5 (in part)

Jones, E.J., McBratney, A.B., 2016. In-situ analysis of soil mineralogy through conjoint use of visible, near-infrared and X-ray fluorescence spectroscopy. In: Hartemink, A.E., Minasny, B. (Eds.), Digital Soil Morphometrics. Springer, Dordrecht.

Other publications and presentations made from this thesis

Abstracts

Jones, E.J., McBratney, A. (2015). *In situ* analysis of soil mineralogy through conjoint use of Vis-NIR and XRF spectroscopy. Digital Soil Morphometrics, University of Wisconsin-Madison, USA, 1st - 4th June, 2015.

Jones, E.J., Minasny, B., McBratney, A.B. (2016). *Predicting profile characteristics with proximal sensors and a spectral soil inference system (SPEC-SINFERS).* New Zealand Society of Soil Science and Soil Science Australia conference - Soil, a Balancing Act Downunder, Queenstown, New Zealand, 12th - 16th December, 2016.

Jones, E.J., Minasny, B., McBratney, A.B. (2017). *Estimating soil profile attributes with proximal sensors and a spectral inference system (SPEC-SINFERS).* Pedometrics, Wageningen, The Netherlands, 26th June – 1st July, 2017.

Jones, E.J., McBratney, A.B., Minasny, B., Cattle, S. (2017). *Soil property sensing with visible near-infrared and portable X-ray fluorescence spectroscopy*. 20th symposium on precision agriculture in Australasia, The University of Sydney, Australia, 14th – 15th August, 2017.

Poster

Jones, E.J. (2015). Proximal sensing in soil profiles. *On the pulse - Research Symposium 2016,* Faculty of Agriculture and Environment, The University of Sydney, Australia, 12th July 2016.

Publications in preparation from this thesis

Chapter 3

Jones, E.J., McBratney, A.B. (2018). Mapping homogeneous spectral response zones in a soil profile

Chapter 4

Jones, E.J., McBratney, A.B. (2018). Designing sampling to quantify soil profile variability

Chapter 5

Jones, E.J., Singh, B., McBratney, A.B. (2018). Checks and mass balances for *in situ* quantification of soil mineral composition

Chapter 6

Jones, E.J., Filippi, P., Minasny, B., McBratney, A.B. (2018). Estimating profile characteristics with proximal sensors and a spectral soil inference system

ACKNOWLEDGMENTS

I would like to thank my supervisors Alex McBratney and Stephen Cattle. For your direction, and your freedom in pursuing this research. For instilling the trust in me to pursue many extracurricular activities, that have also helped me to develop as a researcher.

I would like to thank the Grains Research and Development Corporation (GRDC) for provision of a Grains Industry Research Scholarship to support the project. Without this scholarship the research could not have been as productive.

Budi for your advice and expertise in relation to SPEC-SINFERS, and willingness to answer any soil or statistics questions that I have had. Balwant for advice on mineralogy and XRD that assisted in this research. I want to thank Chook with the shovel for cleaning the pit. Tom for additional statistical questions that I have had.

Uta and Brendan for your time spent advising and training me to using the sensors. Patrick for your help in the field and help in keeping me sane in the office. Jay, for chicken lol. To my first employee, Yumi, for all of your hard work in the laboratory.

For technical support I would like to thank Lori, Glen, Ali, Ivan, Iona and Michael. Things would never run as smoothly without all the great work that you do.

To landowners and property managers who allowed me to sample on their properties: Andrew Rawson, Mal Pritchard, and Ryan Hill and family. Especially, for those who have sacrificed their time in excavated soil pits. Thank you for your time, efforts and interest in the research.

Everyone else that I can think of in unbiased, alphabetical order: Brett, Derek, Kanika, Keith, Kip, Liana, Mahnaz, Mario, Maryam, Milena, Nathan, Phil, Sanjee, Vanessa and Wartini.

And to everyone that I have forgot.

Outside of the university, I would like to thank my mother, father, my direct and extended family, for always being supportive of my pursuits, whatever they may be.

Finally, to Julla, for your patience, for your support and inspiration. *Akbay*!

CONTENTS

1 WHAT IS DIGITAL SOIL MORPHOMETRICS AND WHERE MIGHT IT BE GOING?

1.1 Abstract

A large number of devices exist that are able to provide quantitative and objective representation of soil properties. Many of these devices are able to elucidate properties unattainable to the human eye and may redefine current definitions of "field observable" properties. Possible meanings for, and approaches to, digital soil morphometrics are discussed. Digital soil morphometrics' relationship to other domains of research and practice such as proximal soil sensing and conventional field soil description are explored; with the suggestion that digital soil morphometrics has greatest potential as a special case of proximal soil sensing. The application areas of digital soil morphometrics outside of routine soil description are canvassed and technological gaps are discussed.

1.2 Introduction

The development of morphometrics in the biological sciences enabled the quantitative analysis of form and revolutionised the description and statistical analysis of specimens. To bring a similar revolution to soil description and to unite diverse tools and techniques that are able to provide more objective and quantitative description of soil attributes the subdiscipline

of digital soil morphometrics (DSMorph) has been proposed (Hartemink and Minasny, 2014). In general terms, DSMorph has been defined as the "application of tools and techniques for measuring, mapping and quantifying soil profile attributes and deriving continuous depth functions" (Hartemink and Minasny, 2014). DSMorph is thus at the forefront of technology and innovation in soil science and promises to provide a much needed scientific and technological overhaul to field soil description; a discipline which has been relatively technology stable, possibly averse, for decades. DSMorph tools and techniques have untold pedological, edaphic, and environmental applications.

While mathematical approaches and the computing power to perform morphometric analyses have greatly advanced in recent years, there remain two fundamental issues encumbering the supply of relevant input data. First, labour-intensity, as examining and recording appropriate attributes can be a time-consuming task (Blackith and Rayment, 1971, p. 286). Secondly, subjectivity, as there are always elements to the observation of natural forms that may be considered more of an art than a science (Blackith and Rayment, 1971, p. 1). These issues underlie the development of DSMorph and highlight the value of any technique that may provide less labour-intensive data collection and increased objectivism of soil observations.

The marriage of digital data collection with morphometric approaches is crucial to the success of this new subdiscipline. However, key in recognising DSMorph as a subdiscipline is the identification of which particular attributes or contributions it may deliver that other subdisciplines cannot deliver independently. Digital soil morphometric approaches can better represent variation within observed soil objects, facilitate the quantification of uncertainties and will change the way soils are observed and described. Nonetheless, the offered definition is quite broad and its boundaries need to be better delineated to clearly define what DSMorph is, what it is not and where it might be going.

To better understand DSMorph its component topics of digital data acquisition, soil form and morphometric approaches are deconstructed, and the unique contribution of each is investigated. As innovation in the application of DSMorph tools and techniques as they pertain to routine soil description has recently been reviewed (Hartemink and Minasny,

2014), the second half of this chapter is focused on highlighting a selection of novel and potential applications of DSMorph techniques outside of routine soil description. Technological and information gaps are identified and ways forward discussed.

1.3 Discussion

1.3.1 Conventional field soil description

Before investigating DSMorph further, the need for its development must be understood. Most people would agree that field soil description has largely stagnated following rapid initial development. This development includes the formalisation of procedures for observing and classifying soils (Clarke, 1936; Soil Survey Staff, 1937), as well as the codification of morphological classification of soil structure (Nikiforoff, 1941). This was a significant step towards the standardised observation and reporting of soil attributes mandated by widespread soil survey programs beginning from the turn of the 20th century. Pedology has historically been a hands-on, technology sparse discipline. During this period of standardisation it was noted that the most important tool for the soil observer was the humble spade (Soil Survey Staff, 1937, p. 28). In fact, excluding a small number of qualitative analyses involving HCl or H_2O_2, and basic tests for soil reaction, there were "few reliable field tests of soils that serve a useful purpose in the soil survey" (Soil Survey Staff, 1937, p. 30-2). Fast-forward eight decades and little has changed. The field of soil description has remained largely technologically stable since the 1950s, and to this day the main *modus operandi* in soil description remains trained human observation. Stagnation in itself is not a cause for action; in fact it may signify the successful maturation of a field. However the reliance on human observation brings into question sources of error, as although professional observers may be trained and calibrated to one another, this training cannot completely eliminate inter- or intra-observer variation.

Soil colour, one of the most significant soil attributes, gives an example of how much variation in assessment can be introduced. In addition to differences in the light condition under which colour measurements are being made, it is also well established that physiological differences in the eye mean that not everyone perceives colour in the same

manner, and that colour perception also drifts with aging due to the gradual yellowing of the eye's lens (Billmeyer and Saltzman, 1981, p. 174). The use of reference charts is meant to eliminate the effects of these differences and standardise the reporting of soil colour. However, the Munsell chart is not free from variation. Sánchez-Marañón *et al.* (2005) investigated these ubiquitous colour books and found that manufacturer production differences and non-uniform fading characteristics can significantly affect colour determination. The authors found that visual judgement of soil colour between individuals using the same Munsell chart under heavily controlled conditions was variable. Similar results were obtained by Shields *et al.* (1966), leading them to suggest the use of spectrophotometry to standardise colour observation and eliminate observer variability; a reasonable suggestion that was never implemented. In fact the use of a spectrophotometer system to eliminate observer variation had been suggested another 35 years prior (Carter, 1931). Decades have passed since brewing industries dropped reference charts in favour of spectrophotometric colour standards (ASBC, 1949). It is absurd to imagine modern medical studies and analytical assays using reference charts for quantitative analysis. So why is soil science so slow to change? When observing soils it is important to eliminate, or at least account for, the sources of variability. Thus approaches, such as DSMorph, that can give more objectivity are a useful way forward.

1.3.2 Deconstructing DSMorph

In the appropriation of terms from other scientific disciplines, one must be sure to clearly define how they translate to their new environment. To avoid confusion the component topics of this new subdiscipline need to be deconstructed, to clarify meanings for, and approaches to, DSMorph.

1.3.2.1 Digital data acquisition

The digital component refers not to the devices themselves, but rather to the quantitative data that they can deliver. Data obtained using DSMorph tools are more objective and at times able to measure attributes that are unattainable using traditional methods. To investigate the benefits of these approaches the progress made by observing some soil structural attributes

using digital techniques is investigated. The procurement of digital data is not new, with those who might be considered pioneering digital soil morphometricians looking for new methods to quantify treatment effects on soil surface condition. Currence (1969) for example, quantified soil surface roughness under different tillage treatments using a profilometer system. The automated system was able to record relief information on punch cards at a height resolution of 0.01 inch. Meanwhile O'Callaghan and Loveday (1973) were able to use digitised images to quantify the effect of gypsum application on the length and width of crack skeletons (Fig. 1.1). In both of these studies digital quantification allowed investigation of attributes on a scale that was not feasible using manual techniques, and in doing so discerned treatment effects.

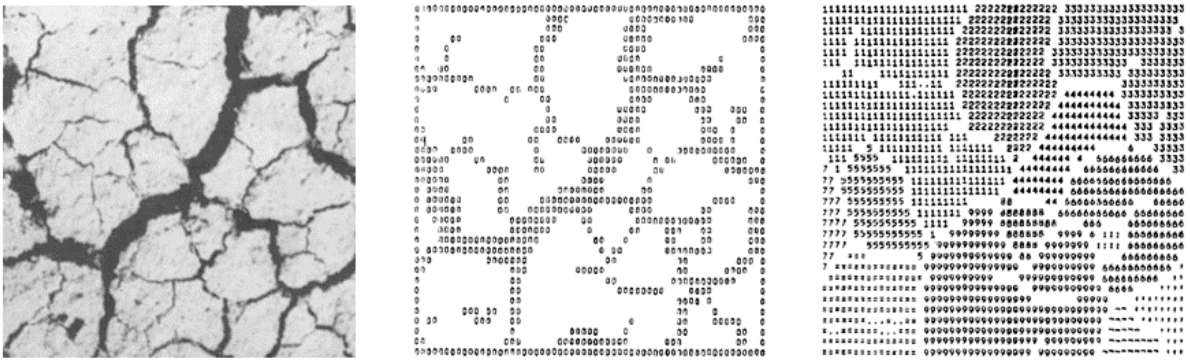


Figure 1.1 Section of original photograph showing a cracked soil surface (left), digitised crack "skeleton" (middle) and "ballooned" discrete peds (right), adapted from (O'Callaghan and Loveday, 1973).

Digital data acquisition of structural properties then extended vertically into the soil profile with a focus on micromorphology. Although not analysed directly in the field, thin sections were taken and image analysers were used to study pore distributions and how they work (Bouma *et al.*, 1977; Murphy *et al.*, 1977). Then computed tomographic scanning was used to investigate undisturbed soil material (Petrovic *et al.*, 1982; Hainsworth and Aylmore, 1983). Mesomophological analysis was introduced to bridge the gap between these micromorphological approaches and field description, which was largely qualitative and macromorphological (Koppi and McBratney, 1991). This meant that instead of measuring discrete points a continuous description of properties such as pore size and porosity could be

presented (Fig. 1.2a). This allowed interpretation of the relationship between structural properties and others such as redoximorphic features (Fig. 1.2c).

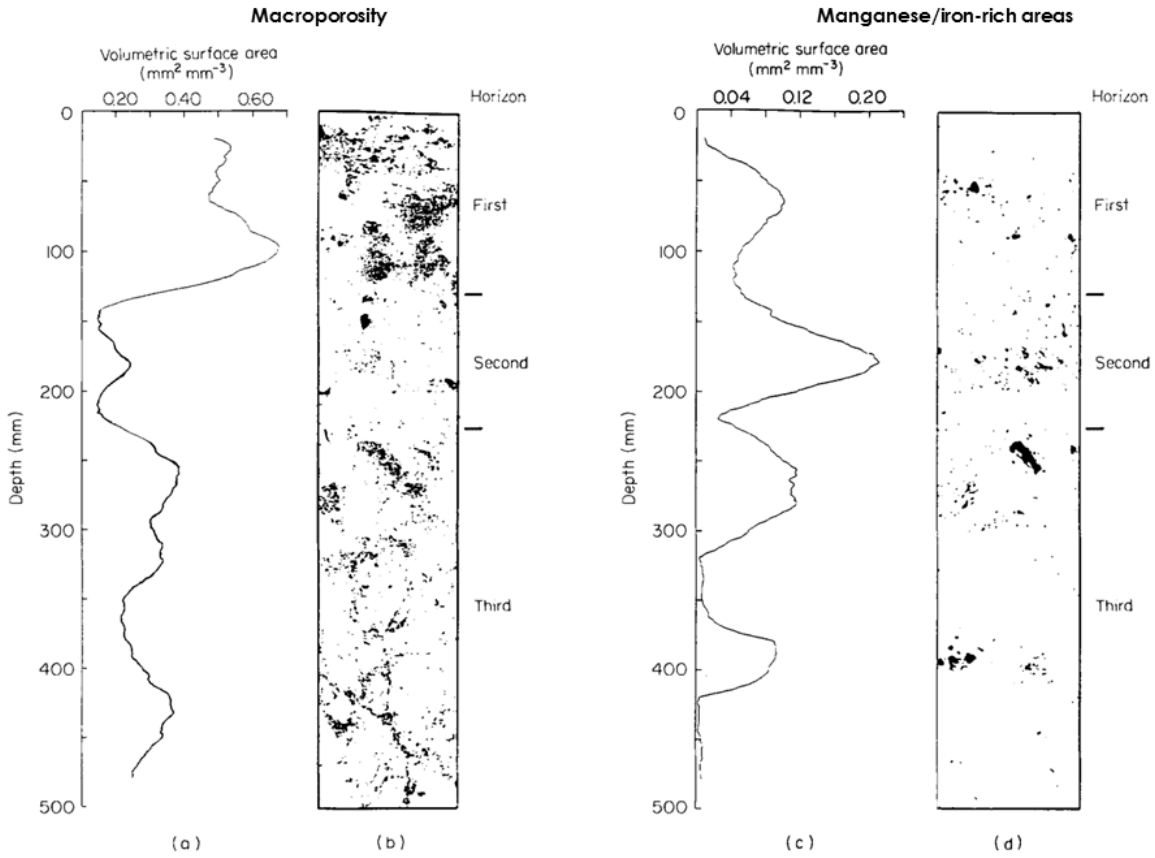


Fig. 1.2 Mesomorphological analysis of a Paleustalf (Koppi and McBratney, 1991): a) smoothed volumetric surface area of macropores; b) image of macropores (black) and soil solids (white) with horizon designations; c) smoothed volumetric surface area of manganese/iron-rich areas; d) image of manganese/iron-rich areas (black) and other soil solids and pores (white) with horizon designations. Images represent a 100 x 500 mm section of the soil profile with a resolution of approximately 0.2 mm. Smoothed lines were produced using a 20 mm moving average.

When representing soils in this continuous manner it is observed that soil properties are not uniform within soil horizons. This may lead us to question the conventional horizon based

representation of soil attributes and to ask: if the tools used to analyse soils are updated, does this mean that the way in which soils are described also needs to be re-addressed?

1.3.2.2 Soil form

"The essential problem in morphometrics is to measure the degree of similarity of two forms" (Blackith and Rayment, 1971, p. 9). However, what does form mean when applied to the soil profile? When assessing profile attributes, the concept of form extends beyond its original biological definition concerning the size and shape of a specimen. For our purposes form encompasses two aspects: geometric, or the disposition of attributes in a 2-, 3-, or 4D space; and multivariate, or the various attributes of interest within the space.

So how is the geometric aspect of form described? As suggested above, currently the geometric aspect of form is not adequately described. When samples are taken to the laboratory, dried, ground and then analysed, only the average value of the sample is reported and information on spatial variability is lost. Some DSMorph devices, such as hyperspectral cameras (Steffens and Buddenbaum, 2013) and laser scanners (Eck *et al.*, 2013), are able to scan in two dimensions and capture some of this vertical and lateral variation. Is this variation adequately captured using point based devices such as visible near-infrared diffuse reflectance (VisNIR) and portable X-ray fluorescence (pXRF) spectroscopy? To adequately capture spatial variation, the method in which soils are described must be redesigned. An expedient solution would be to include multiple vertical transects to allow the characterisation of variability within the observed object. For standardisation of variation over a fixed distance the vertical transects need to be a set distance apart and cover a fixed area, or if you take samples at right angles, within a fixed volume (Fig. 1.3).

Given the mean and the range of values calculated with depth, you can recognise that there is variation within that observed object (Fig. 1.4). The richness of information obtained if soil properties are described in this way can then be manipulated in a model such as a depth function or other laterally isotropic, vertically non-stationary random functions (McBratney and Moran, 1990).

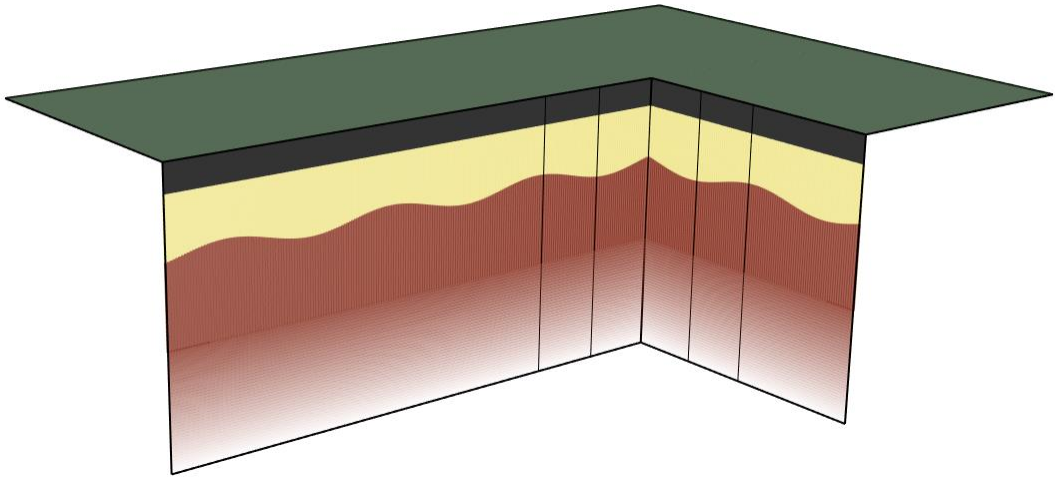


Fig. 1.3 A potential standard soil volumetric object. Three transects are analysed on one wall and others on adjacent wall orthogonal to the first.

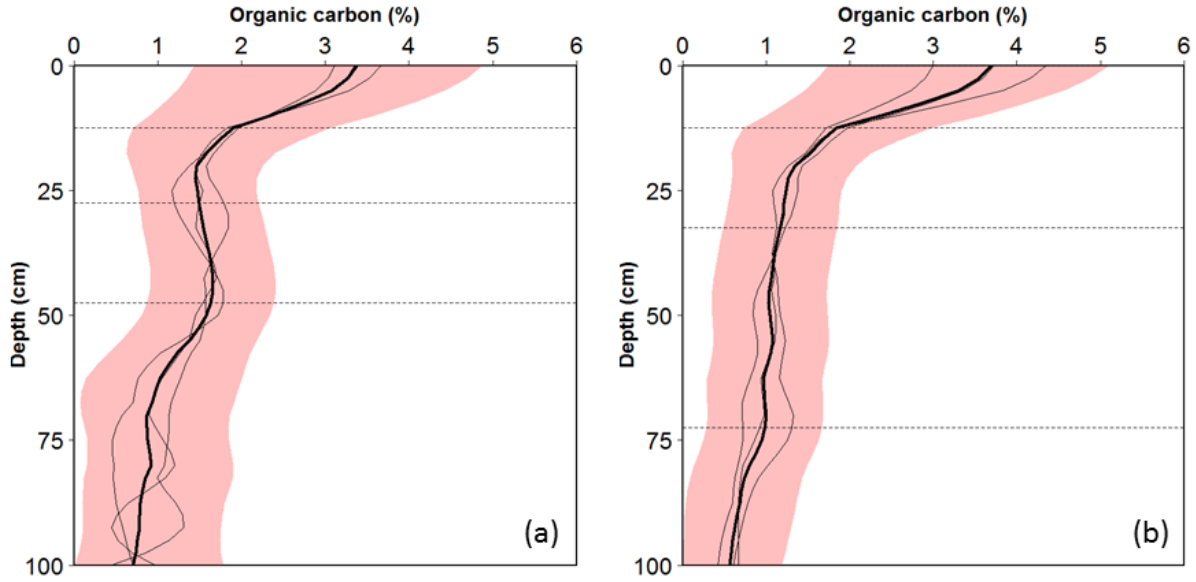


Fig. 1.4 Predicted organic carbon (%) of a) Eutrudept and b) Udipsamment produced from VisNIR readings of pit walls sampled with three vertical transects at 25 mm intervals to a depth of 100 cm. Thin black lines represent predicted values for the three vertical transects; thick black lines indicate the mean predicted value with depth; pink halos indicate the mean 95% prediction interval; and dashed horizontal lines indicate horizon designations.

1.3.2.3 Morphometric analyses

After soil properties have been captured digitally, multivariate statistics can be performed. The attributes themselves being geometric in nature or spatially located render this analysis morphometric. Hole and Hironaka (1960) were able to represent soil profiles in a multi-dimensional space and quantify the degree of similarity between two profiles. Around this time, numerically based taxonomic systems were developing in diverse fields, the key advantages of these being repeatability and objectivity (Sneath and Sokal, 1962). Such systems attempt to remove subjectivity from decision-making processes and should allow different scientists to arrive at the same conclusions, while increasing the accuracy and precision of the results (Bidwell and Hole, 1964). These benefits translate to numeric soil classification systems and such analyses may also play an important role in the development of a universal classification system (Brevik *et al.*, 2015). However, morphometric analyses extend beyond classificatory attempts. As they are able to connect both geometric and morphometric attributes, morphometric analyses may also provide evidence for pedogenetic theories. Stockmann *et al.* (2016), for example, used variation of pXRF-derived geochemical indices with depth to identify if a profile is polygenetic or derived from uniform parent material. Morphometric analysis may also shed light on soil-forming processes, connect properties and processes, and facilitate identification of relationships between properties.

1.3.3 Delineating digital soil morphometrics

One of the mantras of the subdiscipline seems to be to digitally enrich the toolkit of the field pedologist. However, Hartemink and Minasny (2014) also list many laboratory-based techniques, e.g. scanning electron microscopy and X-ray computed tomography, as potential DSMorph tools. While these devices can provide valuable information about soil profile attributes, they will probably not enrich the toolkit of the field pedologist anytime soon. Their inclusion also blurs the definition of what DSMorph is to such an extent that it encompasses laboratory analysis. The power of DSMorph comes with the capacity to objectively quantify soil attributes in the field using methods that have the capacity to increase sampling intervals and more readily quantify spatial variation compared to traditional methods. Digital soil morphometrics may be envisaged at the confluence of pedology, pedometrics and proximal

soil sensing (Fig. 1.5). Thus, in subsequent description, focus is given to techniques that have been performed, or have the potential to be performed, in the field.

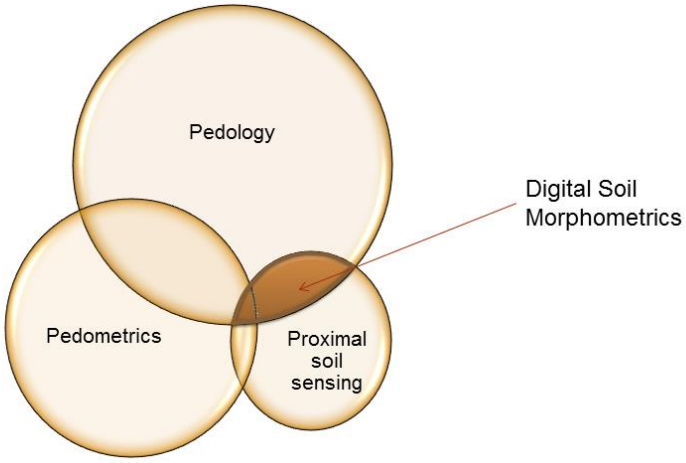


Figure 1.5 Suggested relationship of digital soil morphometrics to pedology, pedometrics and proximal soil sensing.

1.3.4 Novel and potential applications

When incorporating new tools and techniques into the field of soil science, it must be asked: is this simply to update the technology of field soil description, or is this to pose and answer new scientific questions? A review of the applications of DSMorph techniques as they apply to the prediction of attributes commonly used in soil description is given by Hartemink and Minasny (2014). Therefore, the following will highlight some of the peripheral, novel and developing fields that are progressing with potential to benefit from DSMorph techniques. These include continuous depth functions, spectrally derived soil horizons, soil inference systems, adaptive sampling procedures, and monitoring soil change.

1.3.4.1 Horizons or depth functions?

A unique question for DSMorph is to what extent is the distribution of soil properties better described by horizons or depth functions, and to which properties do these pertain?

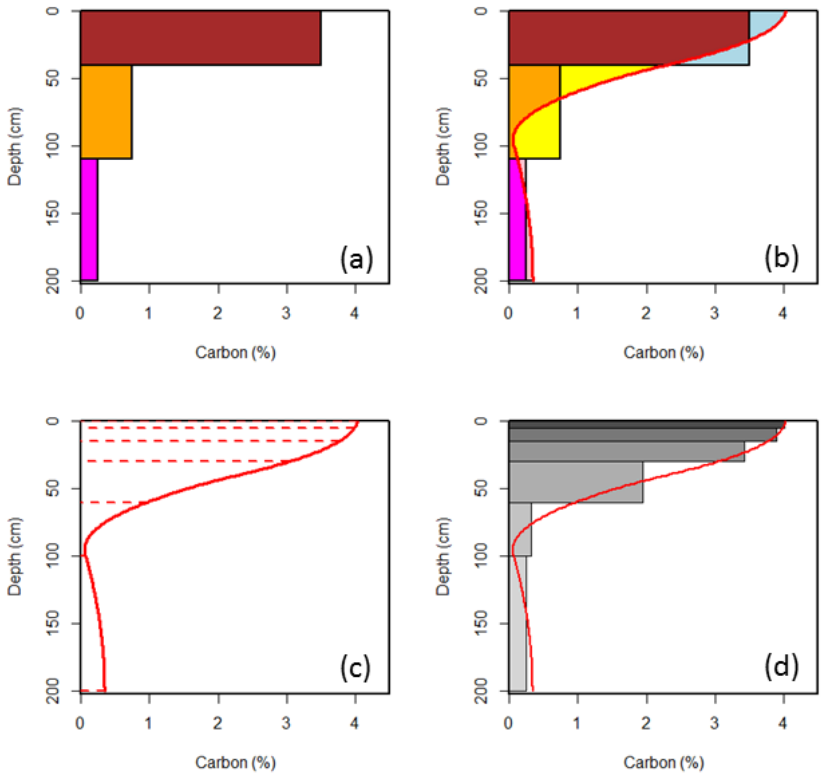


Fig. 1.6 Comparison of current techniques used to represent soil profile data: a) conventional quantitative profile/pedon description; b) fitting mass-preserving spline to horizon data (lambda=0.01); c) fitted spline, horizontal lines indicate GlobalSoilMap depth intervals; d) average spline predicted carbon % fitted to GlobalSoilMap depth intervals.

Most of pedology as it relates to soil description identifies properties based on horizons. This involves describing horizons and identifying the average properties of these horizons, resulting in the representation of discrete property distributions with depth (Fig. 1.6a). For some soil properties it may be a reasonable representation, but for many it is not. For example, Russell and Moore (1968) showed that the decrease in OC with depth in a soil profile is better represented by smooth exponential decay functions. Therefore, mass preserving splines have been fitted to obtained horizon data (Fig. 1.6b,c), signifying a movement away from using discrete horizon values, and towards describing soil depth functions. The superiority of such functions compared to average horizon values at describing the vertical non-stationarity of profiles has been established (Bishop *et al.*, 1999). However, when converting horizon-based data to splines some assumptions need to be made to create

this additional data, which may decrease accuracy. For example, splines invariably dampen actual minima and maxima values, resulting in a smoother predicted distribution (Ponce-Henandez *et al.*, 1986). What can this loss of information mean in terms of process or pedological understanding? In the case of soil permeability models, how would they benefit from continuous measurement of properties such as clay, OC, pore size and connectivity, compared to models based on average horizon values? While the concept of horizons must be conserved as they are a unique feature of soil, it does not mean that they are the only unique feature of soils. Going forward it must be determined whether the collection of horizon-based data continues, for later conversion to continuous depth data. Alternatively, the finer sampling resolution, enabled using DSMorph techniques, could be utilised to capture more spatial variation (such as the example in Fig. 1.4) and create more accurate depth functions directly.

1.3.4.2 Spectrally-derived horizons

Nikiforoff (1931) lamented that "soil horizons and their nomenclature is probably the most confused point in the technic of the description of the soil". Nikiforoff's frustration was that the term "B" horizon was so broad that it did not convey any real pedogenic meaning other than filling its place in the A-B-C horizon succession. At the time it was known that there was great diversity in the horizons from different soil types, but the nomenclature was too simple to account for this richness of interpretation. It was not until much later with the implementation of suffix notation that different B horizons could be succinctly distinguished (Soil Survey Staff, 1951). Is this nomenclature now sufficient to adequately describe the diversity of soil horizons?

Any horizon is a mixture of materials. Using current soil description it is difficult to capture this heterogeneity. With DSMorph techniques the boundaries of horizon identification and classification may be advanced, and measures to describe profiles in a more continuous manner developed. One suggested method involves identifying spectrally homogeneous zones from VisNIR derived fuzzy cluster memberships (Fajardo *et al.*, 2015). This method eliminates observer bias and allows direct investigation of class membership within and between profiles (Fig. 1.7). Could spectrally derived horizons become the new nomenclature to better characterise soils, and interclass memberships the new descriptors? Objective

horizon recognition is also being explored using pXRF (Weindorf *et al.*, 2012; Minasny *et al.*, 2016; Adhikari *et al.*, 2016) and hyperspectral cameras (Steffens and Buddenbaum, 2013).

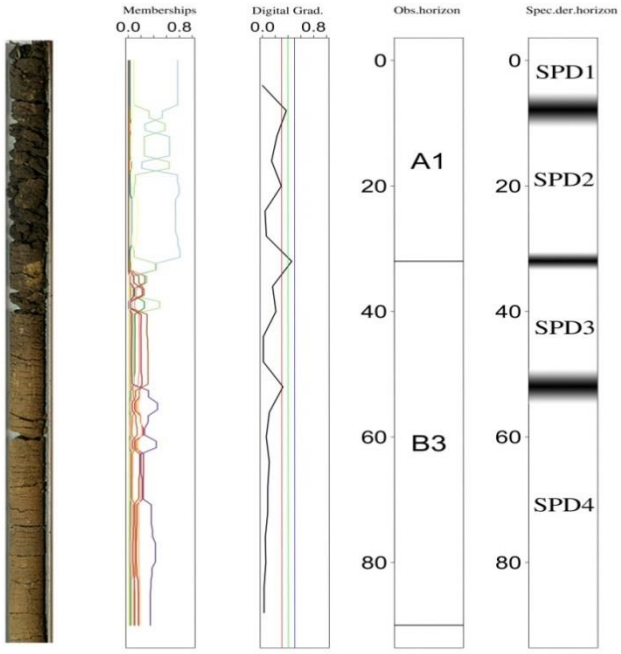


Fig. 1.7 Left to right: photograph of soil core; fuzzy membership classes; digital gradient; horizons observed using conventional techniques; spectrally derived horizons. Taken from (Fajardo *et al.*, 2015).

1.3.4.3 Adaptive sampling

DSMorph techniques have the potential to derive data in the field, but how can this information be used to perform more meaningful operations in the field? A proposed strategy for assessing soil contamination suggests taking advantage by adapting sampling and analysis in real-time (Horta *et al.*, 2015). Calculations were made to find the conditional probability density function of the contaminant and the loss function. From this information, an optimal remediation plan can be made, taking into account both sampling and remediation costs. The method facilitates honing in on contaminated areas, prioritisation of areas of high uncertainty for subsequent focussed sampling and continuous updating of the map until an overall quality criterion is achieved. A similar approach can be envisaged for soil mapping units or profiles, but how to do it? A methodology is required to couple imaging or other techniques with point

based sampling devices to identify the next most valuable point of a soil profile to investigate, and determine when sufficient information has been gained for effective characterisation.

1.3.4.3 Monitoring soil change

The soil environment is not static, though it is sometimes represented as such. Increasing global recognition of issues such as C sequestration (Lal, 2004), provision of ecosystem services (Kreuter *et al.*, 2001), or soil security (McBratney *et al.*, 2014), have generated a corresponding need for increased knowledge on the variability of soil attributes in space and time. Monitoring soil condition indicators, as well as elements of soil degradation such as contamination, loss of organic matter, compaction, acidification and salinisation, are important. The monitoring of soil requires the use of reliable, inexpensive and, at times, non-destructive techniques. These issues have proved troublesome for soil monitoring programs, especially when considering costs of sampling and analysis. As most DSMorph techniques are quantitative and generate a wealth of data, more subtle changes can be detected. Research is needed to identify how the application of DSMorph tools and techniques can improve the efficiency and viability of soil monitoring programs.

1.3.5 Missing Technology

The future of DSMorph is tied to progress in pedology but in particular to proximal soil sensing techniques. Current techniques need to be tested and utilised, new technologies need to be adapted as they arise. Further, overlooked technologies must be resurrected. Large sections of the electromagnetic spectrum are being used, as well as ultrasonics, electrical resistivity and physical measurements but others, such as magnetic susceptibility, appear to be underexploited (Mullins, 1977). Continued investigation of novel technologies is paramount. Current techniques are predominately intrusive. Ideally, all soil attributes of interest would be predicted from above the soil surface, using non-invasive techniques. Ground penetrating radar, γ-radiometrics and electromagnetic induction are three such existing techniques, but they can predict only a few properties of interest. While invasive techniques may fill the gap in the near-term, as expedient intermediaries, the holy grail of soil

observation would be the development of a non-invasive sensor that could quantify all attributes of interest from the soil surface.

In reality, such a device is a long way off, and currently no single sensor or technique has the capacity to accurately predict all attributes of interest. The greatest power in the near future will come from putting the information gained from multiple sensors together. It is this data fusion, combined with soil inference systems, that will provide the most useful information. When combining data from multiple sensors a number of approaches have been utilised. For example, input data may be analysed individually then results combined using a model-averaging procedure (Malone *et al.*, 2014). Spectral data have also first been combined using concatenation and then analysed concurrently (Viscarra Rossel *et al.*, 2006; Wang *et al.*, 2015). Other techniques focus on exploiting the strengths of individual devices. Jones and McBratney (2016) suggest combining VisNIR's ability to provide information on bonding environments with the elemental concentrations reported from an pXRF device to predict soil mineralogy using an integrated chemometric and mass balance approach. The potential data fusion methodologies are myriad, but more focus needs to be given to the quantification of uncertainties. This will offer more valuable input data for soil inference systems that may connect predictions with the wealth of existing soil knowledge, and amplify the number of predicted attributes (McBratney *et al.*, 2006).

1.4 Conclusion

- DSMorph can provide more precise soil properties data with quantified spatial uncertainty than conventional soil description. There are untold pedological, edaphic, and environmental applications to be gained from applying DSMorph tools and techniques.
- If DSMorph is going to digitally enrich the toolkit of the field pedologist, then DSMorph is probably best considered as a special case of proximal soil sensing.
- DSMorph can represent soil profiles in two different ways- depth functions or horizons, but as yet the best approach for the various properties is not clear.
- DSMorph can be used to make field inferences to optimise and adapt sampling in real time.
- DSMorph can enable quantification of change in soil condition and prove useful in soil monitoring programs.

1.5 References

Adhikari, K., Hartemink, A.E., Minasny, B. 2016. Mapping a Profile Wall of a Typic Udipsamments from the Central Sands in Wisconsin, USA. In: Hartemink, A.E., Minasny, B. (Eds.), Digital Soil Morphometrics. Springer, Dordrecht.

ASBC American Society of Brewing Chemists (1949). Beer 10-A Spectrophotometric Color Method. ASBC Methods of Analysis.

Bidwell, O., Hole, F. (1964). Numerical taxonomy and soil classification. Soil Science 97(1), 58-62.

Billmeyer, F. W., Saltzman, M. (1981). Principles of color technology. John Wiley & Sons, New York, NY.

Bishop, T., McBratney, A., Laslett, G. (1999). Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91(1), 27-45.

Blackith, R. E., Rayment, R. A. (1971). Multivariate morphometrics. Academic Press, London.

Bouma, J., Jongerius, A., Boersma, O., Jager, A., Schoonderbeek, D. (1977). The Function of Different Types of Macropores During Saturated Flow through Four Swelling Soil Horizons. Soil Science Society of America Journal 41(5), 945-950.

Brevik, E. C., Calzolari, C., Miller, B. A., Pereira, P., Kabala, C., Baumgarten, A., Jordán, A. (2015). Soil mapping, classification, and pedologic modeling: History and future directions. Geoderma.

Carter, W. (1931). Color analysis of soils with spectrophotometer. American Soil Survey Association Bulletin 12, 169-170.

Clarke, G. R. (1936). The study of the soil in the field. Oxford University Press, London.

Currence, H. D. (1969). Development of a method for measuring and describing soil surface roughness, Iowa State University. Retrospective Theses and Dissertations. Paper 3568.

Eck, D. V., Hirmas, D. R., Giménez, D. (2013). Quantifying Soil Structure from Field Excavation Walls using Multistripe Laser Triangulation Scanning. Soil Science Society of America Journal 77(4), 1319-1328.

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.

Hainsworth, J., Aylmore, L. (1983). The use of computer assisted tomography to determine spatial distribution of soil water content. Soil Research 21(4), 435-443.

Hartemink, A. E., Minasny, B. (2014). Towards digital soil morphometrics. Geoderma 230, 305-317.

Hole, F. D., Hironaka, M. (1960). An experiment in ordination of some soil profiles. Soil Science Society of America Journal 24(4), 309-312.

Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T., McBratney, A., Pallasser, R., Pozza, L. (2015). Potential of integrated field spectroscopy and spatial analysis for

enhanced assessment of soil contamination: A prospective review. Geoderma 241, 180-209.

Jones, E.J., McBratney, A.B., 2016. In-situ analysis of soil mineralogy through conjoint use of visible, near-infrared and X-ray fluorescence spectroscopy. In: Hartemink, A.E., Minasny, B. (Eds.), Digital Soil Morphometrics. Springer, Dordrecht.

Koppi, A., McBratney, A. (1991). A basis for soil mesomorphological analysis. Journal of Soil Science 42(1), 139-146.

Kreuter, U. P., Harris, H. G., Matlock, M. D., Lacey, R. E. (2001). Change in ecosystem service values in the San Antonio area, Texas. Ecological Economics 39(3), 333-346.

Lal, R. (2004). Soil carbon sequestration to mitigate climate change. Geoderma 123(1), 1-22.

Malone, B. P., Minasny, B., Odgers, N. P., McBratney, A. B. (2014). Using model averaging to combine soil property rasters from legacy soil maps and from point data. Geoderma 232–234, 34-44.

McBratney, A., Field, D. J., Koch, A. (2014). The dimensions of soil security. Geoderma 213, 203-213.

McBratney, A. B., Minasny, B., Viscarra Rossel, R. A. (2006). Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. Geoderma 136(1), 272-278.

Minasny, B., Stockmann, U., Hartemink, A.E., McBratney, A.B., 2016. Measuring and modelling soil depth functions. In: Hartemink, A.E., Minasny, B. (Eds.), Digital Soil Morphometrics. Springer, Dordrecht.

Mullins, C. (1977). Magnetic susceptibility of the soil and its significance in soil science–a review. Journal of Soil Science 28(2), 223-246.

Murphy, C. P., Bullock, P., Turner, R. H. (1977). The measurement and characterisation of voids in soil thin sections by image analysis. Part I. Principles and techniques. Journal of soil science 28(3), 498-508.

Nikiforoff, C. (1931). The history of A, B, C. American Soil Survey Association Bulletin 12, 67-70.

Nikiforoff, C. (1941). Morphological classification of soil structure. Soil Science 52(3), 193-212.

O'Callaghan, J. F., Loveday, J. (1973). Quantitative measurement of soil cracking patterns. Pattern Recognition 5(2), 83-98.

Petrovic, A., Siebert, J., Rieke, P. (1982). Soil bulk density analysis in three dimensions by computed tomographic scanning. Soil Science Society of America Journal 46(3), 445-450.

Ponce-Henandez, R., Marriott, F., Beckett, P. (1986). An improved method for reconstructing a soil profile from analyses of a small number of samples. Journal of Soil Science 37(3), 455-467.

Russell, J. S., Moore, A. W. (1968). Comparison of different depth weightings in the numerical analysis of anisotropic soil profile data. Transactions of the 9th International Congress of Soil Science 4, 205–213.

Sánchez-Marañón, M., Huertas, R., Melgosa, M. (2005). Colour variation in standard soil-colour charts. Soil Research 43(7), 827-837.

Shields, J., St. Arnaud, R., Paul, E., Clayton, J. (1966). Measurement of soil color. Canadian Journal of Soil Science 46(1), 83-90.

Sneath, P. H., Sokal, R. R. (1962). Numerical taxonomy. Nature 193(4818), 855-860.

Soil Survey Staff (1937). Soil survey manual. USDA, Washington DC.

Soil Survey Staff (1951). Soil survey manual. USDA, Washington DC.

Steffens, M., Buddenbaum, H. (2013). Laboratory imaging spectroscopy of a stagnic Luvisol profile—High resolution soil characterisation, classification and mapping of elemental concentrations. Geoderma 195, 122-132.

Stockmann, U., Cattle, S. R., Minasny, B., McBratney, A. B. (2016) Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenisis. Catena 139, 220-231.

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131(1–2), 59-75.

Wang, D., Chakraborty, S., Weindorf, D. C., Li, B., Sharma, A., Paul, S., Ali, M. N. (2015). Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. Geoderma 243–244, 157-167.

Weindorf, D. C., Zhu, Y., Haggard, B., Lofton, J., Chakraborty, S., Bakr, N., Zhang, W., Weindorf, W. C., Legoria, M. (2012). Enhanced Pedon Horizonation Using Portable X-ray Fluorescence Spectrometry. Soil Science Society of America Journal 76(2), 522-531.

2 SENSORS, METHODS AND SITE DESCRIPTIONS

A number of methods, analysis sites and proximal soil sensors are shared by the research chapters in this thesis, for brevity they are described in detail in this chapter. The reader will be referred back to this chapter for more detail as required. Methods that are pertinent to a single chapter will be described in detail within that particular chapter.

2.1 Proximal soil sensors

2.1.1 Visible near-infrared diffuse reflectance spectroscopy

Visible near-infrared diffuse reflectance spectroscopy (VisNIR) is a well-established tool for soil investigation (e.g. Dalal and Henry, 1986; Ben-Dor and Banin, 1995). It has shown particular promise for the prediction of soil carbon, texture and CEC (e.g. Islam *et al.*, 2003; Sørensen and Dalsgaard, 2005). Soil information can be gained from VisNIR under a number of modes, most commonly on air-dry and ground (<2 mm) samples in the laboratory. Growing attention is being given to proximal sensing using VisNIR as tine-based implements (Mouazen *et al*., 2007), push-probes (Ben-Dor *et al*., 2008), and point-based sensors *in situ* (Viscarra Rossel *et al*., 2009).

Visible near-infrared diffuse reflectance spectroscopy utilises the absorption of electromagnetic radiation by molecular bonds to discern soil properties. Samples are irradiated with light containing the range of frequencies of interest. This incoming radiation causes molecular bonds in the sample to bend and stretch, and in doing so absorb a portion of the incoming radiation at characteristic wavelengths. The reflected light is received at the detector, and commonly reported as the relative reflectance of a sample compared to a baseline scan of highly reflective substance. The absorbance features in VisNIR spectra are generally attributed to combinations and overtones of fundamental absorption features in the mid-infrared range of the electromagnetic spectrum. As such, while a small number of soil properties have visible absorption features in the VisNIR spectra (Fig. 2.1), for quantitative analysis of a soil VisNIR spectrum, chemometric approaches are required.

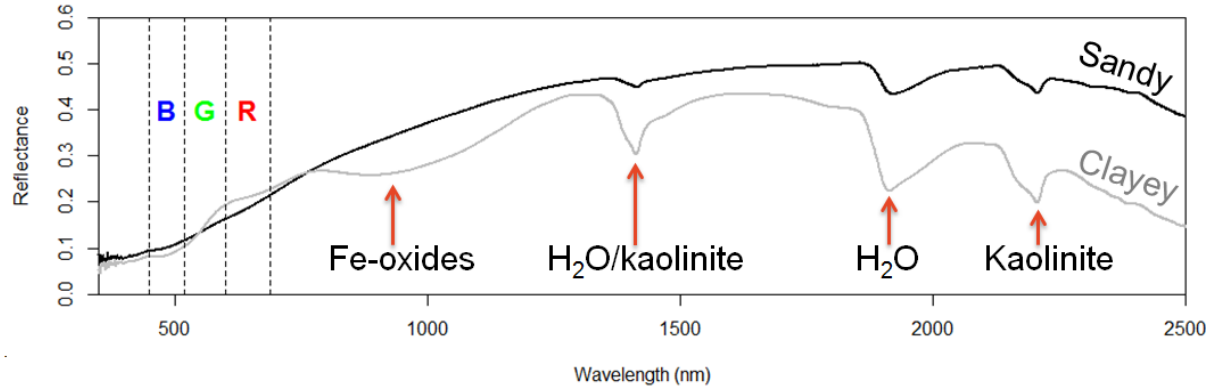


Figure 2.1 Visible near-infrared reflectanc spectra of a sandy, topsoil sample (black) and a clayey, subsoil sample (grey) from the same profile (Site 3). The location of absorption features of Fe-oxides, water and kaolinite are indicated. Red, green and blue colour bands are designated.

2.1.2 Portable X-ray fluorescence spectroscopy

Portable X-ray fluorescence spectroscopy (pXRF) is a relatively new proximal soil sensing device. Portable X-ray fluorescence spectroscopy is able to characterise the elemental composition of a sample, for all elements heavier than Na, i.e. $Z \geq 12$. Applications of pXRF in soil science have included: prediction of heavy metal pollution (Carr et al., 2008); texture (Zhu et al., 2011); and investigation of pedological considerations, such as lithological

discontinuities (Weindorf et al., 2015), and weathering indices (Stockmann et al., 2016). Few studies have investigated lighter elements, such as Al and Si, which are of vital importance to understanding soil physical and chemical properties.

Portable X-ray fluorescence spectroscopy utilises the release of characteristic fluorescence photons from samples. An incident X-ray beam is provided by an X-ray tube. An X-ray tube consists of cathode and an anode in a vacuum sealed envelope. An energy potential is applied at the cathode and electrons pass through a vacuum to the anode. Electrons undergo energy loss as they move from the cathode to the anode, resulting in the generation of X-radiation. The electrical potential across the cathode and anode can be modified to produce X-rays of characteristic energy levels e.g. 10 or 50 keV. The generated X-rays are directed towards a sample, forming the incident X-ray beam.

During fluorescence events an incident X-ray expels a K or L shell electron from its orbit (Fig. 2.2a). This produces a hole in the electron shell and destabilises the electronic structure of the atom. An electron from a higher shell will fall into this lower energy orbit (Fig. 2.2b). Excess energy from this event is released as a fluorescence photon of characteristic energy level (Fig. 2.2c), which is detected by the pXRF detector.

As well as fluorescence events, incident X-ray photons have two other potential interactions with electrons in a sample target:

> **Compton scattering:** a photon hits an electron, ejecting the electron from its shell and losing a fraction of its energy in the scattering event.
>
> **Rayleigh scattering:** a photon hits a strongly bound electron causing it to oscillate in its shell and release energy at the same frequency as the source X-ray photon.

For heavier elements, Compton scattering is effectively zero and only Rayleigh scattering occurs. Conversely, lighter elements have many loosely bound electrons and give rise to a larger proportion of Compton scattering and a reduce proportion of Rayleigh scattering. The addition of light elements to a sample, such as H and O in water molecules, will increase the

amount of Compton scattering, while Rayleigh and fluorescence events are reduced. Compton normalisation, normalising a spectrum based the Compton peak, has thus been developed as a way to compensate for variable moisture contents in a sample.

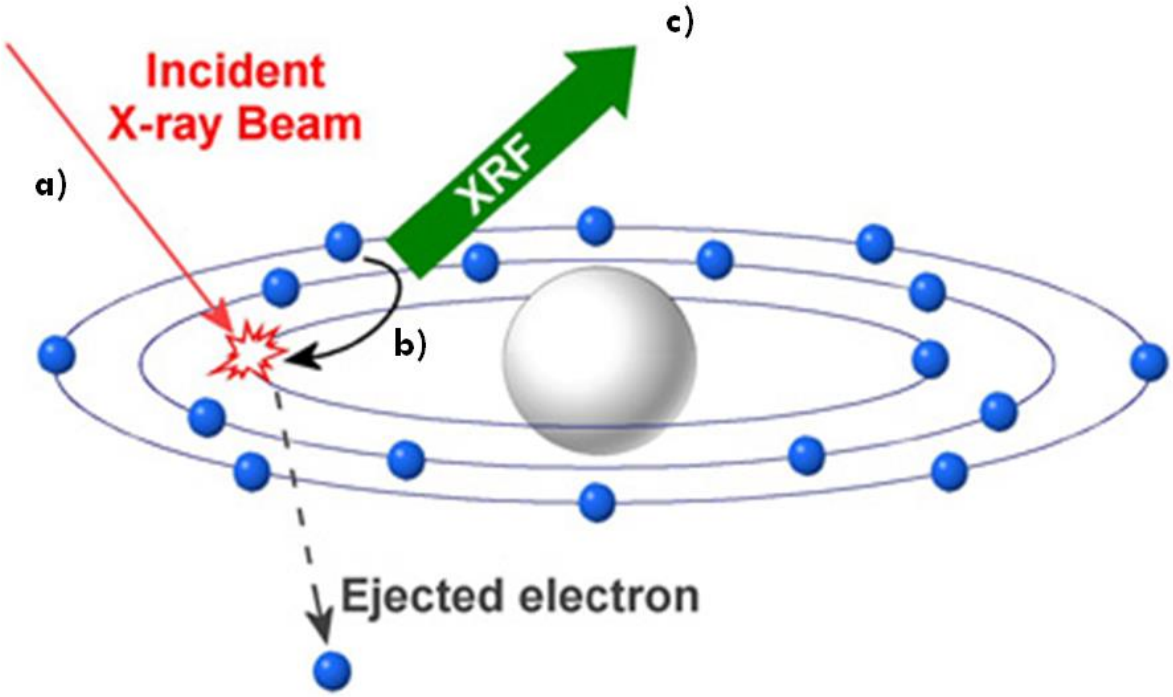


Figure 2.2 A fluorescence event. a) The red X-ray incident beam ejects an electron from its orbit. b) an electron from a higher orbital drops to the lower energy level to give a full orbital shell. c) A fluorescence photon of characteristic energy is released.

Spectra are subject to a number of other processing techniques to remove effects such as: matrix enhancement and absorption of characteristic photons; sum and escape peaks occurring at the detector; and deconvolution of overlapping peaks.

2.2 Methods

2.2.1 Proximal sensing devices

2.2.1.1 Visible near-infrared spectrometer

Visible near-infrared spectra were obtained with an AgriSpec™ device, connected via fibre-optic cable to a contact probe attachment (Analytical Spectral Devices, Boulder, Colorado, USA). Illumination was provided with a halogen lamp inbuilt into the contact probe. A Spectralon® tile (Labsphere Inc., North Sutton, New Hampshire, USA) was used to take a baseline reading before the first measurement and following every 15-20 measurements. Spectralon® is made of polytetraflouroethylene and cintered halon. It has the largest diffuse reflectance values of any known substance over the 350 – 2500 nm wavelength range, with a minimum reflectance of 95% over this range, and greater than 99% in the 400 – 1600 nm wavelength range.

Reflectance data is recorded on three separate detectors: VNIR covering 350 – 1,000 nm; SWIR1 covering 1,000 – 1,800 nm; and SWIR2 covering 1,800 – 2,500 nm respectively. The VNIR detector contains a 512 element silicon photo-diode array. The spectral resolution is 3 nm (full-width-half-maximum) at 700 nm, and the sampling interval is 1.4 nm. The SWIR1 and SWIR2 bandwidths are captured using single Indium Gallium Arsenide (InGaAs) detectors. This means reflectance is captured sequentially for each wavelength, rather than in parallel, as with the VNIR detector array. Each SWIR detector is serviced by an oscillating concave holographic grating, to expose the detectors to different wavelengths of energy. The oscillation period of the gratings is the rate-determining step, requiring 100 ms $scan^{-1}$. The spectral resolution is 10 nm (full-width-half-bandwidth) at 1400 and 2100 nm respectively, and the sampling interval is 2 nm.

Indico® Pro software was used to interface with the spectrometer. Inbuilt algorithms combined and smoothed information from the three detectors and exported a full, 350 - 2,500 nm, spectrum at 1 nm resolution. Spectra were exported as relative reflectance, calculated as the ratio of reflectance, in digital number, from the sample and the baseline reading. The

internal capture rate of this device is 10 Hz and reflectance readings were reported as the average of 40 internal readings.

2.2.1.2 Portable X-ray fluorescence spectrometer

All pXRF analysis was performed with an Olympus Delta Premium pXRF (Olympus, Center Valley, Pennsylvania, USA). The device features an Au anode X-ray tube and a large-area silicon drift detector. Readings were taken in the *Geochem* mode, which is a dual beam configuration that irradiates the sample with X-rays of energy 50 and 10 keV successively. To improve the signal to noise ratio, samples were scanned for 30 s at each energy level. The received spectra are converted into elemental concentrations, ppm or %, based on an inbuilt calibration utilising fundamental parameters. This mode also gives an estimate of the percentage of light elements ($Z \leq 12$) that are present in the sample, based on Compton scattering. Light elements do not have a clearly defined fluorescence emission line but are estimated from the scatter intensity ratio between Rayleigh and Compton scatter peaks. A calibration check was performed daily with a 316 stainless steel alloy clip. The internal calibration check ensures that predictions remain within pre-set tolerances, otherwise an alert is given. A SiO_2 blank and multiple NIST standards were scanned immediately following calibration and again hourly during operation. Scanning the SiO_2 blank facilitates the detection of contamination on the pXRF measurement window. The varying elemental concentrations of the NIST soil standards determine if output elemental concentrations are within an acceptable range and enable tracking of the performance of the device through time (Appendix A1).

2.2.2 *In situ* measurements and sampling

At each site an excavator was used to dig a soil pit approximately 1 m wide, 5 m long and 1.5 m deep in the middle. A suitable 1 m × 1 m section of the soil pit wall was identified, and a smooth surface was prepared by shearing the excess soil with a combination of shovel and asparagus knife. The final shearing was conducted horizontally, perpendicular to the soil surface, and progressing from the top to the bottom of the pit wall to limit surface

contamination from surface debris. Galvanised nails were inserted at 10 cm intervals along transects to guide proximal sensor reading locations.

Proximal soil sensor readings were taken in 2.5 cm increments to give 41 readings over each 1 m transect. Three vertical transects were taken at 0, 50 and 100 cm lateral spacing, as well as three horizontal transects at 0, 50 and 100 cm depth (Fig. 2.2). Portable X-ray fluorescence readings were taken on the vertical transects only. Bulk density cores were taken at 10 cm increments on the 0 and 50 cm vertical transects. Bulk density cores were immediately sealed using vinyl tape to preserve field condition moisture. The bulk density cores had an internal diameter of 4.7 mm and a height of 40 mm. The soils were described using routine soils description and horizon-based samples were taken for laboratory analysis. The rationale for this sampling methodology is outlined in Chapter 4.

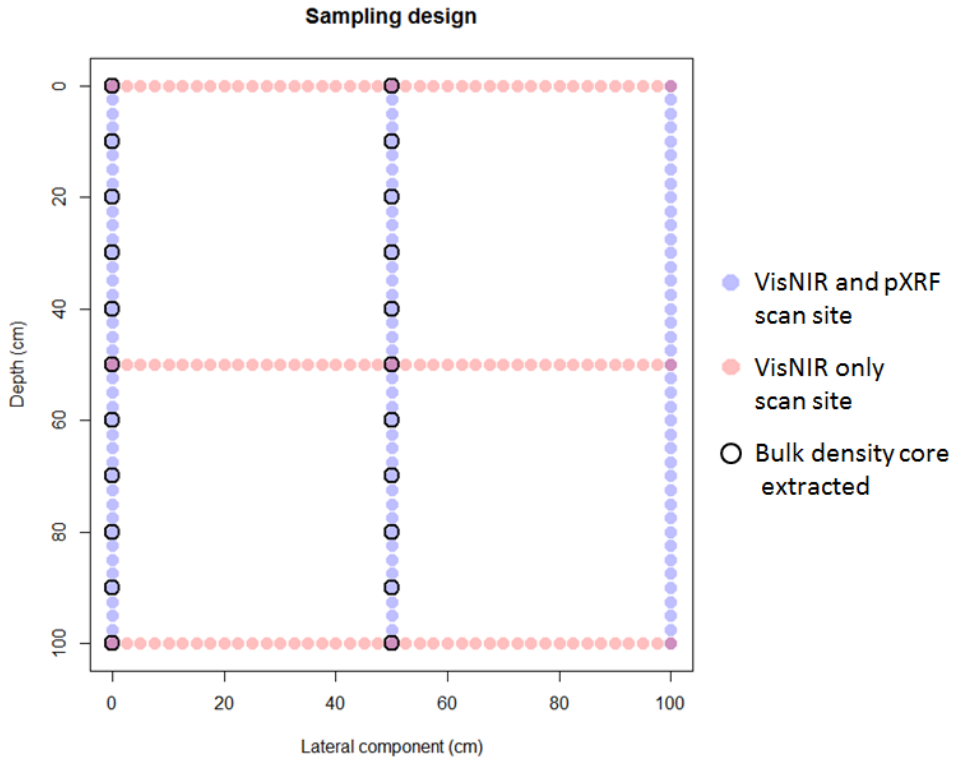


Figure 2.3 Schematic representation of sampling design for the fifteen soil profiles. Sampling sites for visible near-infrared (VisNIR) and portable X-ray fluorescence (pXRF) spectroscopy, and bulk density cores are indicated.

2.2.3 *Ex situ* measurements

Bulk density cores were kept sealed until immediately prior to laboratory scanning, thus allowing soil moisture to equilibrate and minimise surface drying. The bulk density core was scanned using the same VisNIR and pXRF devices as the *in situ* scans. For VisNIR the samples were scanned three times; the mean spectral reflectance was calculated and used for further analysis. Each pXRF reading was taken at the centre of each core to minimise variation in scanning location following drying. The bulk density cores were dried in a 40°C oven for four days, and then rescanned in air-dry, intact condition. The soils were then ground to pass through a 2 mm sieve and rescanned using VisNIR.

2.2.4 Laboratory analyses

All laboratory measurements were conducted on air-dry samples which were ground to pass through a 2 mm sieve, unless stated otherwise. pH was measured in a 1:5 soil to deionised water gravimetric ratio and also in a 1:5 soil to 0.01 M calcium chloride solution using a Mettler Toledo S220 SevenCompact™ pH/Ion meter. Electrical conductivity was measured in a 1:5 soil to deionised water gravimetric ratio using a Mettler Toledo SevenCompact™ conductivity meter.

Particle size analysis was performed using the hydrometer method (Gee and Bauder, 1986). Samples were agitated using end over end shaking in sodium hexametaphosphate solution for 48 h prior to analysis. Sieving was used to isolate fine- and coarse-sand sized particles.

Prior to total carbon and nitrogen analysis a subsample (~10 g) was finely ground (<53 µm) using a Fritsch Mortar Grinder Pulverisette 2 (Fritsch, Germany) for four minutes at a vibrational frequency of 50-60 Hz. Total carbon and nitrogen were then quantified via the combustion method using a Vario Max CNS analyzer (Elementar Analysensysteme GmbH, Hanau, Germany). Organic carbon was obtained using the Walkley and Black method (Walkley and Black, 1934). CO_3^{-2} equivalent was calculated using the rapid titration method (Piper, 1942), as compiled by (Reeuwijk, 1993), for profiles wherein any horizon tested positively for carbonates using 1 M HCl.

Exchangeable cations were derived by two different methods depending on the presence of carbonates and soluble salts. The alcoholic ammonium chloride at pH 8.5 method (Rayment and Lyons, 2011, pp. 307-313), following pre-treatment to remove soluble salts, was used for soils that were found to have carbonates in any horizon. The ammonium acetate method was used for all other soils, as these soils displayed a neutral to acidic pH and negligible quantities of soluble salts. For these soils, exchangeable acidity was also calculated using potassium chloride (Rayment and Lyons, 2011, pp. 321-338).

2.2.5 Data processing and analysis

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

2.2.5.1 Spectral pre-processing

2.2.5.1.1 Splice correction and trimming

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 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.

2.2.5.1.2 Savitzky-Golay filtering

Reflectance readings were converted to absorbance using, A= log(1/R). Data were 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, 1964). This is a progressive function that fits a local polynomial regression of specified order over points lying within the

window. The value at the central point of the window is replaced with the smoothed value. This filter increases the signal to noise ratio of a spectrum, without greatly distorting the signal.

2.2.5.1.3 Standard normal variate transformation

The standard normal variate transformation was used to centre and scale each spectrum individually to a mean of zero and unit variance (Eqn 2.1). This removes the multiplicative interferences of scatter and particle size (Barnes *et al.*, 1989).

Equation 2.1

$$SNV = \frac{\alpha - \bar{\alpha}}{s_\alpha}$$

Where:

α is the spectrum to be transformed

$\bar{\alpha}$ is the mean of the spectrum α

s_α is the standard deviation of spectrum α

2.2.5.2 External parameter orthogonalisation

An external parameter orthogonal (EPO) transformation was used to compensate for the negative effects of variable moisture content when sampling *in situ*. This algorithm was initially developed to remove the effect of temperature variation when estimating the sugar content of fruit using partial least squares regression of the processed and projected spectra (Roger *et al.*, 2003). Its versatility was demonstrated by Minasny *et al.* (2011), who showed that it could also remove the deleterious effect of moisture when estimating SOC in moist and dry samples, without prior knowledge of soil moisture content.

The EPO process identifies areas in the spectra that are affected by soil moisture and projects the spectra into a new space orthogonal to this variation. The projected spectra are effectively

independent of variation in soil moisture, while useful soil information is conserved. The spectra, X, are thought of as the sum of matrices (Eqn 2.2)

Equation 2.2

$$X = XP + XQ + R$$

Where:

P is the projection matrix of the useful part of the spectra

Q is the projection matrix of the not useful part of the spectra, i.e. the part affect by variable soil moisture

R is the residual matrix

To construct P, the difference spectra between moist and dry samples is calculated, $D = X_{moist} - X_{dry}$. Principal component analysis is performed on $D^T D$, and the number of factors is defined to construct Q. P is then constructed by subtracting Q from an identity matrix. The transformed spectra, X^*, are then calculated by multiplying the spectra by P (Eqn 2.3).

Equation 2.3

$$X^* = XP$$

The spectra must be pre-processed in the same manner used to construct P. An EPO projection matrix developed from the same dataset as Minasny *et al.* (2011) was utilised for this thesis.

2.2.5.3 Geochemical ratios

When comparing elemental concentrations under field moist and air-dry scenarios it was observed that geochemical ratios provide more stable metrics (Appendix A2). Seven dominant observable elements, Al, Si, S, K, Ca, Ti and Fe, made up a mean of 99.45 (s.d. = 0.65) of the mass of total identifiable elements (Z>12) for all in situ scans. Reported elemental concentrations that were below the limit of detection for the device were set to zero. This may be a source of error, as limits of detection are affected by the samples ability to produce a fluorescence X-ray, and the energy of the fluorescence X-ray. In general, elements with a smaller atomic number have a larger limit of detection, e.g. ~1% for Al, whereas elements with greater atomic numbers have much lower limits of detection, e.g. <5 ppm for Pb. For each pXRF observation, the ratio was calculated as the mass of an element divided by total mass of the seven dominant observable elements (Eqn. 2.4).

Equation 2.4

$$r_{i,j} = \frac{c_{i,j}}{\sum_j c_{i,j}} \times 100$$

Where:

$r_{i,j}$ is the geochemical ratio of element j in sample i

$c_{i,j}$ is the pXRF observed concentration of element j in sample i

j = {Al, Si, S, K, Ca, Ti, Fe}, i.e. the seven dominant observable elements

2.2.5.4 Cubist models

Predictive spectral models were constructed using Cubist (Quinlan, 1992). Briefly, Cubist utilises rule-based partitioning to split input spectra into subsets with similar characteristics. Rules may be based on a single or multiple wavelengths, and are arranged in a hierarchical structure. Linear regression models are constructed at terminal nodes in the hierarchy, and also at intermediate nodes. Predictions obtained at intermediate nodes are used to smooth predictions at subsequent nodes, and ultimately the final prediction at the terminal node.

A number of other techniques have been tested for spectral calibrations, including, partial least squares regression, random forest neural networks, support vector machines, etc. However, this thesis does not investigate an exhaustive list of models and spectral pre-treatments for modest gains in predictive performance. Instead it focuses on how any calibration model could be used to its full potential.

2.2.6 Validation statistics

A number of validation statistics were calculated to assess the performance of spectral calibration models and SINFERS predictions. While not all are necessary, all are provided so that results can be compared with other studies that offer a diversity of validation statistics.

2.2.6.1 Coefficient of determination

The coefficient of determination (R^2) is a measure of the amount of variance in the dependant variable explained by the independent variable, or variables. It considers the proportion of the residual sums of squares to the total sum of squares (Eqn 2.5).

Equation 2.5

$$R^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n (y_i - \bar{y})^2}$$

Where:

y_i is the i^{th} observed value

$\hat{y}_i$ is the i^{th} predicted value

$\bar{y}$ is the mean of variable y

It is one of the most commonly used metrics when considering model performance. However, it does not consider bias, and it can overstate the performance of a model if predictions do not lie on the 45° line. Values of 1 indicate a perfect fit, 0 indicates no relationship between the

variables. Negative values are possible with this metric and indicate that the mean value provides a better fit for the data than the model predictions.

2.2.6.2 Lin's concordance correlation coefficient (LCCC)

The Lin's Concordance Correlation Coefficient (LCCC) measures the degree of agreement between two variables, such as observed and predicted values (Lin, 1989). Unlike R^2, the LCCC is not immune to bias, and it assesses the fit to the 45° line. This can be demonstrated by considering a simplified example containing a vector of observed values, $y = \{1,2,3,\cdots,1000\}$, and three vectors of predicted values $\hat{y}_a = y, \hat{y}_b = \frac{y}{2}, \hat{y}_c = y + 100$. The R^2 in for each vector of predicted vectors would be 1, as the observed and predicted values form a perfectly straight line and R^2 does not account for bias as stated previously. In comparison LCCC values would be calculated as 1.00, 0.50 and 0.94 for $\hat{y}_a$, $\hat{y}_b$ and $\hat{y}_c$ respectively and is thus a better representation of the accuracy of the predictions. The LCCC is calculated as twice the covariance between the two variables divided by the sum of the variance of each variable and the difference between the mean of each variable (Eqn 2.6).

Equation 2.6

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

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

2.2.6.3 Mean-square error (MSE)

The mean-square error (RMSE) is a scale-dependent measure of accuracy that characterises the difference between observed and predicted values. The RMSE is calculated as the average squared error (Eqn 2.7).

Equation 2.7

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

Where:

y_i is the i^{th} observed value

$\hat{y}_i$ is the i^{th} predicted value

n is the number of observations

2.2.6.4 Root-mean-square error (RMSE)

The root-mean-square error (RMSE) is a scale-dependent measure of accuracy that characterises the difference between observed and predicted values. The RMSE is calculated as the square root of the average squared error (Eqn 2.8).

Equation 2.8

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

Where:

y_i is the i^{th} observed value

$\hat{y}_i$ is the i^{th} predicted value

n is the number of observations

2.2.6.5 Bias

Bias is calculated as the difference between the mean of predicted values and the mean of observed values (Eqn 2.9). It is a useful metric to discern if there is systematic over or under prediction, however it must be interpreted in conjunction with other metrics. For example, if

the mean value was given to all predictions, then the bias would equal zero, however the R^2 would also equal zero.

Equation 2.9

$$bias = \frac{\sum_i^n \hat{y}_i}{n} - \bar{y}$$

Where:

$\hat{y}_i$ is the i^{th} predicted value

$\bar{y}$ is the mean of variable y

2.2.6.6 Bias corrected mean-square error and root-mean-square error

Bias corrected factors first subtract the bias from predictions, see section 2.2.6.4, then calculate bias corrected mean-square error (MSE_c) and bias corrected mean-square error ($RMSE_c$) using the corrected values and MSE and RMSE equations outlined in sections 2.2.6.3 and 2.2.6.4. They offer a best-case scenario for using the model, if the bias correction holds with independent validation sets.

2.2.6.7 Ratio of performance to deviation (RPD)

The ratio of performance to deviation has been used to predict the goodness of fit of NIR calibration (Williams, 1987). It is calculated as the ratio of the standard deviation of a sample to the standard error of prediction (Eqn 2.10).

Equation 2.10

$$RPD = \frac{s}{RMSE}$$

Where:

s is the standard deviation of the independent variable

$RMSE$ is the root-mean-square error of the prediction, as defined in section 2.2.6.4

Some authors have called the use of RPD redundant when R^2 is also provided (Minasny and McBratney, 2007), as the two are directly proportional RPD = $(1\text{-}R^2)^{-0.5}$. Although this relationship only holds if the predicted fall around the 45° line, due to the issues outlined with R^2 in section 2.2.6.1. The RPD is provided in this thesis for comparison with other studies that have utilised it.

2.2.6.8 Ratio of performance to interquartile range (RPIQ)

The RPD has also received criticism as it assumes an underlying normal distribution. To account for the spread of data, without assuming an underlying normal distribution, Bellon-Maurel *et al*., (2010) suggested the ratio of performance to interquartile range (RPIQ).

Equation 2.11

$$RPIQ = \frac{s}{IQR}$$

Where:

s is the standard deviation of the independent variable

IQR is the interquartile range

2.3 Site descriptions

Fifteen soil profiles were analysed that exhibited a diverse range of soil properties and climates from across the state of New South Wales, Australia (Fig. 2.4).

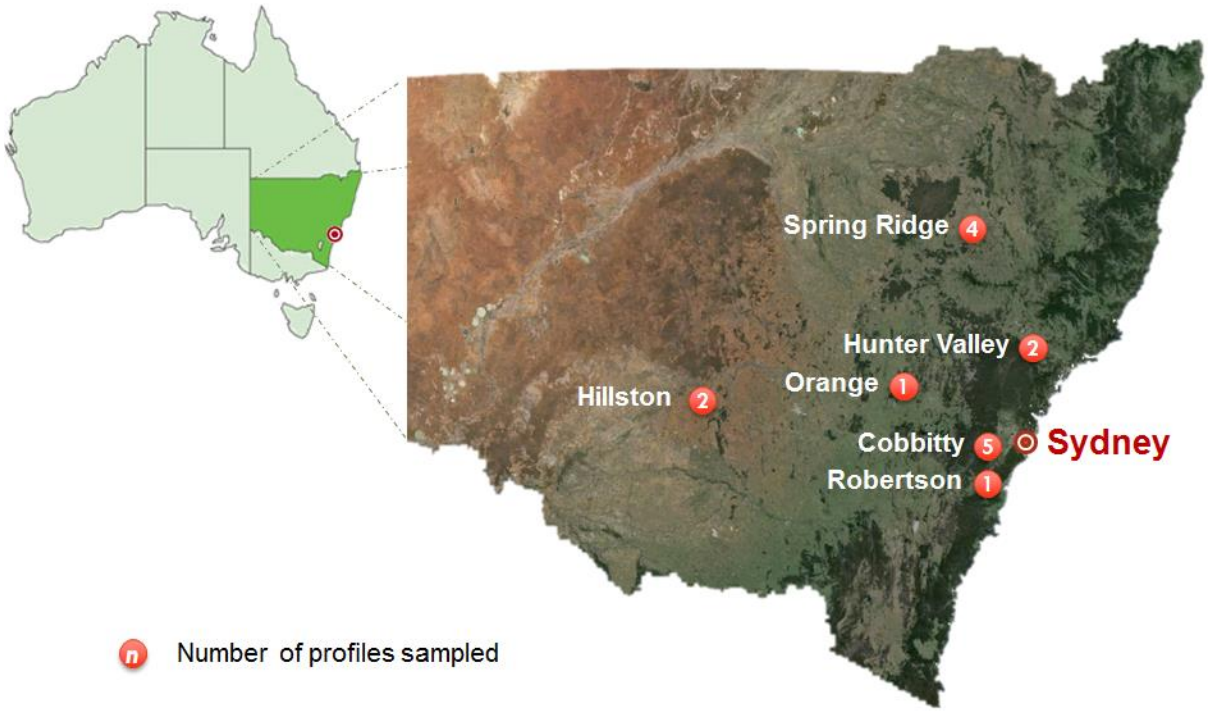


Figure 2.4 Location of soil profiles within the state of New South Wales. Numbers indicate the number of profiles sampled at each site.

The chosen profiles developed from a wide range of parent materials were identified, including marl, shale, mudstone, sandstone, basalt and trachyte. Soils developing from alluvium, residuum and æolian deposits were identified. Sites experienced a range of climates including, semi-arid, temperate and humid subtropical. Mean annual rainfall values ranged from 372 to 963 mm (BOM, 2017).

2.3.1 Description of each location

2.3.1.1 Hunter Valley

Sites 1 and 2 are located in the Hunter Valley, an agricultural and viticultural region. A warm temperate climate is experienced, with a mean annual precipitation of 763 mm (BOM, 2017). Both sites are from naturalised pastures.

2.3.1.2 Cobbitty

Sites 3, 5, 6, 13 and 15 are from University of Sydney experimental farms within a 3 km radius west and north-west of the township of Cobbitty. A humid subtropical climate is experienced, with a mean annual precipitation of 793 mm (BOM, 2017). Sites 3 and 15 are from improved, naturalised pastures. Site 5 is on a natural riverbank position. Sites 6 and 13 are subject to tillage as part of cropping experiments.

2.3.1.3 Spring Ridge

Sites 4, 10, 11 and 12 are from the University of Sydney experimental farm, "Nowley", ~12 km North West of the township of Spring Ridge. A warm temperate climate is experienced, with a mean annual precipitation of 621 mm (BOM, 2017). Sites 4 was from improved, naturalised pastures. Site 10 is from native vegetation that experiences periodic grazing. Sites 11 and 12 are from within cultivated fields. Primary production is wheat and sorghum.

2.3.1.4 Hillston

Sites 7 and 8 are from "Merrowie Station", ~6 km north of the township of Hillston. A semi-arid climate is experienced, with a mean annual precipitation of 372 mm (BOM, 2017). Site 7 is from conserved strip of native vegetation. Site 8 is within a cultivated field. Primary production is irrigated cotton.

2.3.1.5 Orange

Site 9 was located on a property near Nashdale, ~6 km west of the city of Orange. A cool temperate climate is experienced, with a mean annual precipitation of 929 mm (BOM, 2017). Land use at the site is improved, naturalised pasture.

2.3.1.6 Robertson

Site 14 was located on a property ~5 km south-west of the village of Robertson. A cool temperate climate is experienced, with a mean annual precipitation of 963 mm (BOM, 2017). The profile face was formed on an active erosion site, directly adjacent to a cultivated field.

2.3.2 Description of each profile

Each profile was described using routine field observations (section 2.3.2.1-15). Laboratory data are displayed from horizon-based sampling. The 400–750 nm range of VisNIR spectra are displayed for comparison. The spectra are coloured by sample colour estimations derived directly from the spectra (Viscarra Rossel *et al.*, 2009). Briefly, the average reflectance from red (R, 600–690 nm), green (G, 520-600 nm) and blue (B, 450-520 nm) ranges of the spectra are scaled appropriately and used to construct the colours in RGB space. Elemental composition from pXRF is displayed as the ratio of Si, Al, S, K, Ca, Ti and Fe with depth. The three individual vertical transects are displayed as thin lines, to give an appreciation of lateral variability. The average of the three transect is displayed as the wider line.

2.3.2.1 Field observations, sensor readings and laboratory data of Site 1

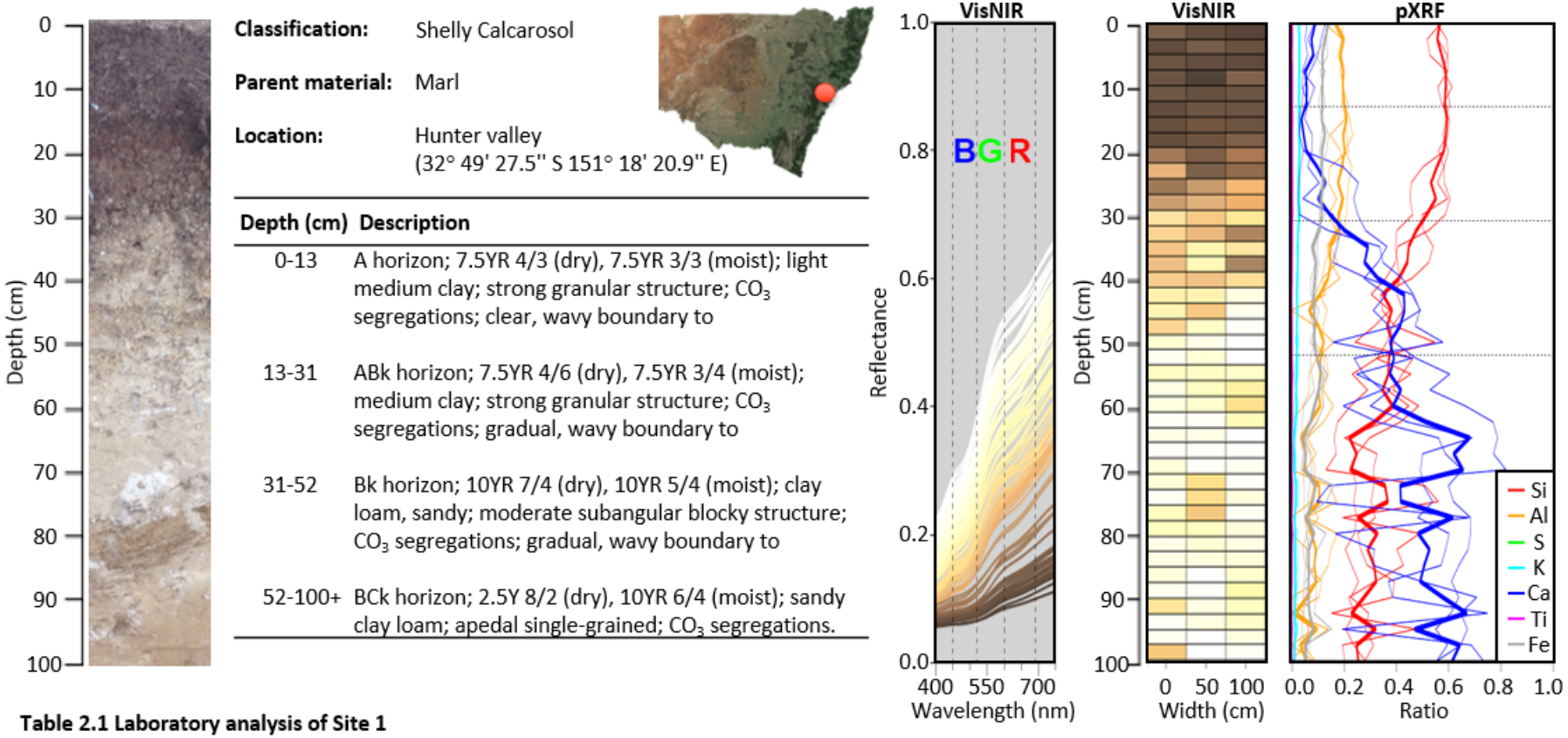


Table 2.1 Laboratory analysis of Site 1

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations (cmol$_{(+)}$ kg^{-1})					CEC (cmol$_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	13	47.1	17.3	35.6	16.5	19.1	8.22	7.49	0.14	2.90	0.184	2.64	2.6	16.0	0.64	0.7	0.04	-	17.0
ABk	13	31	50.1	15.6	34.3	22.3	12.0	8.26	7.54	0.11	1.96	0.109	1.65	3.1	15.0	0.18	0.5	0.02	-	16.0
Bk	31	52	37.7	18.8	43.5	39.1	4.4	8.44	7.72	0.12	2.44	0.058	0.88	19.0	13.0	0.01	0.4	0.02	-	13.0
BCk	52	100	22.0	20.2	57.8	40.6	17.2	8.77	7.76	0.09	5.64	0.032	0.40	35.0	12.0	0.03	0.4	0.02	-	12.0

2.3.2.2 Field observations, sensor readings and laboratory data of Site 2

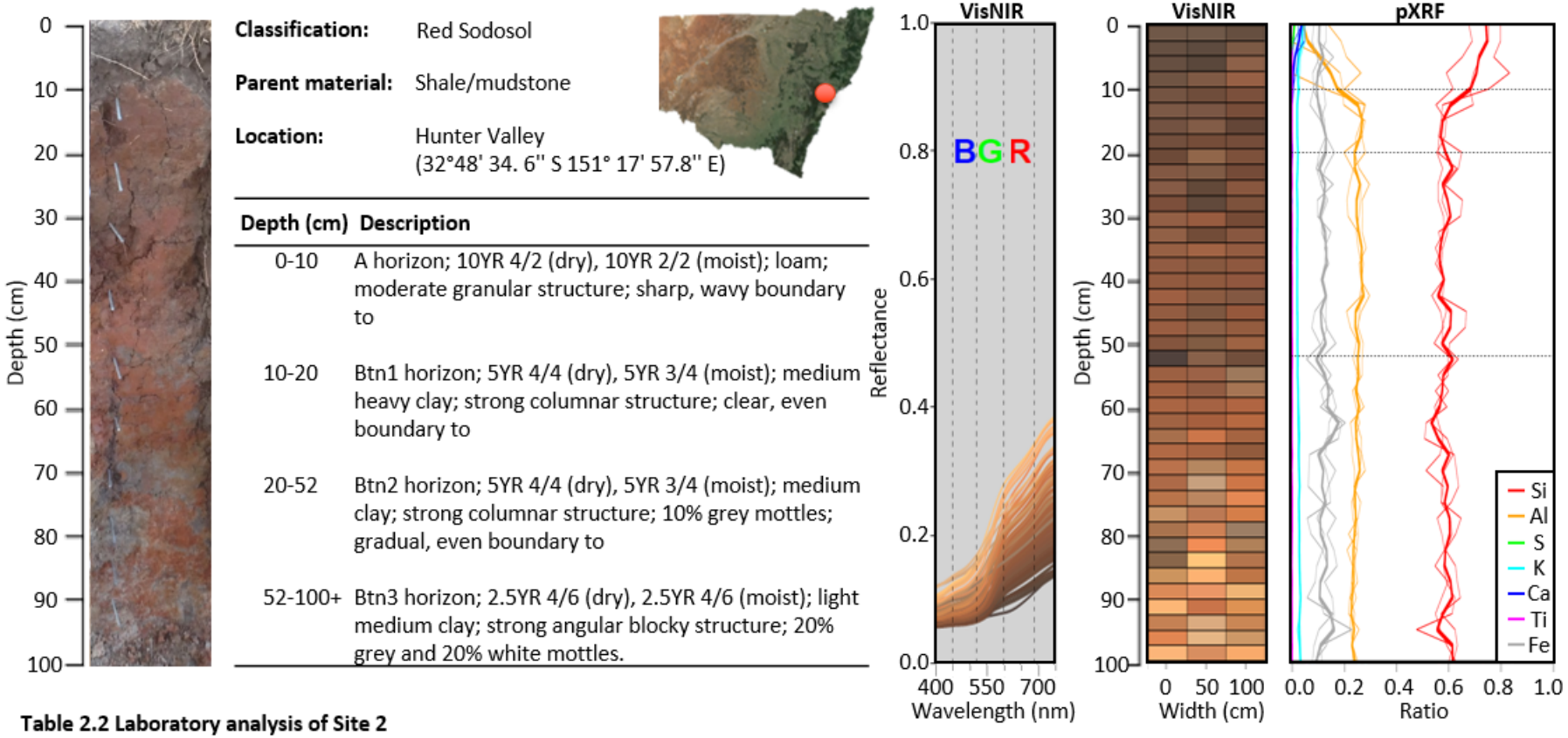
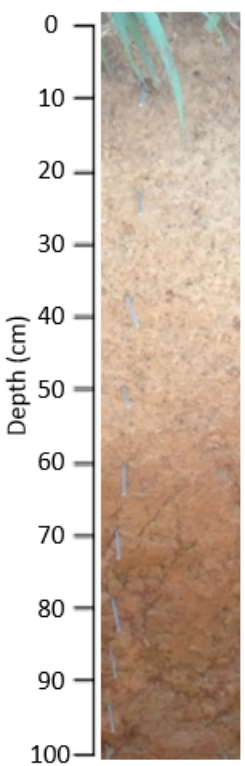


Table 2.2 Laboratory analysis of Site 2

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	10	23.7	16.7	59.6	9.9	49.6	5.23	4.49	0.13	4.15	0.265	4.07	-	4.7	0.34	4.5	0.44	0.6	10.5
Btn1	10	20	64.5	5.4	30.1	26.0	4.0	5.77	4.46	0.08	0.96	0.049	0.71	-	3.7	0.49	12.0	1.30	1.5	18.6
Btn2	20	52	59.5	8.1	32.4	30.7	1.7	5.53	4.29	0.11	0.77	0.041	0.53	-	1.8	0.36	12.0	1.90	2.1	18.3
Btn3	52	100	53.0	10.0	37.0	15.5	21.5	4.79	3.93	0.43	0.37	0.039	0.25	-	0.4	0.27	11.0	3.70	2.1	17.8

2.3.2.3 Field observations, sensor readings and laboratory data of Site 3

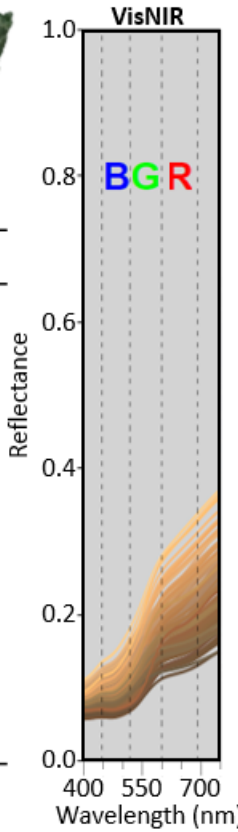


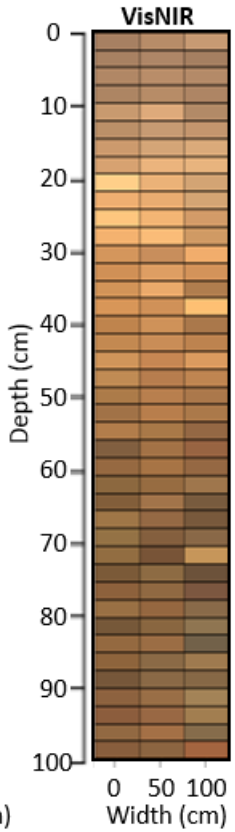
Classification: Brown Kurosol

Parent material: Ashfield shale

Location: Cobbitty
(33° 59' 44.9" S 150° 39' 11.9" E)

Depth (cm)	Description
0-16	A horizon; 10YR 6/4 (dry), 10YR 3/4 (moist); sandy loam; moderate granular structure; 2-4 mm diameter magnetite/maghemite; clear, wavy boundary to
16-33	E horizon; 7.5YR 6/4 (dry), 7.5YR 4/4 (moist); sandy clay loam; strong subangular blocky structure; gradual, wavy boundary to
33-47	Bt1 horizon; 7.5YR 5/4 (dry), 7.5YR 4/6 (moist); medium heavy clay; strong angular blocky structure; gradual, wavy boundary to
47-100+	Bt2 horizon; 7.5YR 5/6 (dry), 7.5YR 4/6 (moist); medium heavy clay; strong angular blocky structure; 25% red and 25% grey mottles.





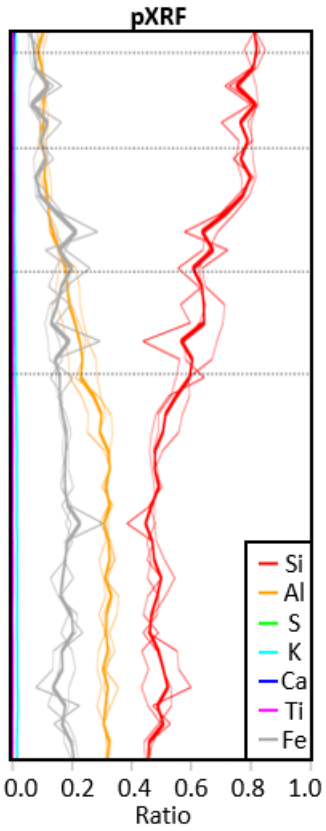


Table 2.3 Laboratory analysis of Site 3

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations (cmol$_{(+)}$ kg^{-1})					CEC (cmol$_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	16	17.9	13.6	68.4	35.3	33.1	5.99	4.84	0.06	0.96	0.043	0.65	-	1.2	0.51	1.1	0.12	0.2	3.1
E	16	33	24.2	11.7	64.1	25.8	38.2	5.69	4.57	0.06	0.69	0.033	0.47	-	0.9	0.28	2.0	0.17	0.5	3.9
Bt1	33	47	57.4	10.4	32.1	20.3	11.8	5.43	4.22	0.05	0.81	0.044	0.62	-	1.5	0.42	5.2	0.41	1.3	8.9
Bt2	47	100	66.7	11.4	21.9	19.8	2.1	5.34	4.00	0.07	0.37	0.002	0.27	-	0.2	0.15	5.1	0.78	7.2	13.5

2.3.2.4 Field observations, sensor readings and laboratory data of Site 4

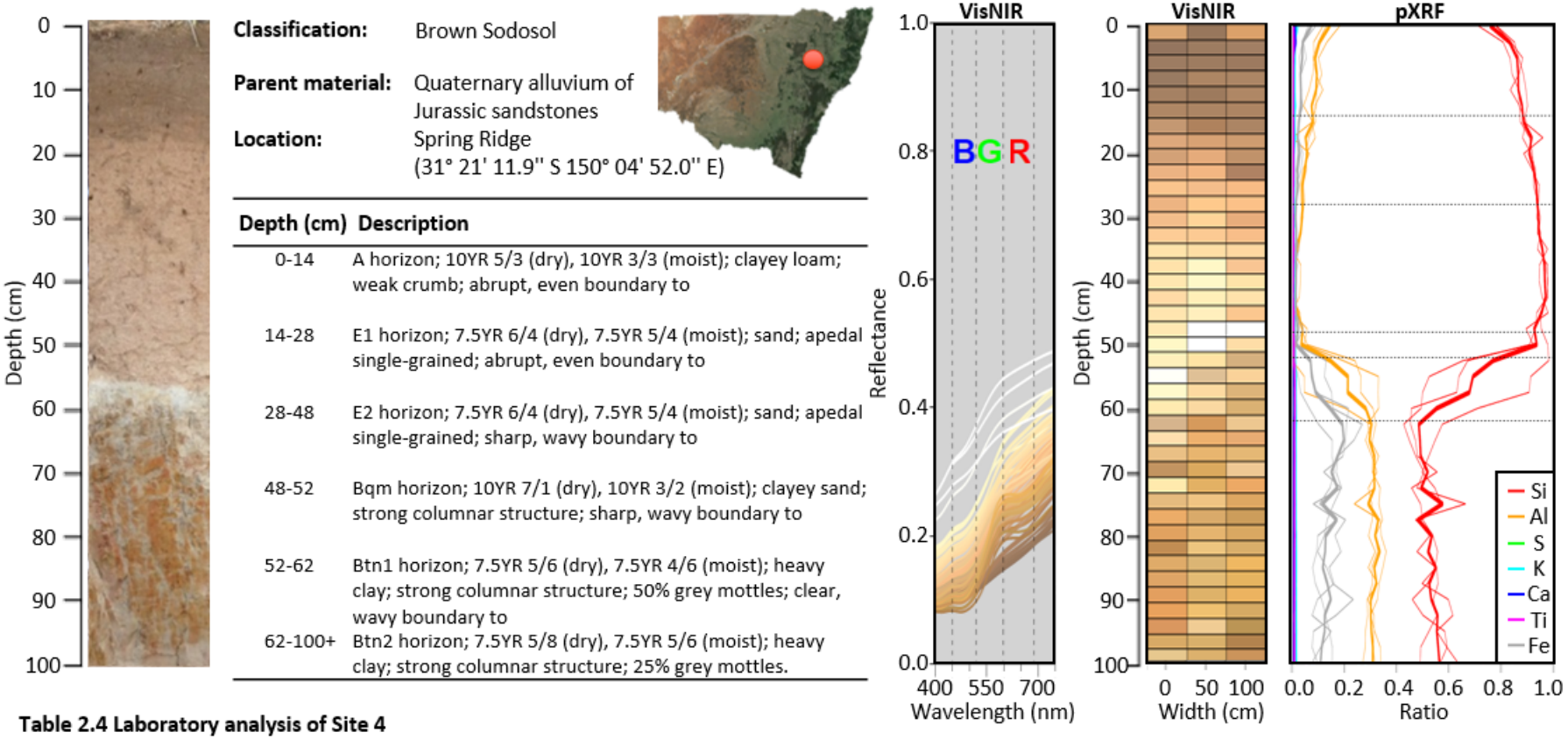


Table 2.4 Laboratory analysis of Site 4

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	14	8.4	2.5	89.1	14.4	74.7	6.09	4.83	0.05	0.86	0.038	0.74	-	1.4	0.09	0.8	0.19	0.1	2.6
E1	14	28	5.4	1.0	93.6	20.0	73.6	6.51	4.96	0.01	0.19	0.004	0.00	-	0.4	0.03	0.2	0.04	0.0	0.6
E2	28	48	4.8	1.0	94.2	18.5	75.7	6.23	5.25	0.01	0.09	0.002	0.00	-	0.1	0.00	0.1	0.06	0.0	0.3
Bqm	48	52	16.2	9.4	74.4	29.9	44.5	7.15	5.60	0.03	0.10	0.004	0.00	-	0.2	0.00	0.6	0.23	0.0	1.0
Btn1	52	62	80.4	2.3	17.3	5.1	12.2	6.20	4.93	0.10	0.25	0.005	0.20	-	1.6	0.18	11.0	2.80	0.2	15.5
Btn2	62	100	65.4	1.6	33.1	8.4	24.7	6.53	5.11	0.09	0.13	0.003	0.00	-	1.0	0.14	8.8	2.60	0.0	12.5

2.3.2.5 Field observations, sensor readings and laboratory data of Site 5

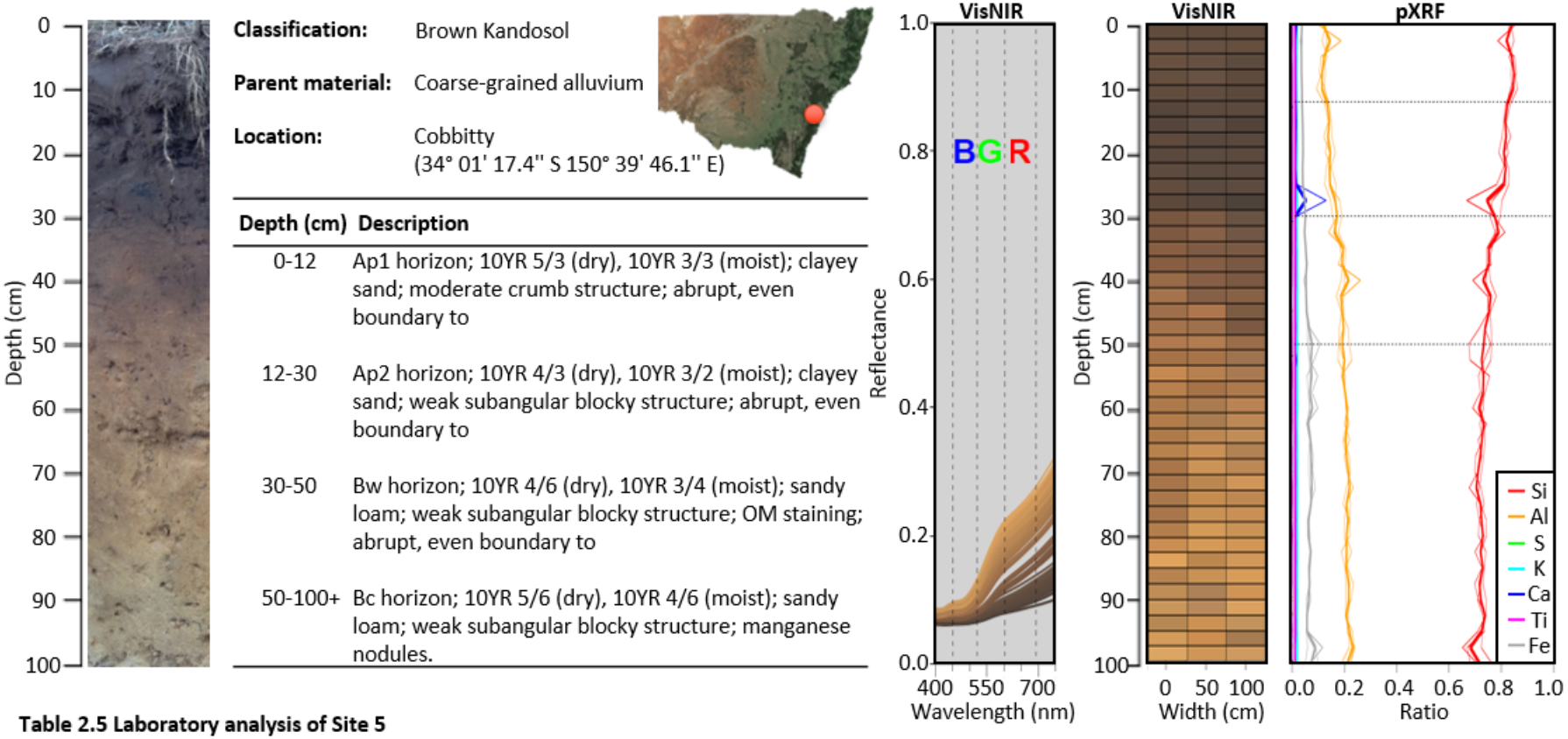
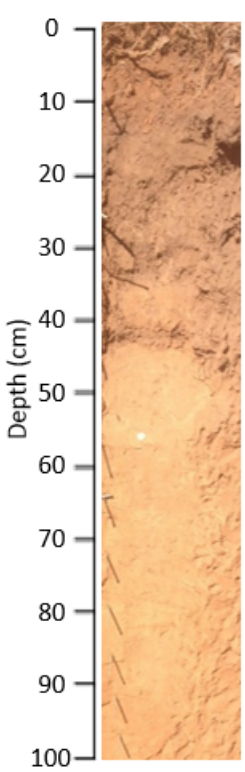


Table 2.5 Laboratory analysis of Site 5

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
Ap1	0	12	10.2	3.6	86.3	13.6	72.7	6.44	5.58	0.07	0.54	0.034	0.52	-	1.9	0.29	0.4	0.00	0.0	2.6
Ap2	12	30	10.7	4.8	84.5	13.5	71.0	6.73	5.76	0.13	0.41	0.023	0.32	-	2.5	0.23	0.1	0.04	0.0	2.9
Bw	30	50	14.4	3.7	81.9	13.2	68.7	6.72	5.68	0.03	0.38	0.015	0.20	-	2.0	0.25	0.4	0.02	0.1	2.9
Bc	50	100	16.6	4.5	78.9	11.9	67.0	6.74	5.95	0.03	0.18	0.016	0.18	-	2.0	0.11	0.4	0.03	0.0	2.6

2.3.2.6 Field observations, sensor readings and laboratory data of Site 6



Classification: Red Tenosol

Parent material: Coarse-grained alluvium

Location: Cobbitty
(34° 01' 35.8" S 150° 39' 43.6" E)

Depth (cm)	Description
0-16	A horizon; 7.5YR 4/3 (dry), 7.5YR 3/2 (moist); sandy clay loam; moderate crumb structure; abrupt, wavy boundary to
16-31	Bw1 horizon; 7.5YR 4/4 (dry), 7.5YR 3/4 (moist); clay loam, sandy; strong angular blocky structure; clear, wavy boundary to
31-48	Bw2 horizon; 7.5YR 4/6 (dry), 5YR 3/4 (moist); clay loam, sandy; strong angular blocky structure; gradual, even boundary to
48-73	2Bw1 horizon; 5YR 4/6 (dry), 5YR 3/4 (moist); sandy clay loam; strong angular blocky structure; gradual, even boundary to
73-100+	2Bw2 horizon; 7.5YR 4/6 (dry), 5YR 3/4 (moist); sandy loam; moderate subangular blocky structure.

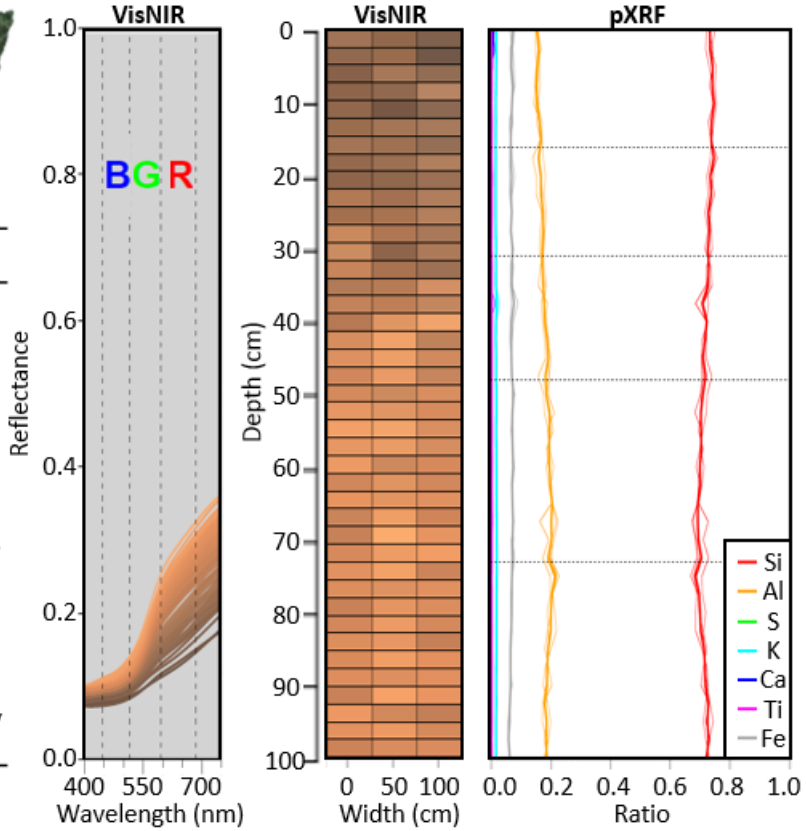
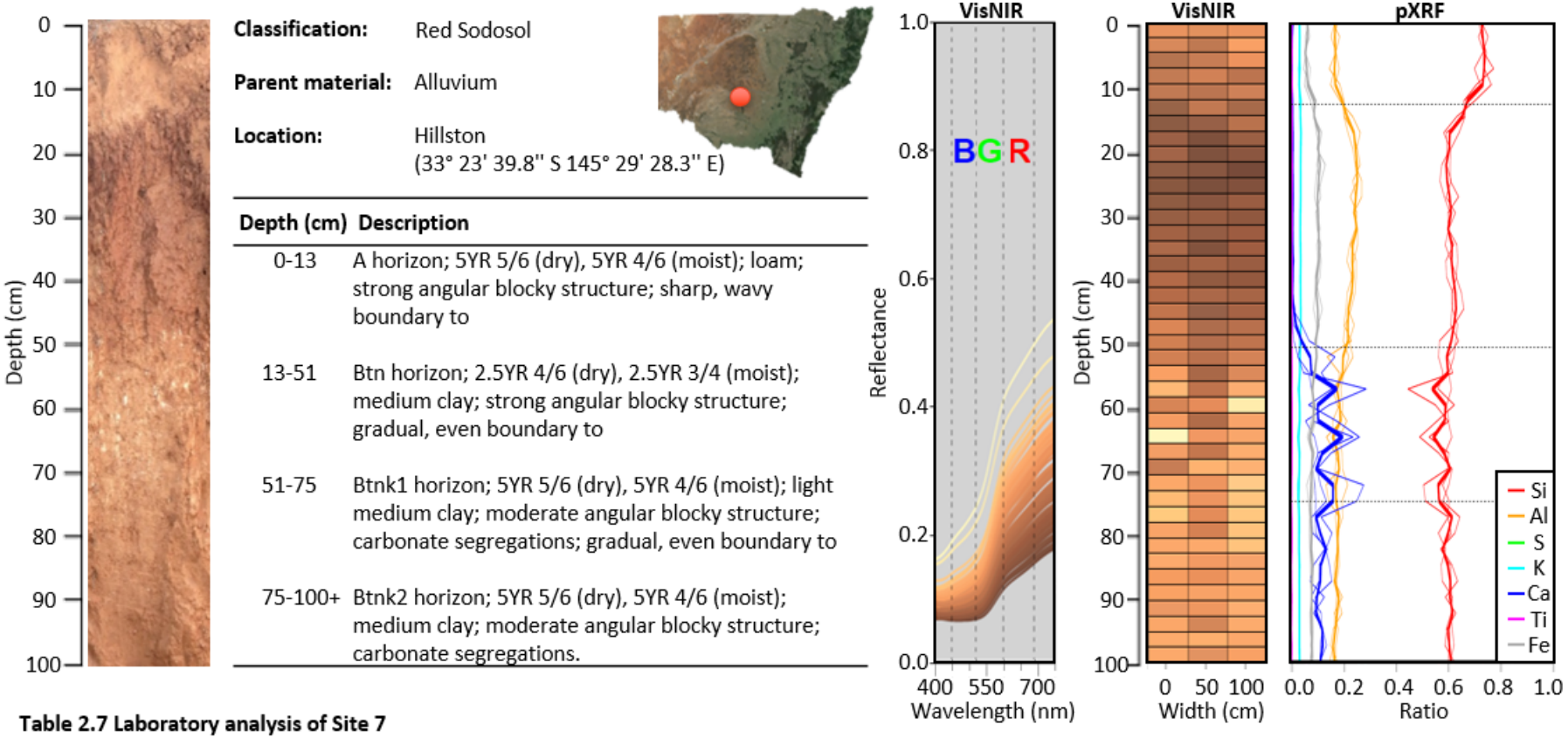


Table 2.6 Laboratory analysis of Site 6

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	16	18.4	12.5	69.1	49.5	19.6	6.09	5.08	0.10	3.26	0.199	2.97	-	5.3	1.30	2.7	0.16	0.0	9.5
Bw1	16	31	21.1	11.5	67.4	42.1	25.3	5.55	4.74	0.08	1.16	0.051	0.90	-	1.7	1.00	1.2	0.14	0.2	4.3
Bw2	31	48	22.8	7.8	69.4	46.5	22.8	6.34	5.17	0.06	0.36	0.017	0.24	-	1.1	1.10	0.7	0.20	0.0	3.0
2Bw1	48	73	18.4	5.5	76.1	35.6	40.6	6.42	5.24	0.05	0.19	0.007	0.00	-	0.9	1.00	0.4	0.13	0.0	2.4
2Bw2	73	100	12.4	4.4	83.2	35.8	47.4	6.60	5.38	0.06	0.11	0.005	0.00	-	0.2	1.10	0.1	0.07	0.1	1.6

2.3.2.7 Field observations, sensor readings and laboratory data of Site 7



Classification: Red Sodosol

Parent material: Alluvium

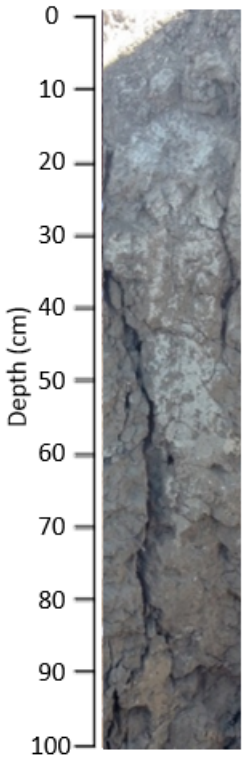
Location: Hillston
(33° 23' 39.8'' S 145° 29' 28.3'' E)

Depth (cm)	Description
0-13	A horizon; 5YR 5/6 (dry), 5YR 4/6 (moist); loam; strong angular blocky structure; sharp, wavy boundary to
13-51	Btn horizon; 2.5YR 4/6 (dry), 2.5YR 3/4 (moist); medium clay; strong angular blocky structure; gradual, even boundary to
51-75	Btnk1 horizon; 5YR 5/6 (dry), 5YR 4/6 (moist); light medium clay; moderate angular blocky structure; carbonate segregations; gradual, even boundary to
75-100+	Btnk2 horizon; 5YR 5/6 (dry), 5YR 4/6 (moist); medium clay; moderate angular blocky structure; carbonate segregations.

Table 2.7 Laboratory analysis of Site 7

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	13	17.7	6.4	75.9	37.3	38.6	6.38	5.73	0.11	0.30	0.027	0.32	0.0	2.4	0.79	1.6	0.02	-	4.8
Btn	13	51	56.2	4.8	39.0	20.8	18.2	7.43	7.31	0.20	0.42	0.036	0.35	0.0	7.0	0.84	7.1	1.20	-	16.0
Btnk1	51	75	55.0	5.8	39.2	25.6	13.5	9.57	8.30	0.29	1.43	0.002	0.00	10.0	6.0	0.64	7.5	2.20	-	16.0
Btnk2	75	100	52.7	5.1	42.2	28.2	14.0	9.32	8.29	0.53	1.21	0.001	0.00	8.9	5.5	0.61	7.7	2.20	-	16.0

2.3.2.8 Field observations, sensor readings and laboratory data of Site 8



Classification: Grey Vertosol

Parent material: Alluvium

Location: Hillston (33° 24' 33.1" S 145° 33' 58.0" E)

Depth (cm)	Description
0-2	A horizon; 10YR 5/1 (dry), 10YR 4/1 (moist); light medium clay; strong granular structure; sharp, even boundary to
2-7	Bss1 horizon; 10YR 5/1 (dry), 10YR 4/2 (moist); medium heavy clay; strong lenticular and angular blocky structure; gradual, even boundary to
7-28	Bss2 horizon; 10YR 3/1 (dry), 10YR 4/1 (moist); medium heavy clay; strong lenticular structure; carbonate segregations; gradual, even boundary to
28-100+	Bn horizon; 10YR 6/2 (dry), 10YR 5/3 (moist); medium heavy clay; strong angular blocky structure; carbonate segregations.

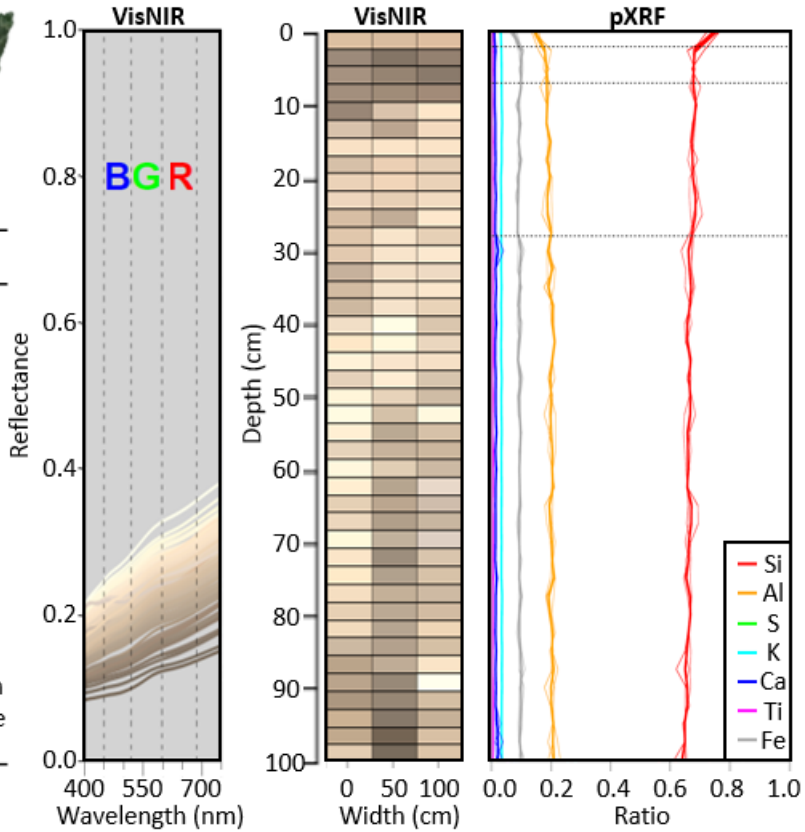
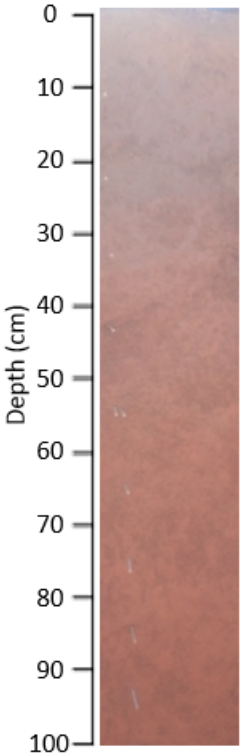


Table 2.8 Laboratory analysis of Site 8

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations (cmol$_{(+)}$ kg^{-1})					CEC (cmol$_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	2	60.9	12.8	26.3	22.5	3.8	6.85	6.35	0.21	0.66	0.051	0.73	0.0	9.5	1.10	6.9	0.32		18.0
Bss	2	7	61.9	12.7	25.3	22.0	3.3	7.27	6.27	0.08	0.58	0.045	0.60	0.0	10.0	0.72	7.2	0.38	-	18.0
Bssk	7	28	67.0	12.1	21.0	17.9	3.1	8.51	7.43	0.19	0.52	0.006	0.19	0.9	10.0	0.38	6.2	1.10	-	18.0
Bnk	28	100	63.6	13.5	22.9	17.9	5.0	8.81	7.75	0.27	0.17	0.002	0.00	1.1	8.5	0.49	5.3	2.40	-	17.0

2.3.2.9 Field observations, sensor readings and laboratory data of Site 9



Classification: Red Dermosol

Parent material: Polygenetic, æolian over Residuum, incl. trachyte

Location: Orange (33° 17' 21.0" S 149° 01' 35.2" E)

Depth (cm)	Description
0-18	A horizon; 10YR 5/3 (dry), 10YR 3/4 (moist); silty loam; strong subangular structure; clear, even boundary to
18-40	Bt horizon; 7.5YR 5/6 (dry), 10YR 4/6 (moist); light medium clay; strong subangular structure; infilled macropores; clear, even boundary to
40-68	2Bt1 horizon; 5YR 4/6 (dry), 5YR 3/4 (moist); medium clay; strong polyhedral structure; conspicuous trachyte saprolite ~300 – 400 mm diameter; diffuse, even boundary to
68-100+	2Bt2 horizon; 5YR 4/6 (dry), 5YR 3/4 (moist); medium heavy clay; strong polyhedral structure; conspicuous trachyte saprolite ~300 – 400 mm diameter.

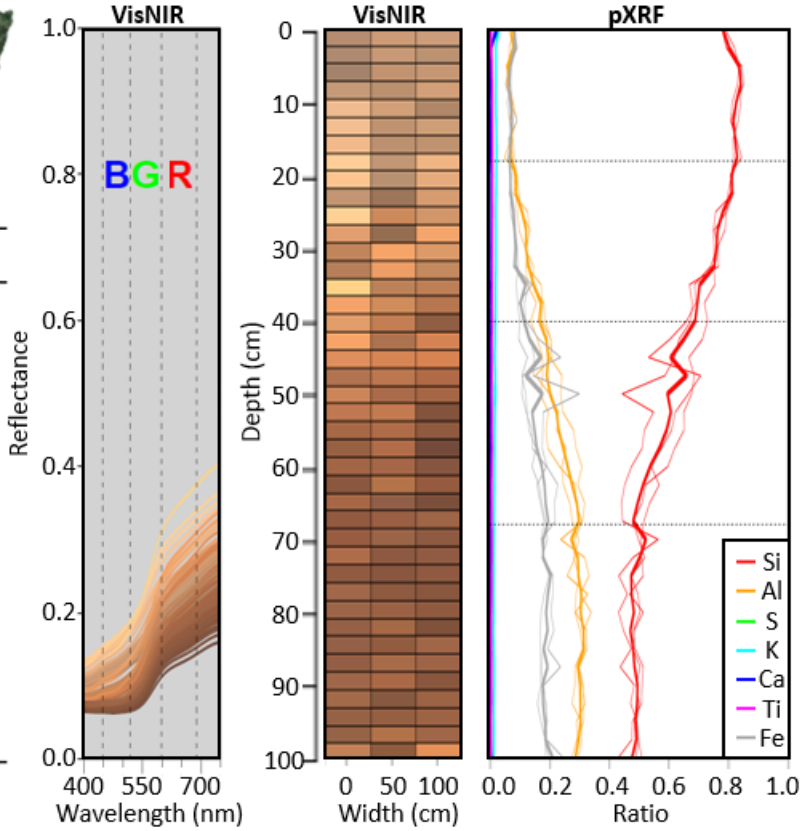
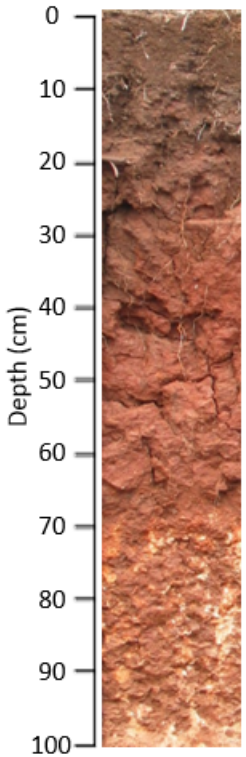


Table 2.9 Laboratory analysis of Site 9

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	18	14.4	26.1	59.5	52.1	7.3	5.98	5.09	0.05	2.16	0.127	1.76	-	5.0	0.59	0.9	0.00	0.0	6.5
Bt	18	40	30.9	21.7	47.4	41.7	5.7	6.28	5.24	0.02	0.59	0.029	0.46	-	4.3	0.61	0.9	0.00	0.1	5.9
2Bt1	40	68	53.4	14.5	32.1	26.6	5.5	6.34	5.40	0.03	0.39	0.022	0.33	-	6.9	0.37	2.1	0.08	0.0	9.4
2Bt2	68	100	62.9	14.0	23.1	19.0	4.1	6.59	5.60	0.01	0.29	0.003	0.23	-	8.5	0.69	3.2	0.06	0.0	12.4

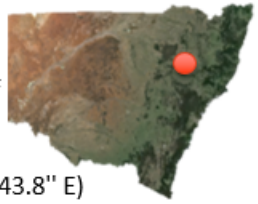
2.3.2.10 Field observations, sensor readings and laboratory data of Site 10



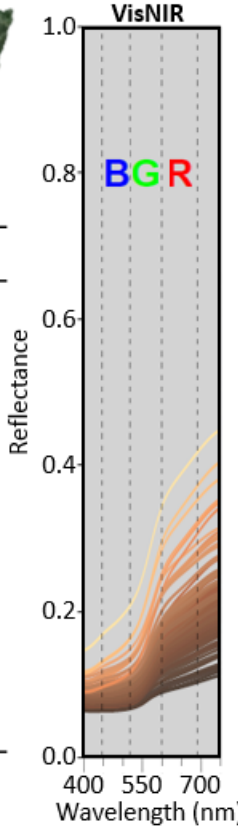
Classification: Red Chromosol

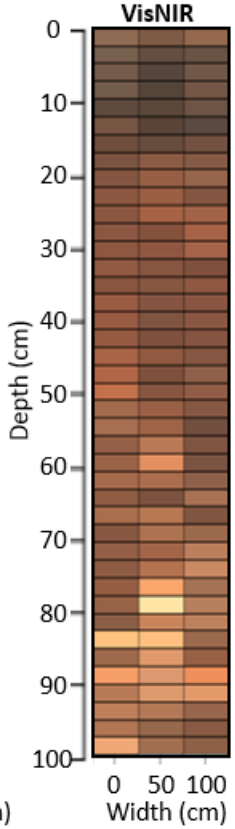
Parent material: Weathering/colluvium of mudstones

Location: Spring Ridge (31° 21' 16.0'' S 150° 06' 43.8'' E)



Depth (cm)	Description
0-10	A11 horizon; 7.5YR 3/4 (dry), 5YR 2.5/2 (moist); light clay; strong angular blocky structure; abrupt, even boundary to
10-30	A12 horizon; 5YR 3/3 (dry), 5YR 3/3 (moist); light medium clay; strong angular blocky structure; abrupt, even boundary to
30-53	Btkss horizon; 5YR 4/6 (dry), 2.5YR 4/6 (moist); medium clay; strong lenticular and angular blocky structure; carbonate segregations; gradual, wavy boundary to
53-97	Btk1 horizon; 5YR 4/4 (dry), 2.5YR 4/6 (moist); light clay; moderate subangular blocky structure; carbonate segregations; gradual, wavy boundary to
97-100+	Btk2 horizon; 5YR 4/4 (dry), 2.5YR 4/6 (moist); light medium clay; moderate subangular blocky structure; carbonate segregations.





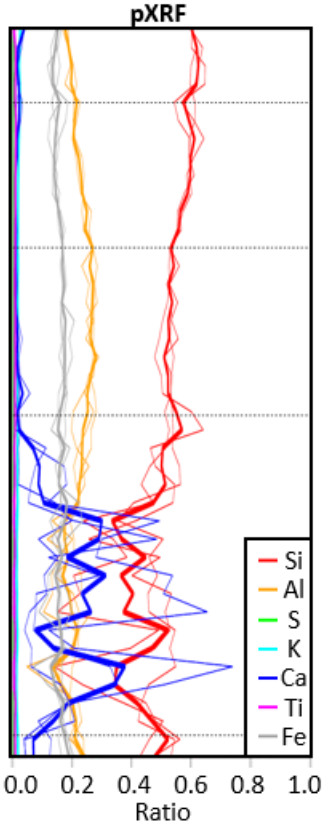
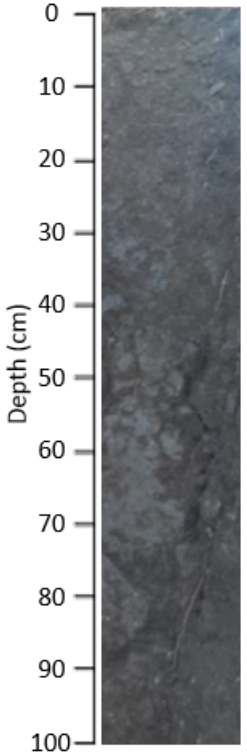


Table 2.10 Laboratory analysis of Site 10

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations (cmol$_{(+)}$ kg^{-1})					CEC (cmol$_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A11	0	10	38.5	2.7	58.8	15.0	43.8	6.61	6.13	0.16	1.35	0.082	1.43	0.0	7.0	1.40	5.2	0.02	-	14.0
A12	10	30	37.6	5.4	57.1	10.7	46.3	7.38	6.46	0.05	0.72	0.033	0.79	0.0	6.5	0.44	5.0	0.03	-	12.0
Btss	30	53	53.1	4.9	41.9	8.6	33.3	8.51	7.75	0.14	0.42	0.024	0.30	1.0	8.0	0.28	9.9	0.17	-	18.0
Btk1	53	97	52.5	8.7	38.8	12.2	26.6	8.34	7.94	0.61	2.44	0.005	0.23	16.0	8.5	0.25	12.0	0.26	-	21.0
Btk2	97	100	35.4	5.4	59.2	12.8	46.4	8.80	8.11	0.27	0.34	0.002	0.00	2.0	6.5	0.16	11.0	0.48	-	18.0

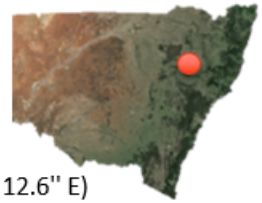
2.3.2.11 Field observations, sensor readings and laboratory data of Site 11



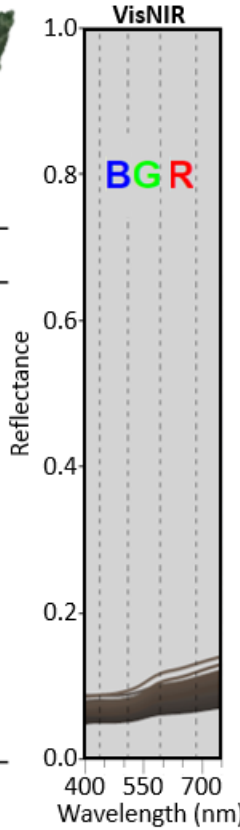
Classification: Black Vertosol

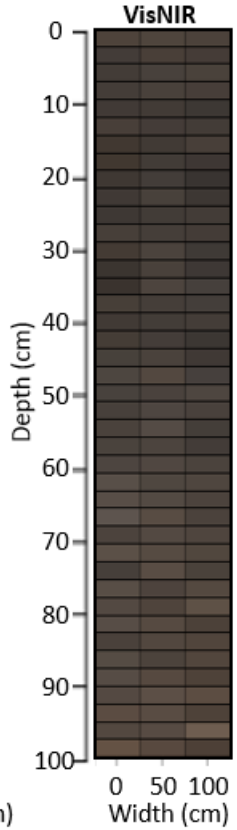
Parent material: Weathered Jurassic basalt

Location: Spring Ridge (31° 21' 34.3'' S 150° 08' 12.6'' E)



Depth (cm)	Description
0-5	Ak horizon; 5YR 3/1 (dry), 5YR 2.5/1 (moist); medium heavy clay; strong granular structure; carbonates; abrupt, even boundary to
5-22	Bk1 horizon; 5YR 3/1 (dry), 5YR 2.5/1 (moist); medium clay; strong angular blocky structure; carbonates; gradual, even boundary to
22-73	Bkss horizon; 5YR 3/1 (dry), 5YR 2.5/1 (moist); medium heavy clay; strong lenticular structure; carbonates; gradual, wavy boundary to
73-100+	Bk2 horizon; 7.5YR 3/2 (dry), 7.5YR 2.5/2 (moist); medium heavy clay; strong lenticular structure; carbonates.





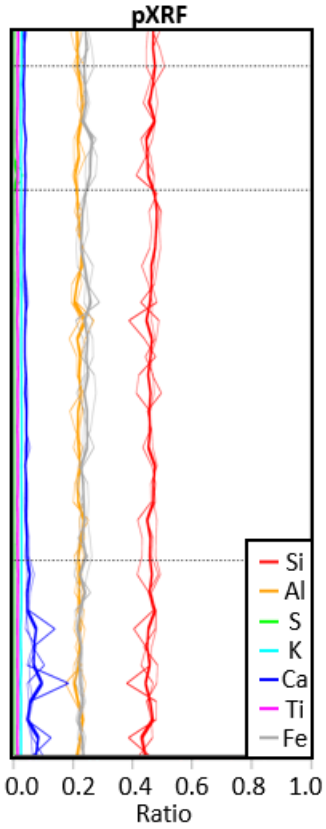
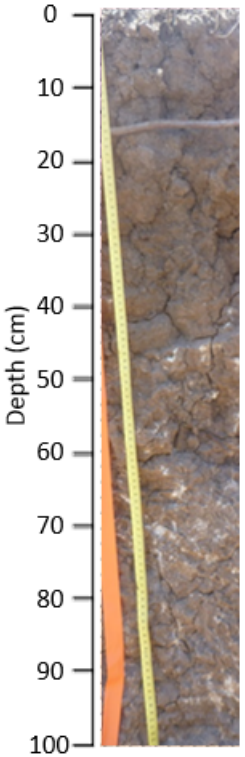


Table 2.11 Laboratory analysis of Site 11

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
Ak	0	5	63.4	16.6	20.0	10.6	9.4	7.40	6.49	0.08	1.89	0.117	1.98	1.2	13.0	1.40	13.0	0.22	-	28.0
Bk1	5	22	71.7	13.8	14.5	9.7	4.9	7.56	6.55	0.06	1.59	0.073	1.54	1.1	12.0	0.64	12.0	0.29	-	25.0
Bkss	22	73	69.9	14.8	15.3	9.4	5.9	8.37	7.03	0.08	1.53	0.054	1.32	1.2	11.0	0.36	13.0	0.83	-	25.0
Bk2	73	100	67.0	10.6	22.4	13.0	9.3	8.77	7.94	0.17	1.13	0.008	0.59	3.5	12.0	0.44	15.0	1.00	-	28.0

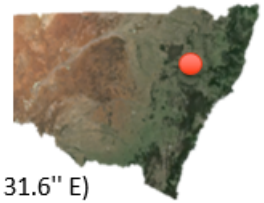
2.3.2.12 Field observations, sensor readings and laboratory data of Site 12



Classification: Brown Vertosol

Parent material: Basalt and mudstone colluvium

Location: Spring Ridge (31° 19' 54.0" S 150° 05' 31.6" E)



Depth (cm)	Description
0-30	Apn horizon; 10YR 4/2 (dry), 10YR 3/2 (moist); medium clay; strong angular blocky structure; clear, even boundary to
30-39	Bknss horizon; 10YR 4/2 (dry), 10YR 3/2 (moist); medium heavy clay; strong lenticular and angular blocky structure; carbonates; clear, even boundary to
39-66	Bkn horizon; 10YR 4/2 (dry), 10YR 4/3 (moist); medium clay; strong lenticular and polyhedral structure; carbonates; gradual, wavy boundary to
66-100+	Bny horizon; 7.5YR 4/3 (dry), 7.5YR 4/3 (moist); medium clay; strong lenticular and polyhedral structure; gypsum.

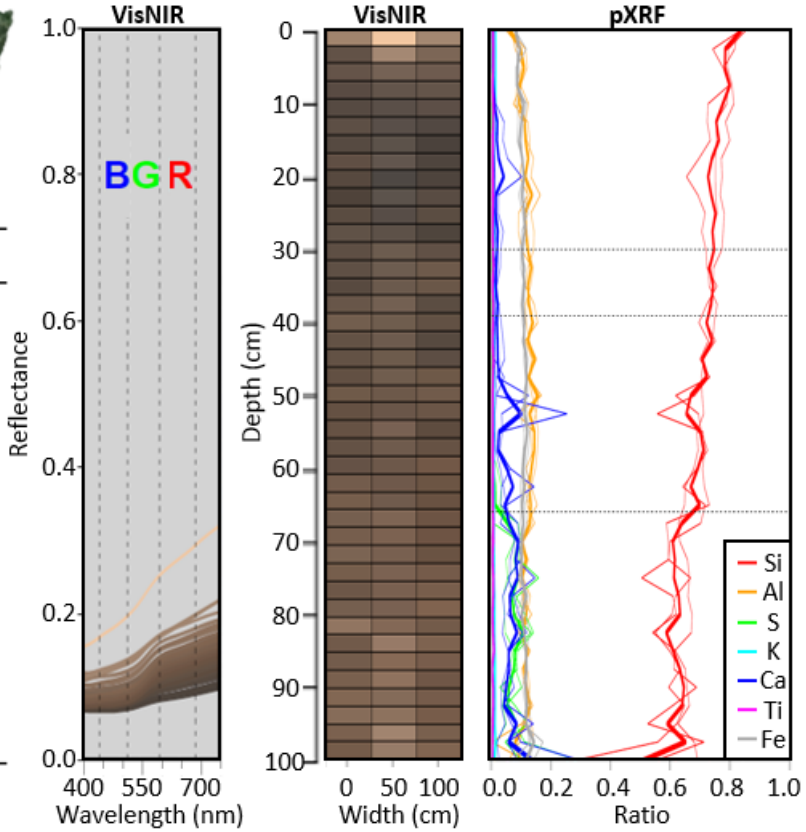
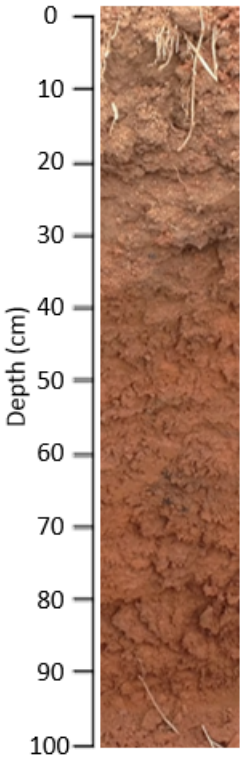


Table 2.12 Laboratory analysis of Site 12

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations (cmol$_{(+)}$ kg^{-1})					CEC (cmol$_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
Apn	0	30	49.5	16.3	34.2	10.2	24.1	7.18	5.88	0.07	1.01	0.050	0.78	0.0	7.5	0.33	9.1	1.20	-	18.0
Bknss	30	39	58.4	14.6	26.9	7.7	19.2	9.24	8.09	0.25	0.52	0.021	0.37	1.5	9.0	0.22	11.0	2.10	-	22.0
Bkn	39	66	60.9	13.2	25.9	6.0	19.9	9.28	8.17	0.28	0.42	0.006	0.29	0.8	8.5	0.18	12.0	2.60	-	23.0
Bny	66	100	70.3	6.6	23.0	5.4	17.6	7.76	7.83	3.49	0.13	0.002	0.00	0.8	18.0	0.19	11.0	1.50	-	31.0

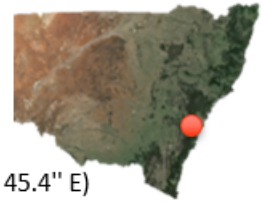
2.3.2.13 Field observations, sensor readings and laboratory data of Site 13



Classification: Red Dermosol

Parent material: Fine-grained alluvium

Location: Cobbitty
(34° 01' 25.2'' S 150° 39' 45.4'' E)



Depth (cm)	Description
0-23	Ap horizon; 7.5YR 4/4 (dry), 7.5YR 3/4 (moist); clay loam; moderate subangular blocky structure; abrupt, even boundary to
23-45	Bt1 horizon; 2.5YR 4/6 (dry), 2.5YR 3/4 (moist); light medium clay; strong angular blocky structure; gradual, even boundary to
45-100+	Bt2 horizon; 2.5YR 4/6 (dry), 2.5YR 3/4 (moist); light medium clay; strong angular blocky structure.

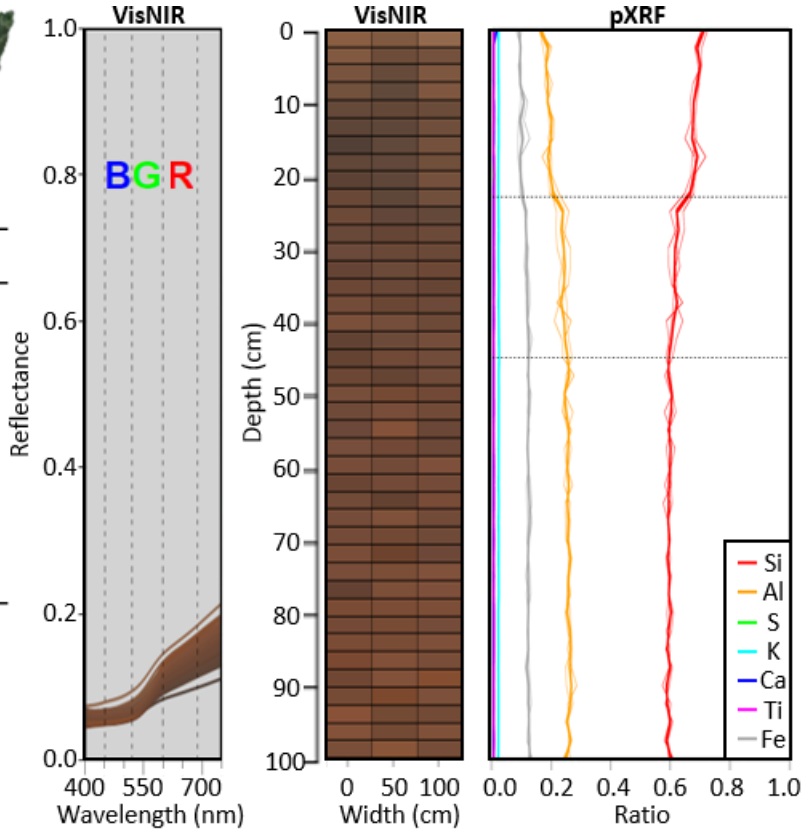
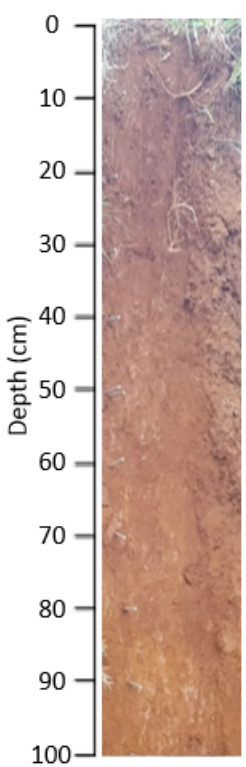


Table 2.13 Laboratory analysis of Site 13

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
Ap	0	23	28.2	17.8	54.0	10.4	43.6	6.40	5.41	0.06	1.06	0.059	0.79	-	4.9	0.72	1.3	0.11	0.0	7.1
Bt1	23	45	37.2	18.6	44.2	20.2	24.0	7.05	6.03	0.03	0.54	0.031	0.43	-	5.7	0.24	1.4	0.14	0.0	7.5
Bt2	45	100	38.9	18.0	43.1	9.3	33.8	6.94	6.19	0.04	0.22	0.006	0.00	-	5.5	0.12	1.3	0.20	0.0	7.2

2.3.2.14 Field observations, sensor readings and laboratory data of Site 14

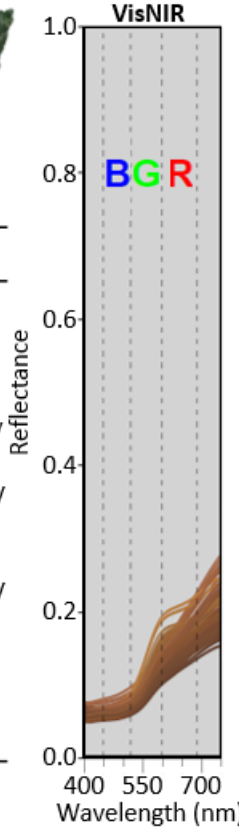


Classification: Red Ferrosol

Parent material: Weathered basalt

Location: Robertson
(34° 36' 40.5" S 150° 32' 12.6" E)

Depth (cm)	Description
0-10	A horizon; 5YR 3/4 (dry), 2.5YR 2.5/3 (moist); loam; strong granular structure; abrupt, even boundary to
10-34	Bt horizon; 5YR 4/6 (dry), 5YR 3/4 (moist); silty clay loam; strong granular structure; clear, wavy boundary to
34-60	Btw1 horizon; 5YR 4/6 (dry), 5YR 3/4 (moist); silty clay loam; strong polyhedral structure; gradual, wavy boundary to
60-83	Btw2 horizon; 5YR 4/6 (dry), 5YR 3/4 (moist); silty clay loam; strong angular blocky and polyhedral structure; gradual, wavy boundary to
83-100+	Btw3 horizon; 7.5YR 4/6 (dry), 5YR 3/4 (moist); silty clay loam; strong angular blocky and polyhedral structure.



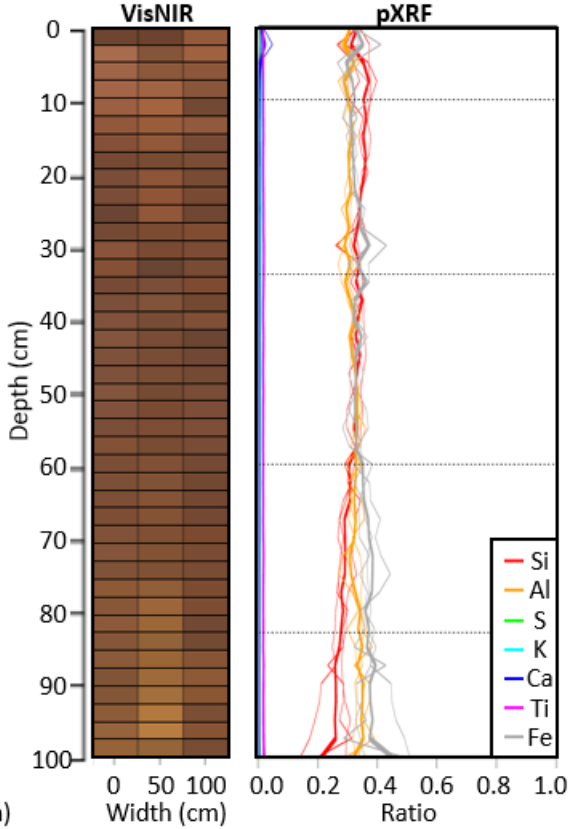
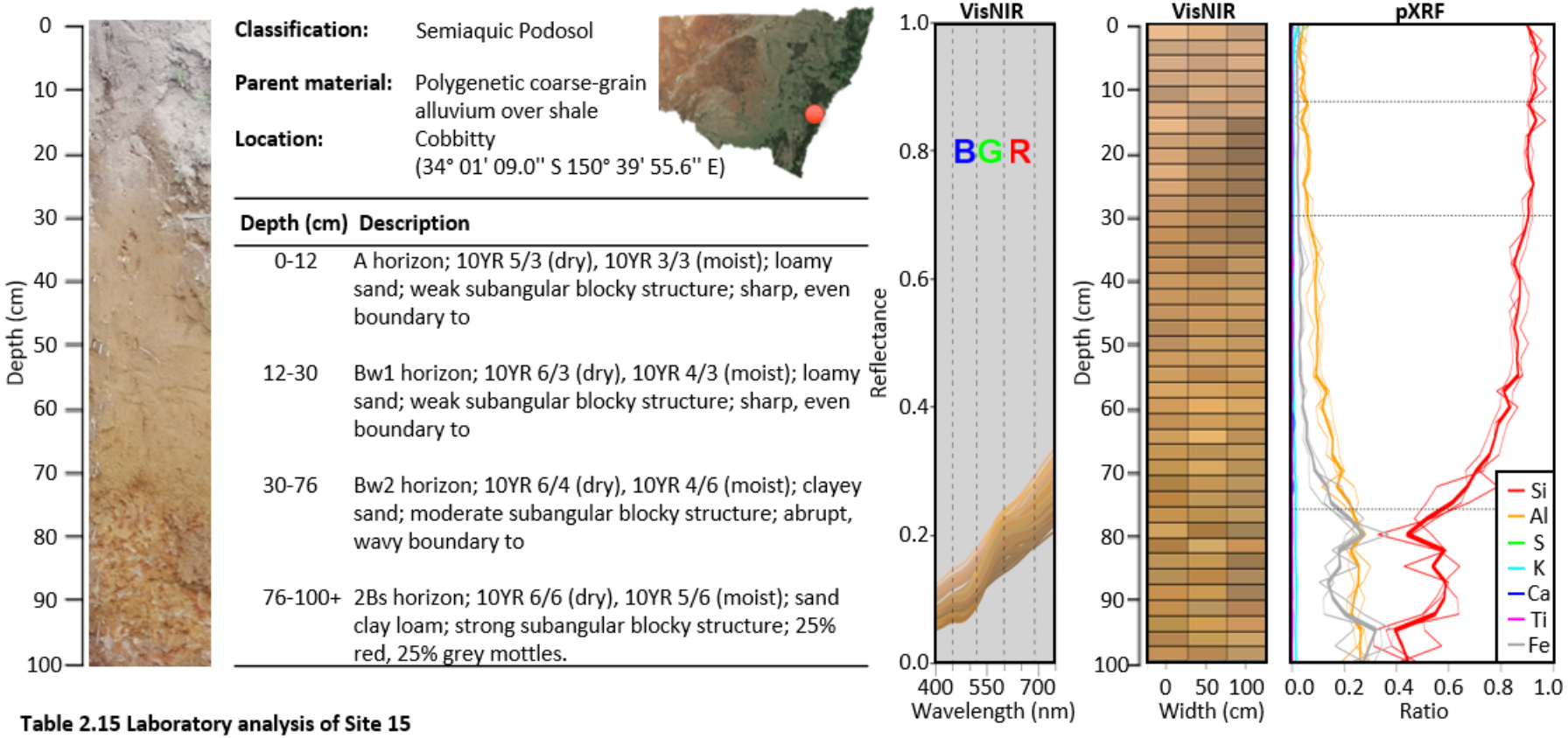


Table 2.14 Laboratory analysis of Site 14

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations (cmol$_{(+)}$ kg^{-1})					CEC (cmol$_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	10	26.2	19.6	54.2	21.9	32.2	5.36	4.51	0.10	6.06	0.544	7.47	-	4.8	1.10	2.0	0.06	1.1	9.1
Bt	10	34	50.0	23.6	26.4	15.5	10.9	5.14	4.17	0.03	3.51	0.205	3.08	-	0.9	0.20	0.4	0.07	3.0	4.5
Btw1	34	60	58.4	21.1	20.6	11.5	9.0	5.16	4.17	0.03	1.99	0.104	1.76	-	0.8	0.09	0.1	0.08	4.0	5.0
Btw2	60	83	53.5	26.2	20.4	15.0	5.4	5.12	4.18	0.03	1.38	0.043	1.01	-	0.9	0.05	0.1	0.10	3.7	4.9
Btw3	83	100	48.1	28.4	23.6	9.1	14.5	4.97	4.08	0.03	0.98	0.005	0.74	-	0.8	0.04	0.3	0.11	5.5	6.7

2.3.2.15 Field observations, sensor readings and laboratory data of Site 15



Classification: Semiaquic Podosol

Parent material: Polygenetic coarse-grain alluvium over shale

Location: Cobbitty (34° 01' 09.0" S 150° 39' 55.6" E)

Depth (cm)	Description
0-12	A horizon; 10YR 5/3 (dry), 10YR 3/3 (moist); loamy sand; weak subangular blocky structure; sharp, even boundary to
12-30	Bw1 horizon; 10YR 6/3 (dry), 10YR 4/3 (moist); loamy sand; weak subangular blocky structure; sharp, even boundary to
30-76	Bw2 horizon; 10YR 6/4 (dry), 10YR 4/6 (moist); clayey sand; moderate subangular blocky structure; abrupt, wavy boundary to
76-100+	2Bs horizon; 10YR 6/6 (dry), 10YR 5/6 (moist); sand clay loam; strong subangular blocky structure; 25% red, 25% grey mottles.

Table 2.15 Laboratory analysis of Site 15

Horizon	Upper depth (cm)	Lower depth (cm)	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	F. sand (g 100 g^{-1})	C. sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	TC (g 100 g^{-1})	TN (g 100 g^{-1})	OC (g 100 g^{-1})	CO_3 eq. (g 100 g^{-1})	Exch. cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})
															Ca	K	Mg	Na	Al	
A	0	12	7.7	3.9	88.4	12.2	76.2	5.26	4.21	0.02	0.89	0.049	0.77	-	0.5	0.13	0.1	0.00	0.5	1.1
Bw1	12	30	9.2	5.6	85.3	14.7	70.6	5.62	4.46	0.02	0.40	0.016	0.46	-	0.6	0.10	0.1	0.00	0.4	1.2
Bw2	30	76	12.6	4.5	82.8	11.2	71.6	5.86	4.66	0.02	0.23	0.007	0.20	-	0.5	0.11	0.2	0.00	0.2	1.0
2Bs	76	100	23.9	10.2	65.9	9.9	55.9	5.35	4.08	0.02	0.11	0.002	0.00	-	0.2	0.08	1.7	0.07	2.5	4.5

2.4 References

Barnes, R. J., Dhanoa, M., Lister, S. (1989). Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra, 43.

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(2), 364-372.

Ben-Dor, E., Heller, D., Chudnovsky, A. (2008). A Novel Method of Classifying Soil Profiles in the Field using Optical Means. Soil Science Society of America Journal 72(4), 1113-1123.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, P., Roger, J-M., McBratney, A.B., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC Trends in Analytical Chemistry 29, 1073-1081.

Bureau of Meteorology (BOM). Australia's official weather forecasts & weather. Available at: http://bom.gov.au [Accessed 8 Oct. 2017].

Dalal, R., Henry, R. (1986). Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. Soil Science Society of America Journal 50(1), 120-123.

Gee, G. W., Bauder, J. W. (1986). "Particle-size analysis," in Methods of Soil Analysis: Part 1. Physical and Mineralogical Methods II, ed A. Klute (Madison: Soil Science Society of America), 383–411.

Islam, K., Singh, B., McBratney, A. (2003). Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Soil Research 41(6), 1101-1114.

Lin, I.K.L. (March 1989). "A concordance correlation coefficient to evaluate reproducibility". Biometrics. International Biometric Society. 45 (1): 255–268.

Minasny, B., McBratney, A.B. (2007) Why calculating RPD is redundant. Available from: https://www.researchgate.net/publication/256464655_Why_calculating_RPD_is_redundant [accessed Oct 07 2017].

Minasny, B., McBratney, A. B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L., Joalland, S. (2011). Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma 167–168(0), 118-124.

Mouazen, A. M., Maleki, M. R., De Baerdemaeker, J., Ramon, H. (2007). On-line measurement of some selected soil properties using a VIS–NIR sensor. Soil and Tillage Research 93(1), 13-27.

Piper, C. S. (1942). Soil and plant analysis. Hans Publishers, Bombay.

Quinlan, J. R. (1992). "Learning with continuous classes", 5th Australian joint conference on artificial intelligence.

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

Rayment, G. E., Lyons, D. J. (2011). Soil chemical methods: Australasia, 3. CSIRO publishing.

Reeuwijk, L. P. (1993). Procedures for soil analysis. International Soil Reference and Information Centre.

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.

Savitzky, A., Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 36(8), 1627–1639.

Sørensen, L., Dalsgaard, S. (2005). Determination of clay and other soil properties by near infrared spectroscopy. Soil Science Society of America Journal 69(1), 159-167.

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, 253-266.

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(1), 29-38.

Williams, P.C. 1987. Variables affecting near-infrared reflectance spectroscopic analysis. In: Near-Infrared Technology in the Agricultural and Food Industries (eds P. Williams & K. Norris), pp. 143–167. American Association of Cereal Chemists Inc., Saint Paul, MN.

3 MAPPING HOMOGENEOUS SPECTRAL RESPONSE ZONES IN A SOIL PROFILE

3.1 Abstract

Homogeneous spectral response zones represent relatively uniform regions of soil which 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 condition, as experienced 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 condition. Spectra underwent routine pre-processing then transformed using either an EPO or DS projection matrix. 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 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 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 (RMSE: EPO = 11.8, DS = 15.4, PP = 38.5) and also in identifying homogeneous spectral response zones that corresponded to field observed horizons.

3.2 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 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-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 36,654 air-dried and ground samples from Australia. Horizon centroids in canonical space 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, varying between 85 and 130 cm depth, at 2 cm increments 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 dimensionality 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 EPO, DS, global moisture modelling and spiking are viable approaches to reduce the negative 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 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 evaluated EPO and DS 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.

3.3 Methods

3.3.1 Site description

The study site was located on Westwood Farm, an experimental property owned by the University of Sydney, 3 km northwest of Cobbitty, NSW, Australia (33°59'44.9"S 150°39'11.9"E) (Fig. 3.1). 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 (Walker, 1960). The mineralogy of the clay fraction of this soil is commonly 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 intensive grazing on naturalised kikuyu (*Pennisetum clandestinum*) and paspalum (*Paspalum dilatatum*) grasses. This site was also used as "Site 3" in future chapters, see sections 2.3.1.2 and 2.3.2.3 for further details.

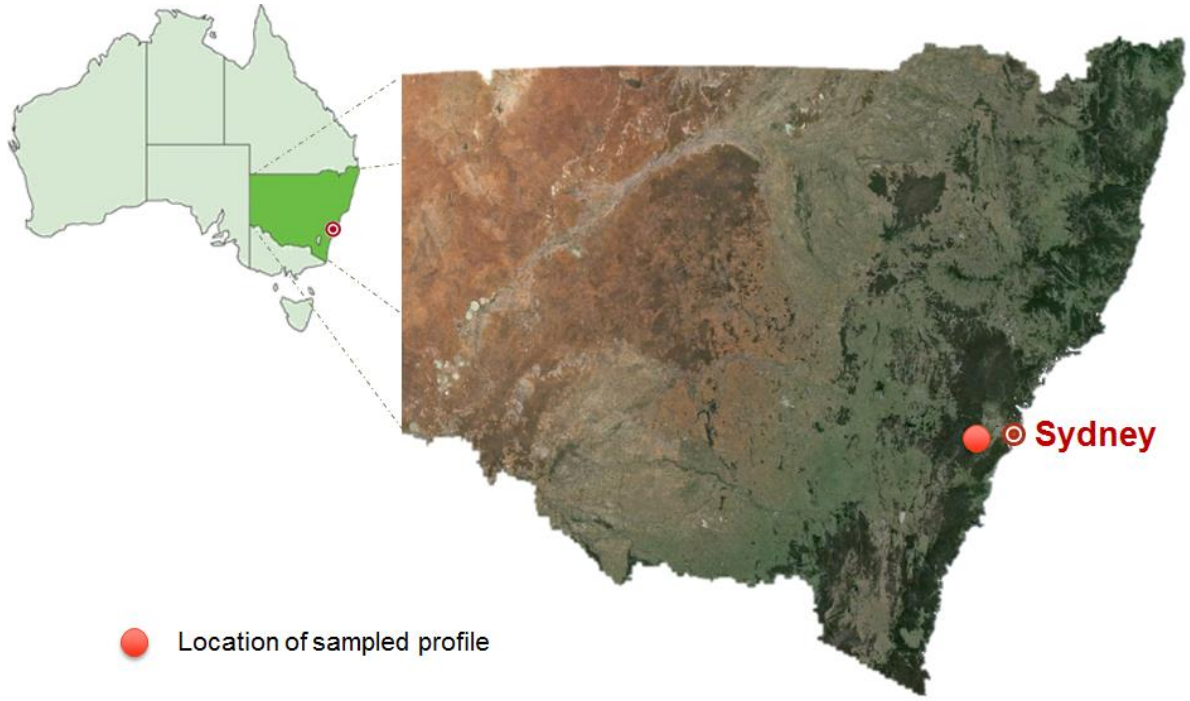


Figure 3.1 Location of the sampled profile in relation to Sydney within the state of New South Wales, Australia.

3.3.2 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 (Isbell, 2016). Notable features of the soil include an abrupt textural contrast from sandy clay loam in the E horizon to medium heavy clay in the Bt1 horizon (Table 3.1). The Bt1 horizon was also found to be strongly acidic, pH (1:5 H_2O) < 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 3.2).

A 1 m × 1 m sampling region was delineated on the pit wall and sheared to a smooth surface (Fig. 3.2). The final shearing was conducted in a horizontal direction, progressing vertically 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.

Table 3.1 Field observations

Horizon	Depth (cm)	Transition	Boundary shape	Munsell (moist)	Texture†	Structure	Mottling and inclusions	Carbonates (1 M HCl)
A	0-16	abrupt	even	10YR 3/4	L	Moderate Granular	Magnetite/Maghemite	negative
E	16-33	clear	wavy	7.5YR 4/4	SCL	Strong subangular blocky	Magnetite/Maghemite	negative
Bt1	33-47	gradual	wavy	7.5YR 4/6	MHC	Strong angular blocky	-	negative
Bt2	47-100+	-	-	10YR 4/6	HC	Strong angular blocky	20% red, 20% grey mottles	negative

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

Table 3.2 Horizon-based bulk soil properties, including a 0-2 cm topsoil sample.

Horizon	Depth (cm)	OC (g 100 g^{-1})	TN (g 100 g^{-1})	Clay (g 100 g^{-1})	Silt (g 100 g^{-1})	Sand (g 100 g^{-1})	pH_w	pH_c	EC (dS m^{-1})	Each. Cations ($cmol_{(+)}$ kg^{-1})					CEC ($cmol_{(+)}$ kg^{-1})	ESP (%)
										Ca	K	Mg	Na	Al		
-	0-2	3.39	0.238	16	17	36	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	0.043	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	0.033	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	0.044	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	0.002	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; TN – total nitrogen; pH_w – pH (1:5 H_2O); pH_c – pH (1:5 $CaCl_2$); EC – electrical conductivity; CEC – cation exchange capacity; ESP – exchangeable sodium percentage.

Figure 3.2 Digital 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.

3.3.3 In situ scanning, sample collection and *ex situ* scanning

Visible near-infrared (VisNIR) readings were taken in 2.5 cm increments to give 41 readings over each 1 m transect. Ten vertical transects were taken at 10 cm lateral spacing, as well as three horizontal transects at 0, 50 and 100 cm depth (Fig. 3.3). Bulk density cores were extracted on a 10 cm grid for further scanning under laboratory conditions (Fig. 3.3). Samples 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 condition moisture. The flat soil surface at the top of bulk density cores was scanned with VisNIR in the laboratory under field-moist condition. The cores were then air-dried at 40°C until constant weight was achieved, and the surface was scanned again to give an air-dry reading. A subset of the sample was ground to pass through a 2 mm sieve for laboratory analysis. The remaining intact soil was heated again at 40°C, reweighed and then

heated to 105°C to obtain the oven-dried weight for calculation of gravimetric moisture content.

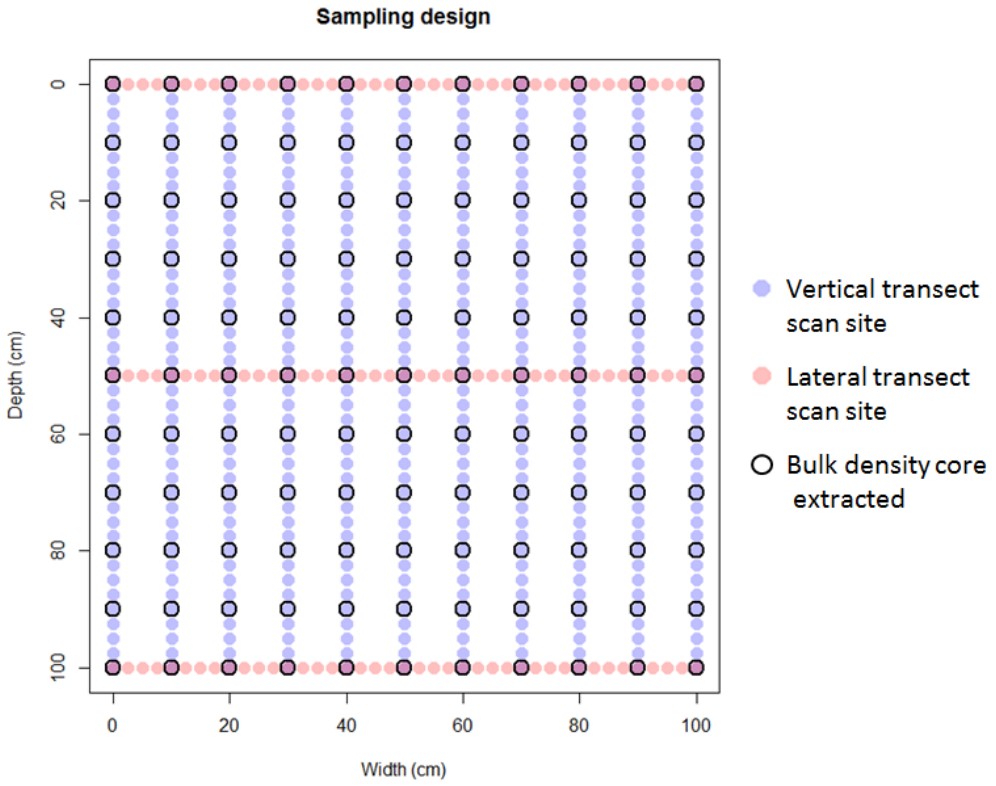


Figure 3.3 Schematic representation of sampling design of the soil profile.

3.3.4 Constructing the projection matrices

A single library was used to construct both EPO and DS transformation matrices, as it has been observed that predictive calibrations are sensitive to the geographical range of the calibration library (Sudduth and Hummel, 1996), and this is more widely applicable to moisture correction factors. The DS transfer matrix was constructed as per Wang *et al.*, (1991) (Fig. 3.4a) and the EPO projection matrix following Minasny *et al.*, (2012) (Fig. 3.4b). A detailed description of EPO is given in section 2.2.5.2.

Structural differences in the EPO and DS matrices are immediately evident. However, features around the 1,400 and 1,800 nm water absorption bands can be identified in both.

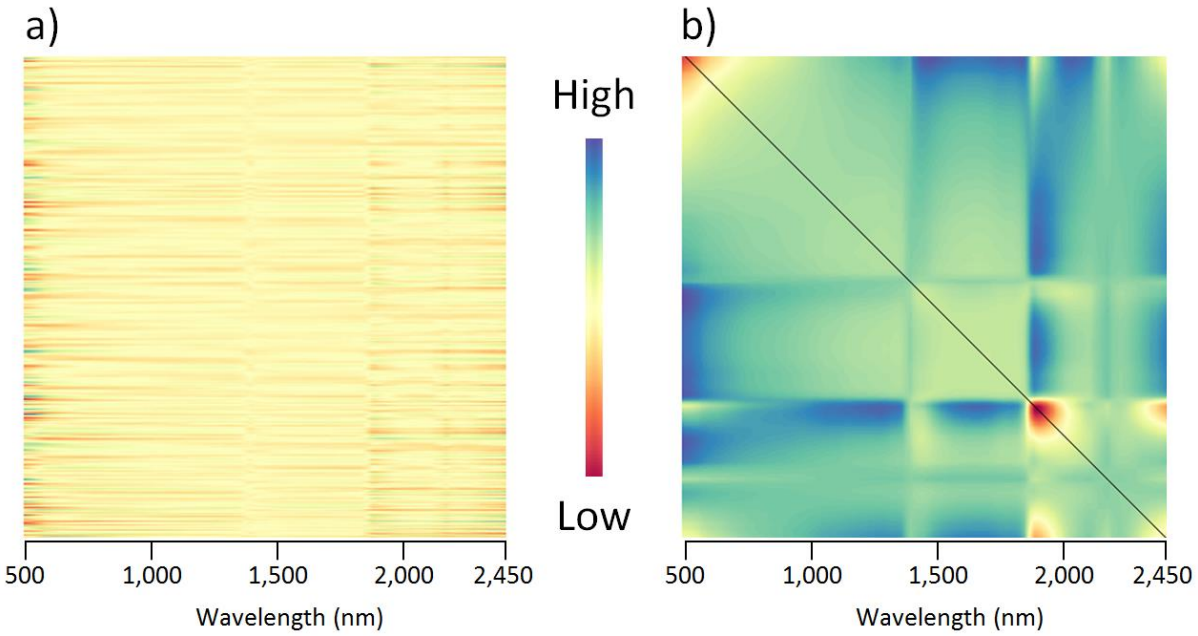


Figure 3.4 Comparison of matrix structure: a) direct standardisation transfer matrix; and b) external parameter orthogonalisation projection matrix.

3.3.5 Data processing

3.3.5.1 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 is driven by a large amount of autocorrelation between wavelengths 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 spectra 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.

3.3.5.2 *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 (Eqn 3.1), so as to minimise the within cluster variance (MacQueen, 1967).

Equation 3.1

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

Where:

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. The number of clusters was set equal to four, i.e. the same number of soil horizons observed, for this investigation to standardise the analysis. 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.

3.3.6 Statistics

Differences between PP, DS and EPO spectra under field moist and air-dry condition were assessed by calculating the RMSE of PC scores projected into the PC space of *in situ* scans. Qualitative assessment of homogeneous spectral response zones was provided by comparison of the distribution of classes on the profile with field observed horizons.

3.4 Results and discussion

3.4.1 Moisture variability of the soil profile

Gravimetric soil moisture content varied widely within the profile (Fig. 3.5). 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.

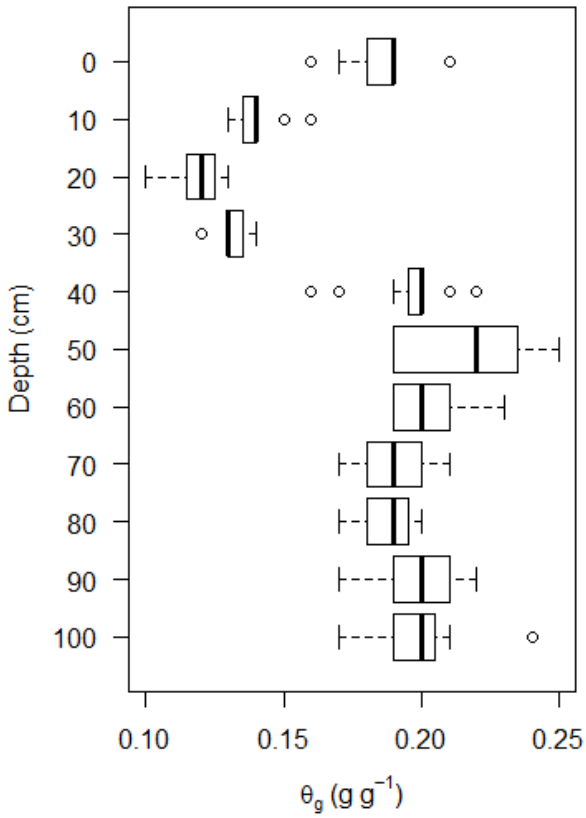


Figure 3.5 Box plots displaying the distribution of gravimetric moisture content by depth.

3.4.2 Moisture and treatment effects on spectra

Spectra taken on field condition cores had reduced reflectance than those taken on air-dry or ground samples, as noted previously (Bowers and Hanks, 1965) (Fig. 3.6a). 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. 3.6b). 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. 3.6c). 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. 3.6d). Differences between the topsoil and subsoil samples also appeared to be accentuated.

3.4.3 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. 3.7).

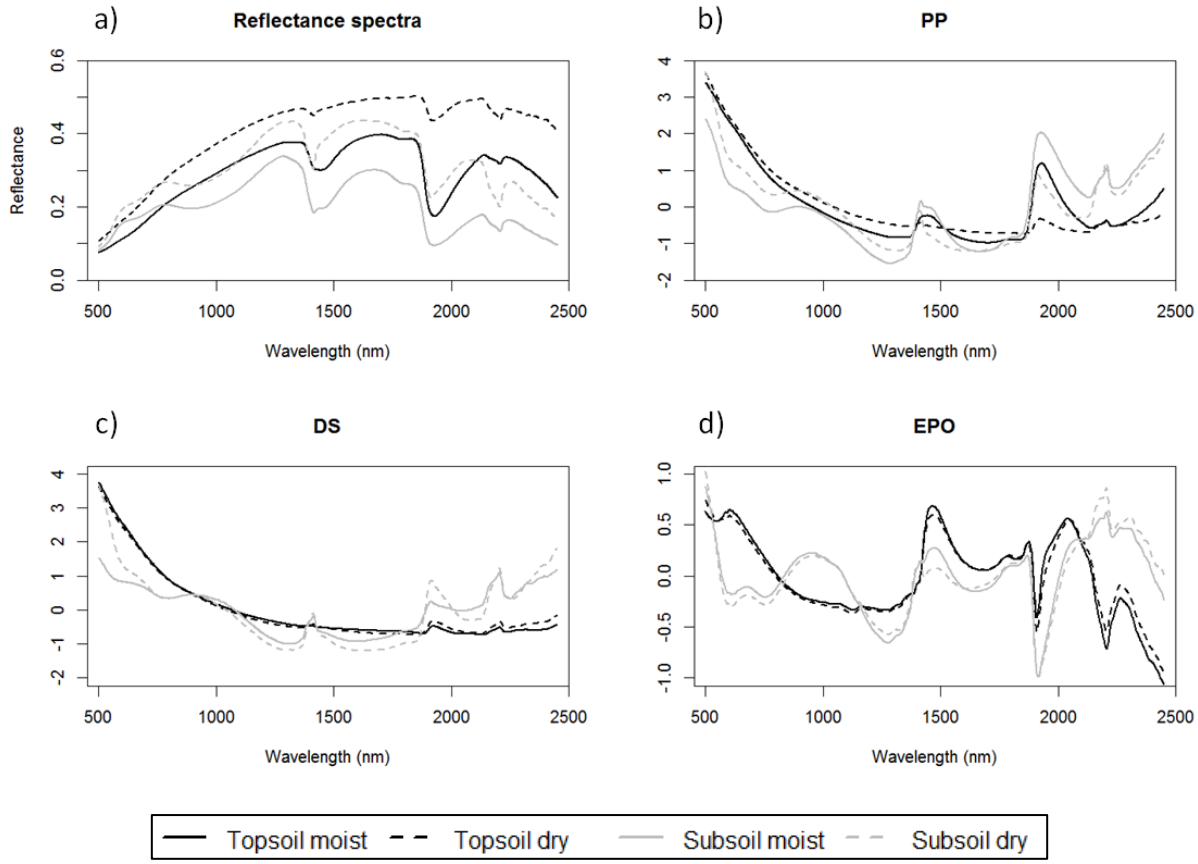


Figure 3.6 Comparison of a representative topsoil and subsoil sample scanned in field moist and air-dried condition: a) trimmed and splice corrected reflectance spectra (500-2450 nm); b) pre-processed (PP) spectra; c) direct standardisation (DS) approach whereby the moist sample is corrected to resemble 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.

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. 3.8).

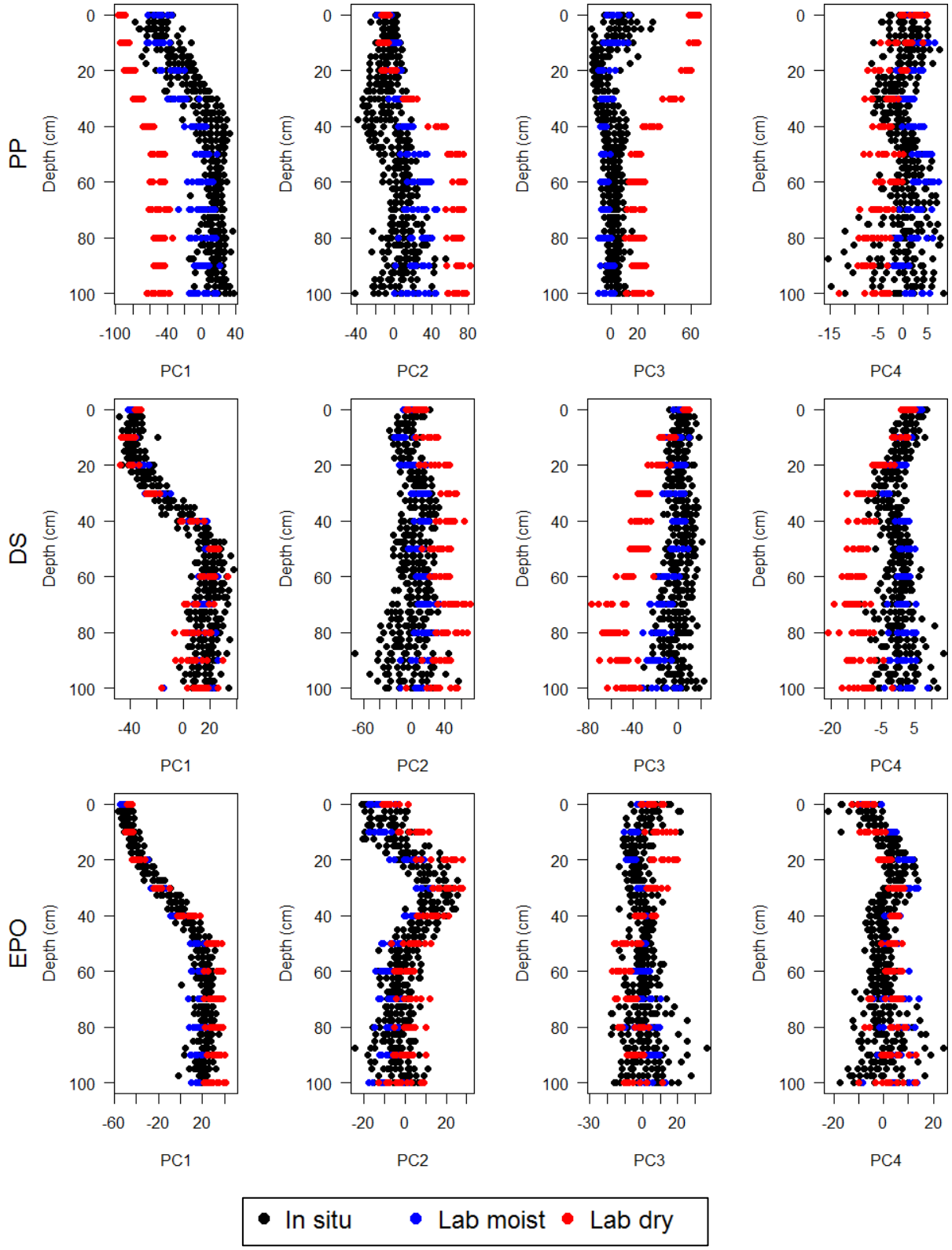


Figure 3.7 Principal component scores for VisNIR spectra obtained *in situ* (black), field moist in the laboratory (blue) and air-dry in the laboratory (red).

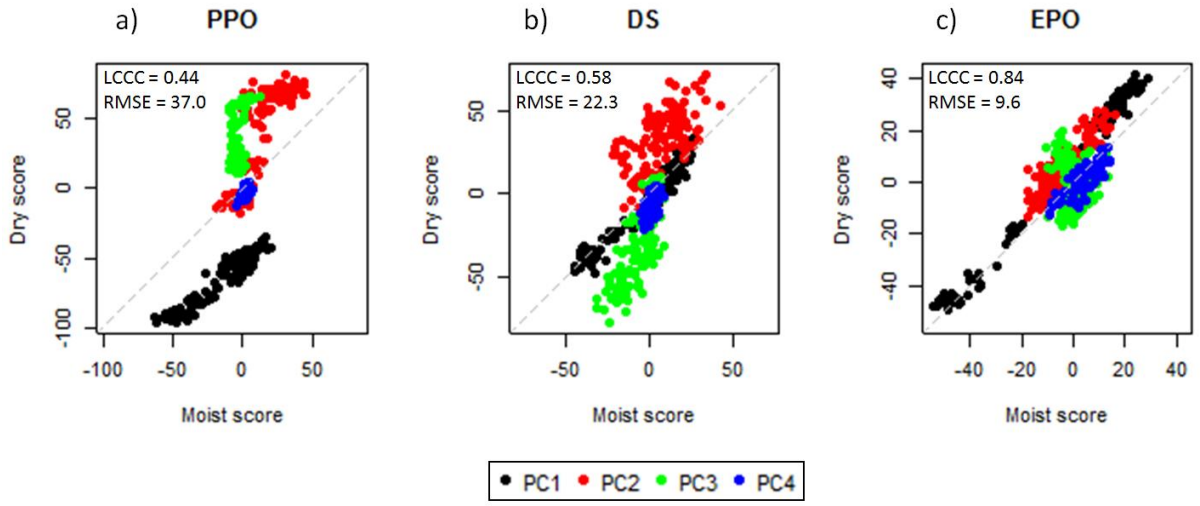


Figure 3.8 The first four PCs of VisNIR spectra under moist and air-dry condition: a) pre-processing only; b) direct standardisation; c) external parameter orthogonalisation.

3.4.4 Comparison of clusters to observed horizons

The four clusters 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. 3.9). 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).

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.

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.

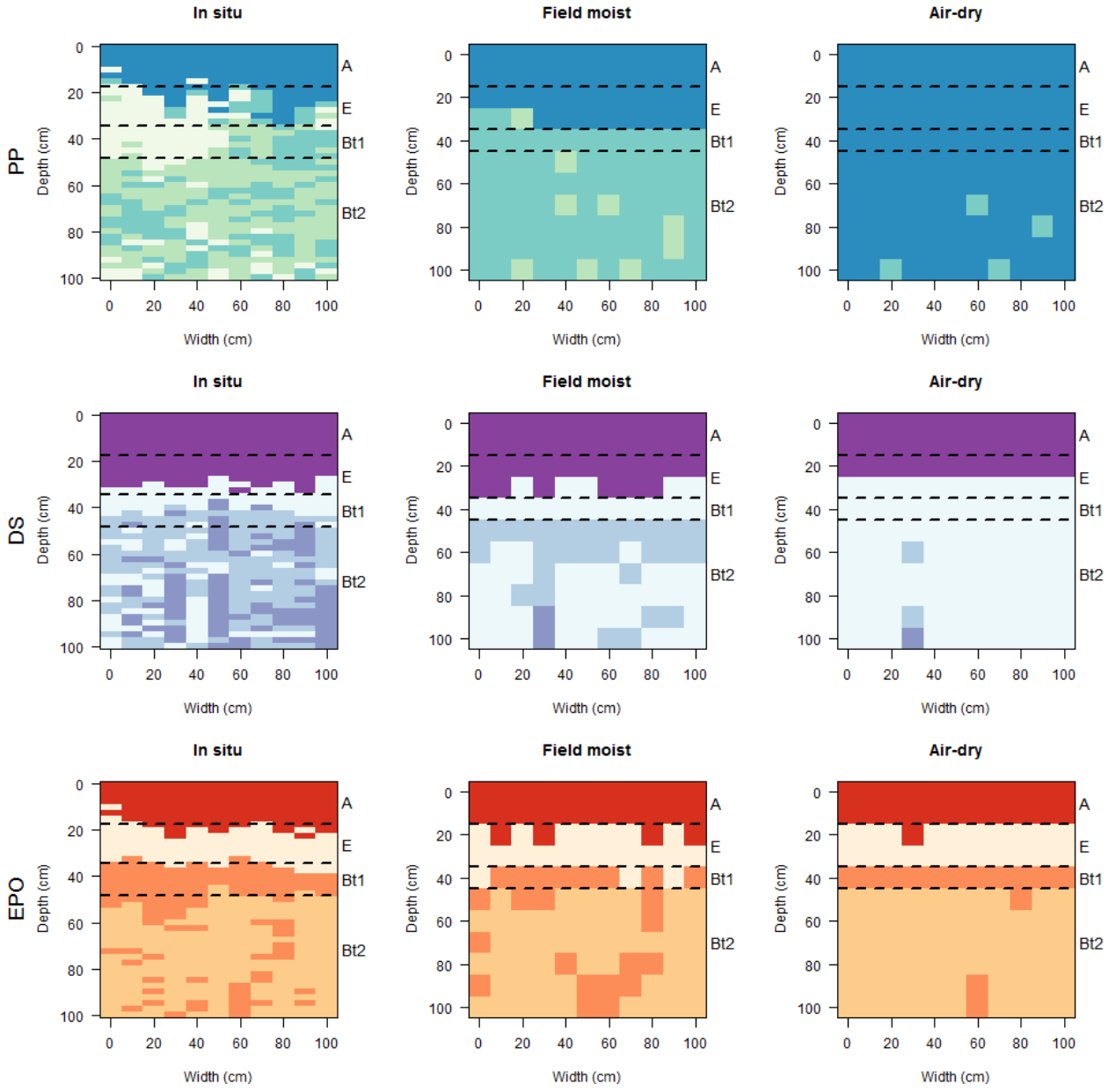


Figure 3.9 The distribution of classes identified by *k*-means clustering of the first four PC scores of spectra following: i) PP – pro-processed only spectra; ii) DS – direct standardisation; and EPO – external parameter orthogonalisation. Field observed horizon boundaries are indicated by dashed horizontal lines. Horizon designations are indicated.

The success of horizon identification by *k*-means clustering of VisNIR spectra is attributable to horizon delineation being 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 100 g^{-1} for between the soil surface and 0.37 g 100 g^{-1} in the Bt2 horizon, while clay ranged from 16 to 67 g 100 g^{-1} (Table 3.2).

Clusters often did not translate to contiguous zones on the soil profile. Associations of clusters 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).

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.

3.4.4.1 Evaluation of DS and EPO

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 samples' moisture content is required. Any method to ascertain soil moisture that requires drying a sample fundamentally 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 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.

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.

3.5 Conclusion

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 clusters consistent with field observed horizons were better expressed under field moist and air-dry condition.

3.6 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 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.

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.

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

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 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.

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.

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.

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 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.

4 DESIGNING SAMPLING TO QUANTIFY SOIL PROFILE VARIABILITY

4.1 Abstract

A range of field deployable soil sensors are available to investigate soil profiles and their variability *in situ*. New sampling methodologies are required to use these sensors efficiently while sufficiently characterising profile variability. To better inform sampling methodologies the vertical and lateral variability of a soil profile was explored. A 1 m x 1 m surface of a soil profile was intensively sampled on 2.5 cm increment transects using visible near-infrared diffuse reflectance spectroscopy (VisNIR). The profile was physically sampled on a 10 cm grid and characterised in terms of organic carbon, total nitrogen, pH, clay content, bulk density, gravimetric soil moisture and electrical conductivity. Visible near-infrared diffuse reflectance spectroscopy and portable X-ray fluorescence spectroscopy (pXRF) scans were taken in field moist, air-dried, and ground (<2 mm) condition. The vertical and lateral variance of soil properties and sensor readings were characterised by constructing directional variograms. The junction of the lateral sill with the vertical modelled variogram occurred below the sampling interval of 10 cm for laboratory and pXRF data. As extrapolating below the minimum sampling interval can be unpredictable, the high resolution VisNIR information was used to investigate below this interval. *In situ* VisNIR spectra were a suitable proxy for

soil properties, as the principal components of the VisNIR spectra were correlated with soil properties. A 4.3 cm increment in the vertical component was found to experience the equivalent variance in soil properties as an entire 1 m lateral cross section. This finding informed the development of a sampling methodology that focussed on intensive vertical sampling, while still capturing the spatial variability of the profile. Following analysis of an additional fourteen profiles, the vertical increment was refined to 4.1 cm, while lateral variability was characterised with a range of 38.1 cm.

4.2 Introduction

Understanding soil spatial variability is paramount to understanding pedogenic processes and for reaching the potential of the soil resource. New devices are available that allow more efficient investigation of soil properties and their variability at the profile level (Hartemink and Minasny, 2014). These devices will change the way soils are described and reported on.

A soil profile is inherently anisotropic (Bathke and Cassel, 1991). Multiple gradients traverse a soil profile, including thermal, redox and moisture potentials, and input of organic substrates (Jenny, 1994). Generally, these gradients apply perpendicular to the soil surface. Differential influences of these gradients form layers in the soil referred to as horizons. Soil horizons are layers of relatively uniform soil properties, whereby variability within a layer is smaller than between layers. Soil variability must be considered when developing a sampling methodology. There are many scales of soil variability. Spatial variability is expressed at the aggregate (Cruvinel *et al*., 1993), field (Cambardella *et al*., 1994), landscape (Cook *et al*., 1996) and global scale (Arrouays *et al*., 2014). The investigation of profile-scale variability, relevant to the use of proximal soil sensors, warrants further investigation.

The support size of proximal sensors provides a minimum sampling interval, below which inbuilt redundancies will prevail. Soils vary more vertically than laterally at the profile-scale. Thus, a single vertical transect through a profile may encompass a significant quantity of the total variance of a profile. However, describing a soil profile involves analysing more than a single vertical transect, as there is also known variation in the lateral component. Initial

studies analysing soil profiles with proximal sensors have sampled with equal vertical and lateral intensity (e.g. Adhikari *et al*., 2016; Grauer-Gray and Hartemink, 2016). However, this may be an inefficient method.

Another consideration is the lateral variability of soils as a function of depth. If more variability is expressed at the soil surface, or at depth, then sampling intensity can be scaled accordingly to efficiently capture the variability. There are multiple factors affecting variability with depth. Topsoils may be more homogeneous due to bioturbation and/or cultivation, meanwhile, subsoils are more directly influenced by parent material, preferential flow paths and localised organic input from root exudates.

Uncertainties prevail in terms of an efficient vertical and lateral sampling intensity. To guide effective sampling methodologies using proximal soil sensors, better understanding of soil spatial variability at a relevant scale is required. This study investigated the development of an improved sampling methodology by exploring profile-scale variability in terms of soil properties and proximal soil sensor readings. Vertical and lateral variability were characterised independently, as well as topsoil and subsoil variability.

4.3 Methods

4.3.1 Proximal sensor scanning, sampling and laboratory analysis

This research was conducted in two parts. An initial investigation was conducted by intensively sampling a single soil profile. Findings from this initial investigation were used to formulate a sampling methodology for an additional fourteen soil profiles.

4.3.1.1 Preliminary investigation site, intensively sampled profile

The initial sampling was conducted on the same soil profile and sampling methodology as used in Chapter 3. Briefly, a 1 m x 1 m surface of a soil pit wall was sheared to a smooth surface and a 10 cm sampling grid was outlined with galvanised nails. The profile was

sampled *in situ* using VisNIR at 2.5 cm increments along 1 m transects. Vertical transects was space 10 cm apart, to give eleven separate transects in the profile. Lateral transects were scanned at 0, 50 and 100 cm depth respectively. Bulk density cores were taken on the 10 cm grid. The bulk density cores were transported to the laboratory and rescanned using VisNIR and also pXRF in field moist, air-dry, and air-dry and ground (<2 mm) condition. The sampling grid was characterised in terms of TC, TN, pH (1:5 H_2O), pH (1:5 $CaCl_2$), and EC, as outlined in section 2.2.4. However, clay content was calculated using the pipette method (Gee and Bauder, 1986) for this investigation. Measurement of bulk density and gravimetric moisture content were also obtained from the bulk density cores.

4.3.1.2 Additional fourteen sites

Results from the preliminary investigation were used to construct a revised sampling methodology that was used at an additional fourteen profiles. The fourteen profiles represented a diverse selection of soils from across the state of New South Wales, Australia. For a full description of the profiles see section 2.3. For these additional sites, VisNIR spectra were recorded on three vertical transects at 0, 50, 100 cm lateral spacing, and lateral transects were taken at 0, 50 and 100 cm depth, as described in section 2.2.2. Horizon-based sampling was conducted, and samples were characterised for TC, TN, pH (1:5 H_2O), pH (1:5 $CaCl_2$), EC, texture, OC, CO_3 equivalent, CEC and exchangeable cations, as outlined in section 2.2.4.

4.3.1.3 Spectral pre-processing and EPO transformation

To reduce the negative effects of variable soil moisture and facilitate comparison of field condition and air-dry VisNIR spectra, all spectra were subject to routine pre-processing and EPO transformation, as outlined in section 2.2.5.1 and section 2.2.5.2. Dimensionality reduction was achieved through PCA, as described in section 3.3.5.1. Laboratory scans in field condition, air-dry, and air-dry and ground (<2 mm), where all projected into the PC constructed utilising their corresponding *in situ* scans for direct comparison.

4.3.2 Variogram construction and evaluation

Variograms were used to characterise profile-scale variability. Variograms describe the spatial dependence of a variable, i.e. the expected magnitude of variance between two observations as a function of the distance between them. Empirical variograms are constructed for observations $z_i, i = 1,2, \dots, k$ by summation of the variance between pairs of observations, $N(h)$, separated by a given lag distance h (± lag tolerance) and dividing by twice the number of pairs of observations (Eqn 4.1).

Equation 4.1

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} \left| z_i - z_j \right|^2$$

Empirical variograms are used as estimators of an underlying theoretical variogram which must be calculated for geostatistical applications such as kriging and stochastic simulations (Oliver and Webster, 1990). Model functions used to fit empirical variograms must be conditional negative semi-definite (McBratney and Webster, 1986). Common examples are circular, spherical, exponential, Gaussian- and Matérn-based models. These models are often used to described the spatial dependence of the observations in terms of a sill, range and nugget. Where:

- The sill is the limit of the semivariance. Representing the maximum semivariance over the study area.
- The range is the lag distance at which the sill value is first acheived. Autocorrelation is essentially zero beyond the range.
- The nugget describes significant semivariance at zero lag distance. Theoretically the variogram should pass through the origin, as at a lag distance of zero no semivariance is expected. Nuggets exist due to semivariance at lag distances smaller than the minimum sampling spacing, and measurement error.

To characterise vertical and lateral variation independently, the two components must first be separated. This can be achieved through the use of directional variograms (St-Onge and

Cavayas, 1995). Directional variograms are commonly used when describing relationships that are direction dependent, such as the dispersion of point source pollution by prevailing winds (e.g. Potoglou and Kanaroglou, 2005). Directional variograms restrict the search criterion for valid lag distance pairs to a directional band specified by an azimuth direction, angular tolerance and bandwidth (Fig. 4.1).

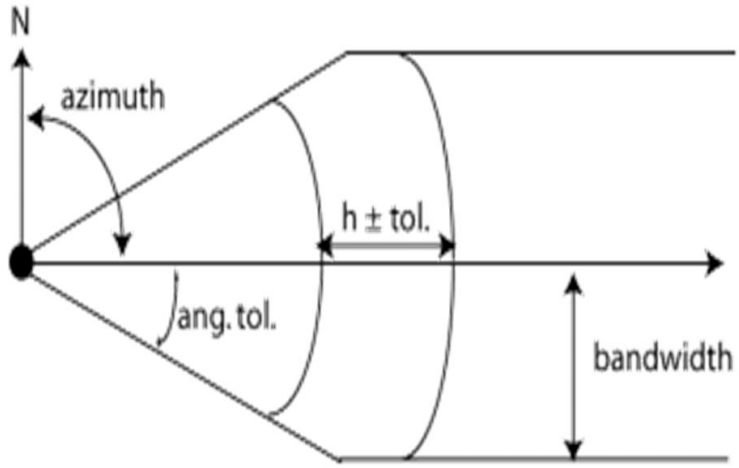


Figure 4.1 Schematic representation of the directional band used to isolate pairs when constructing an empirical directional variogram.

To isolate vertical and lateral transects the directional angle, a proxy azimuth from a vertical reference plane towards the soil surface, is set to 0° for vertical components or 90° for lateral components. Angular tolerance and bandwidth are both set equal to 0. This isolates each transect for variance calculations which are then combined for construction of the empirical variogram.

Empirical directional variograms were constructed for principal component scores using the *variog()* function from the "geoR" package (Ribeiro and Diggle, 2016). Model functions were constructed from the empirical variograms using the *variofit()* function from the same package.

4.3.2.1 Comparing vertical and lateral variability

To characterise the relationship between vertical and lateral variance, the point of intersection of the lateral sill with the modelled vertical variogram was identified for each principal component score. The points of intersection were then weighted by the amount of variation explained by each principal component, to give an estimation of the deviation in the vertical component required to experience the same amount of variation as that experienced across an entire 1 m lateral cross-section of the profile.

4.3.3 Statistics

4.3.3.1 Coefficient of variation

To explore soil variability with depth the coefficient of variation (CV) was calculated for laboratory derived properties and pXRF reported elemental composition on individual lateral transects.

The coefficient of variation is a unitless measure developed to characterise the dispersion of a dataset. It is a useful technique to standardise and compare variability measures where mean values vary significantly, such as mean values of organic carbon in topsoil versus subsoil samples. The CV of a sample is calculated as the ratio of the standard deviation, s, to the mean, $\bar{x}$, of a dataset (Eqn 4.2).

Equation 4.2

$$CV = \frac{s}{\bar{x}}$$

4.3.3.2 Pearson's correlation coefficient

Pearson's correlation coefficient (r) was used to describe the relationship between principal component scores of VisNIR spectra and laboratory derived data to determine if they are a suitable proxy for describing soil variability.

Pearson's correlation coefficient is a widely used measure to describe the linear correlation between two variables. Pearson's correlation coefficient is calculated as the covariance of two variables divided by the product of their standard deviations (Eqn 4.3). Values range between +1 for perfect positive correlation to -1 for a perfect negative correlation.

Equation 4.3

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where:

x_i, y_i are the i^{th} observed values for x and y respectively

$\bar{x}$, $\bar{y}$ are the mean values of variable x and y respectively

4.4 Results and discussion

4.4.1 Laboratory and pXRF data

Soil properties were found to vary vertically and laterally. No property exhibited vertical stationarity, and only pH (1:5 $CaCl_2$) exhibited a monotonic change with depth (Fig. 4.2). Summary statistics of the laboratory derived and pXRF data may be found in Table 4.1.

Total carbon and TN were highly correlated. Their largest values were observed at the soil surface before decreasing rapidly to 30 cm. Values then increased slightly to 40 cm before decreasing again. This increase is likely driven by the corresponding increase in clay at this point, as clay particles offer a number of mechanisms to stabilise organic matter particles and protect them from decomposition (Six, *et al*., 2002). All horizons tested negative for the presence of carbonates, therefore total carbon is equivalent to organic carbon.

pH (1:5 $CaCl_2$) displayed high variability in the topsoil, with a substantial decrease in variability with depth. pH (1:5 H_2O) displayed high variability throughout the profile and

increased over the 70 to 100 cm depth range. This increase in pH (1:5 H_2O) is thought to be due to decreasing EC values found over this depth range. Smaller levels of EC reduce the exchangeable acidity that is displaced into the solution and accentuate differences between pH measured in deionised water and $CaCl_2$ solution (Minansy *et al*., 2011).

Clay content was small in the top 10 cm, with an average of 11.8 g 100 g^{-1}. Clay content then rose sharply in the argillic horizon, reaching a mean value of 68.1 g 100 g^{-1} at 60 cm, before decreasing gradually with depth.

Bulk density showed an inverse relationship with clay, except for the soil surface, where high OC contributed to the smallest observed bulk density of 1.32 g cm^{-3}. Values increased rapidly to a maximum of 1.62 g cm^{-3} at 20 cm. Bulk density then decreased to a local minimum between 50 – 60 cm before increasing slightly. Gravimetric soil moisture and EC both showed inverse relationships to bulk density, although soil moisture remained relatively stable below 60 cm.

Al, Si, K and Fe accounted for 99.8% of the mass of all observable elements in the profile. The parent material of the site is Ashfield Shale, a Triassic sedimentary rock comprising black mudstones and grey shales (Howard,1969). The pXRF reported elements showed high levels of Fe levels and negligible Ca, reflecting the known geochemistry of this shale parent material. Conserved relationships, observed using pXRF, between parent materials and soil geochemistry have previously been demonstrated by Stockmann *et al*. (2016)

Al, K and Fe were all correlated with clay to varying degrees. Al values had the strongest correlation with clay. Average Fe values increased with depth, although variation was much greater in both the topsoil and deep subsoil. This was attributed to the presence of magnetite/maghemite inclusions in the topsoil and heavy mottling in the subsoil, which is attributed to dissolution and precipitation of Fe resulting in spatial disaggregation (Rabenhorst and Parikh, 2000). Conversely, Si was inversely correlated with clay content.

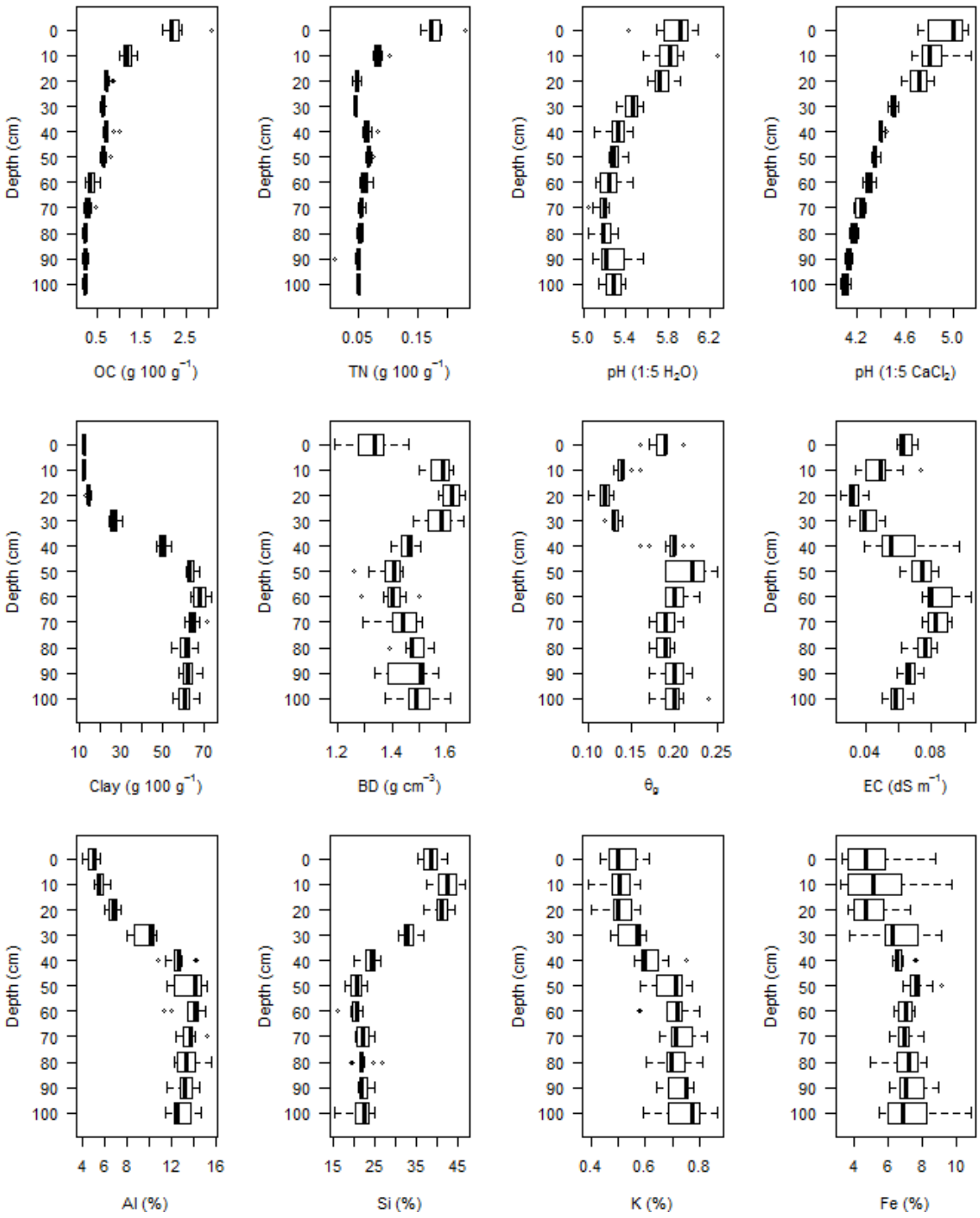


Figure 4.2 Box plots displaying the vertical distribution of soil organic carbon, total nitrogen, pH (1:5 H_2O), pH (1:5 $CaCl_2$), clay, bulk density, gravimetric soil moisture, electrical conductivity, pXRF observed Al, Si, K and Fe.

Table 4.1 Summary statistics of laboratory and pXRF data of air-dry samples. IQR = interquartile range.

	Laboratory data								pXRF (%)			
	OC	TN	pH_w	pH_c	Clay	BD	θ_g	EC	Al	Si	K	Fe
Minimum	0.17	0.01	5.0	4.1	11.0	1.19	0.10	25.4	4.0	15.5	0.39	3.2
Maximum	3.06	0.23	6.3	5.2	73.7	1.67	0.25	103.5	15.6	46.7	0.87	10.8
Median	0.62	0.05	5.3	4.4	57.4	1.48	0.19	64.4	12.4	24.1	0.64	6.7
IQR	0.47	0.02	0.4	0.5	48.5	0.15	0.06	25.8	6.4	15.5	0.17	1.7
Mean	0.69	0.07	5.4	4.4	44.9	1.48	0.18	62.9	10.9	28.1	0.64	6.5

4.4.1.1 Coefficient of variation with respect to depth

Individual properties exhibited variation of CV with depth. In the top 50 cm of the profile Fe and EC had the largest CV. Below 50 cm, Fe and OC had the largest CV. pH (1:5 $CaCl_2$), pH (1:5 H_2O), clay and BD all had a low CV value which were relatively stable with depth. OC, TN, gravimetric soil moisture, EC and all pXRF elements had relatively high CV values. The pXRF elements had relatively high CV values when scanning intact samples, as the spatial heterogeneity of the sample is reserved. Rescanning of the samples following grinding and passing through a 2 mm sieve greatly reduced the CV, except for Fe the topsoil. A global median lateral CV value of 7.2% for was identified for laboratory derived data and pXRF of air-dry samples. If samples were ground and sieved prior to pXRF scanning this value was reduced to 5.6%. The median CV value is 5.2% when considering laboratory data alone.

No trend with depth was observed with the CV values of lateral transects when considering all properties (Fig. 2.3). This demonstrates that lateral variance was consistent in both the

topsoil and subsoil. Therefore, sampling intensity in the topsoil and subsoil must be equivalent to capture the full spatial variability of a profile.

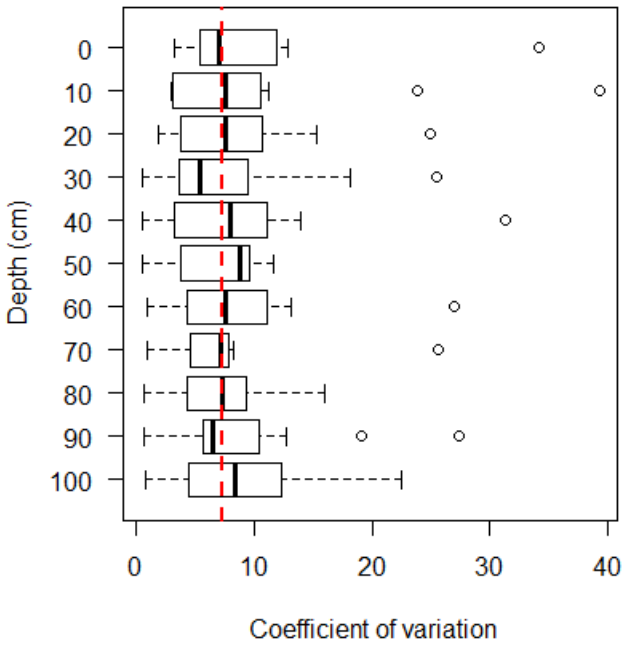


Figure 4.3 Box plots displaying coefficient of variation values for laboratory and $pXRF_{air\text{-}dry}$ data with depth. The dotted red line indicates the global median lateral coefficient of variation value of 7.2% for comparison.

4.4.1.2 Variogram construction and evaluation

The variance of the laboratory data had not stabilised to a sill over the depth of the profile for vertical transects. This is not surprising, given the large variation of soil properties with depth. The junction of the sill of lateral variograms with the modelled vertical theoretical variogram occurred below 10 cm in all laboratory data and pXRF elements, except pH (1:5 H_2O) and Fe where it occurred at ~12 cm (data not presented). Extrapolating variogram models below this minimum distance is likely to introduce errors. Therefore, to characterise variance below the 10 cm interval, the VisNIR spectra was explored.

4.4.2 Intensively sampled profile

4.4.2.1 Correlation with laboratory data

The first four principal components of the *in situ* VisNIR spectra explain 97.3% of the total variance of the spectra (Fig. 4.4). The use of EPO reduced the deleterious effects of variable moisture such that projected VisNIR scans under moist, dry and ground condition had very similar correlations with laboratory data (Table 4.2). Although PC1 is correlated with gravimetric soil moisture, this is most likely due to the relationship between clay and soil moisture. The PCs of the intensively sampled VisNIR spectra were therefore used as proxies for soil properties; given that they were correlated with soil properties and that the PC scores were stable under variable moisture and surface conditions.

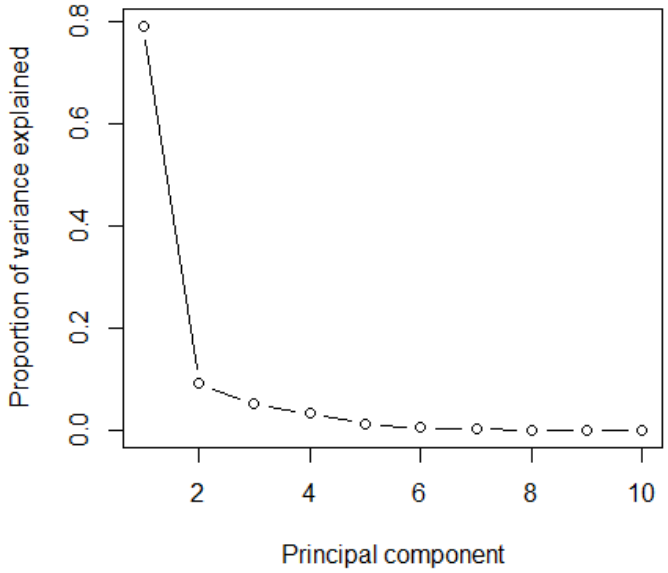


Figure 4.4 Proportion of variance explained by the first ten principal components of the intensively sampled profile.

Table 4.2 Pearson correlation coefficient (r) of soil properties and PC1 under variable scanning conditions.

	Property							
Condition	OC	TN	pH_w	pH_c	Clay	BD	θ_g	EC
Field moist	-0.80	-0.57	-0.86	-0.92	0.98	-0.25	0.66	0.61
Air-dry	-0.77	-0.52	-0.86	-0.91	0.99	-0.30	0.71	0.64
Ground	-0.75	-0.53	-0.86	-0.87	0.98	-0.29	0.68	0.63

4.4.2.2 Variogram construction and evaluation

Similarly to the laboratory data, the variance of PC1 and PC2 also did not produced a sill over the 1 m vertical transects. However, valid variograms could still be fit to the data. The point of interection of the laterial sill with the vertical variogram model was found to occur at 3.4, 7.7, 10.7 and 7.4 cm for the first four PCs respectively (Fig. 4.5).

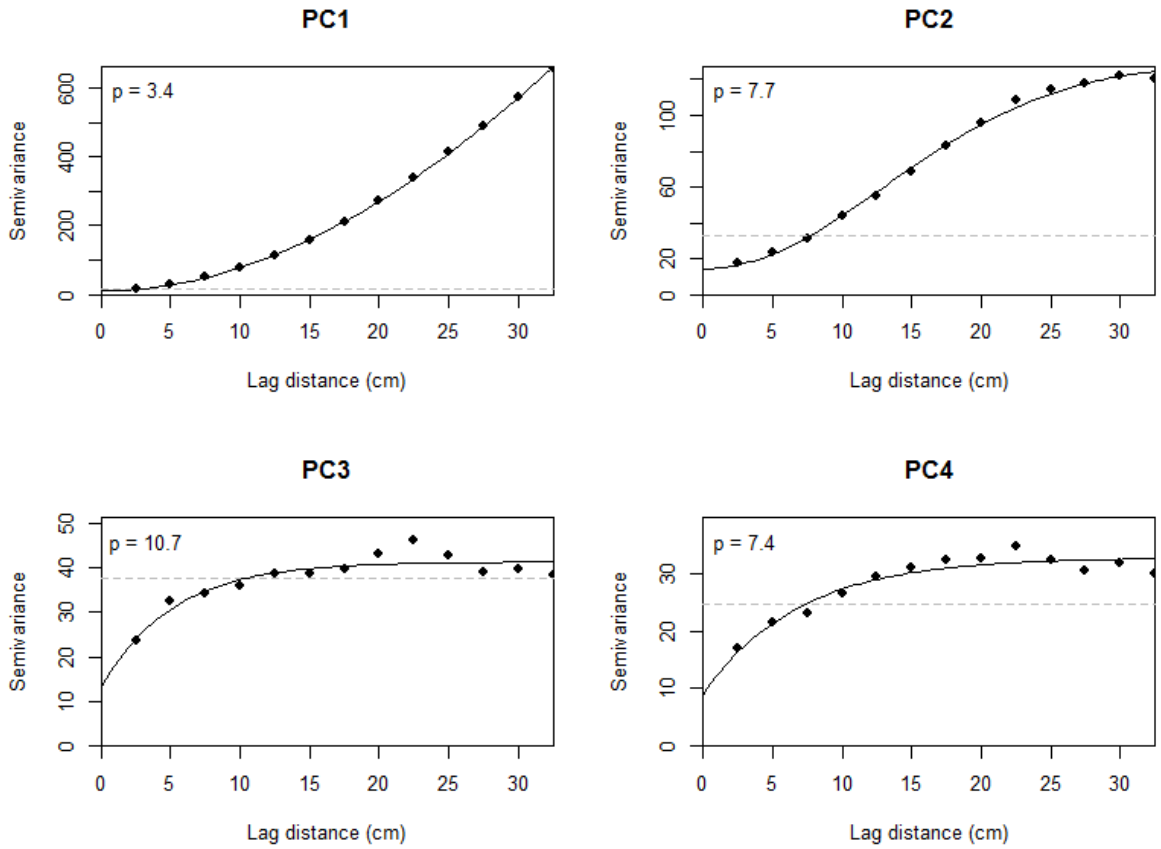


Figure 4.5 Empirical variograms and fitted models for vertical directional variograms. The sill of the equivalent lateral variogram is indicated by the dotted grey line with the point of intersection, p, inscribed in the upper left corner. A Gaussian function was used to fit the model for PCs 1 and 2, a Matérn function was used for PCs 3 and 4.

Weighting results by the amount of variance explained by each PC gives 4.3 cm as the vertical increment required to experience an equivalent amount of variation in soil properties as you would experience across an entire 1 m lateral cross-section. This suggests an optimal sampling grid would have vertical transects space 1 m apart with observations 4.3 cm apart

on each transect, or some multiple thereof. This information was combined with the observation that soil properties have similar variance in both the topsoil and subsoil, to construct the revised sampling methodology that was applied to an additional fourteen soil profiles. Corresponding scans from the intensively sampled profile were also utilised, to give observations at fifteen soil profiles in total.

4.4.3 Fifteen soil profiles

4.4.3.1 Correlation with laboratory data

The first four principal components accounted for 95.0% of all variation in the spectra from the fifteen profiles (Fig. 4.6).

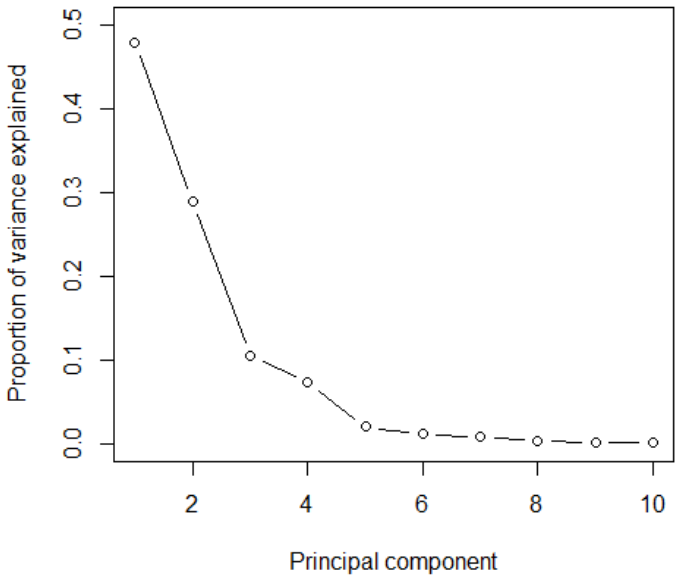


Figure 4.6 Proportion of variance explained by the first ten principal components for the fifteen soil profiles.

As with the intensively sampled profile, the PCs of the VisNIR spectra from all fifteen profiles were correlated with soil properties, reaffirming their use as a proxy for soil properties (Fig. 4.7). Horizon aggregated PC scores were used for the correlation calculation in this instance.

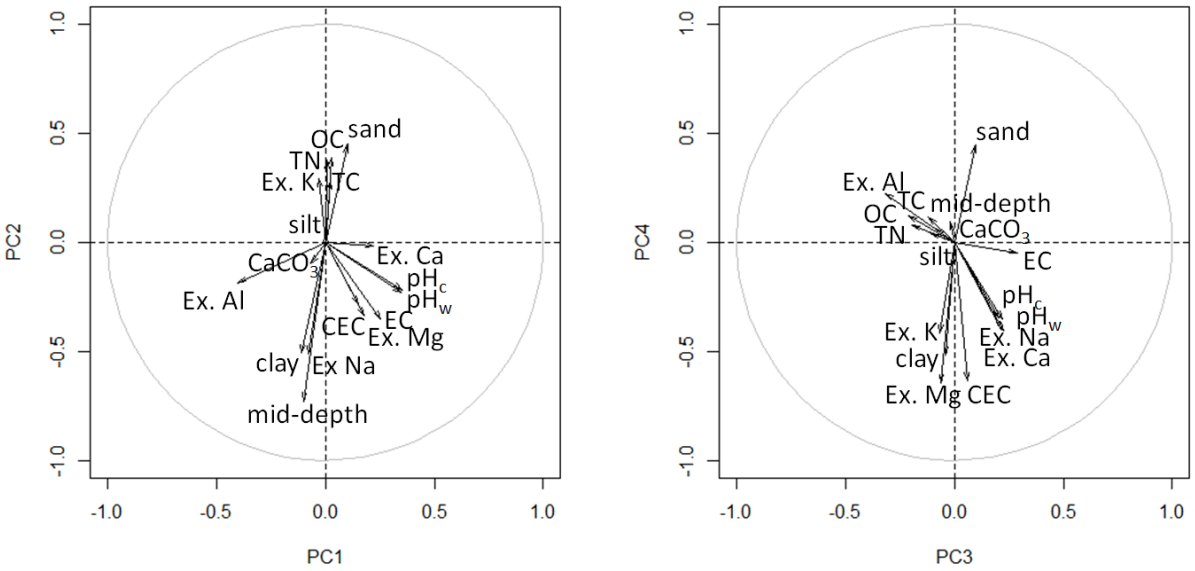


Figure 4.7 Biplot of the first four principal components from *in situ* scans of all fifteen soil profiles.

PC1 was negatively correlated with exchangeable Al and positively correlated with pH. PC2 describes a depth component. It is positvely correlated with sand, OC and TN all of which are more strongly expressed in the topsoil. Conversely negative correlations are seen with mid-depth and clay. PC3 is postively correlated with pH and some base cations and negatively correlated with exhangeable Al, OC and TN. PC4 contains some information related to soil fertility, as it is positively correlated with sand and exchangeable Al and negatively correlated with clay, CEC and exchangeable base cations.

4.4.3.2 Variogram construction and evaluation

The range of lateral variograms was found to be equivalent to 24.7, 33.6, 83.9 and 77.3 cm for the first four PCs respectively (Fig. 4.8). The weighted mean of the ranges was 38.1 cm. This represents the distance at which samples are no longer correlated, at the profile-scale. As such, sampling at a lateral increment much smaller than this is redundant, as sample sites will likely be correlated. Further, sampling at a lateral distance more than double this distance, will likely miss variability.

It should be noted that this analysis considers the entire 1 m lateral cross-section when calculating lateral sills and ranges. Normally when constructing empirical variograms, the maximum distance will be limited to one-third of the maximum distance between samples. This is to limit the effect of a small number of observations at large distances on the overall variogram form. However, in this analysis, with the use of directional variograms, there are a total of 45 observations at the maximum distance, and observation of the empirical variograms indicate that inclusion of the entire 1 m transect did not hindered the construction of empirical variograms.

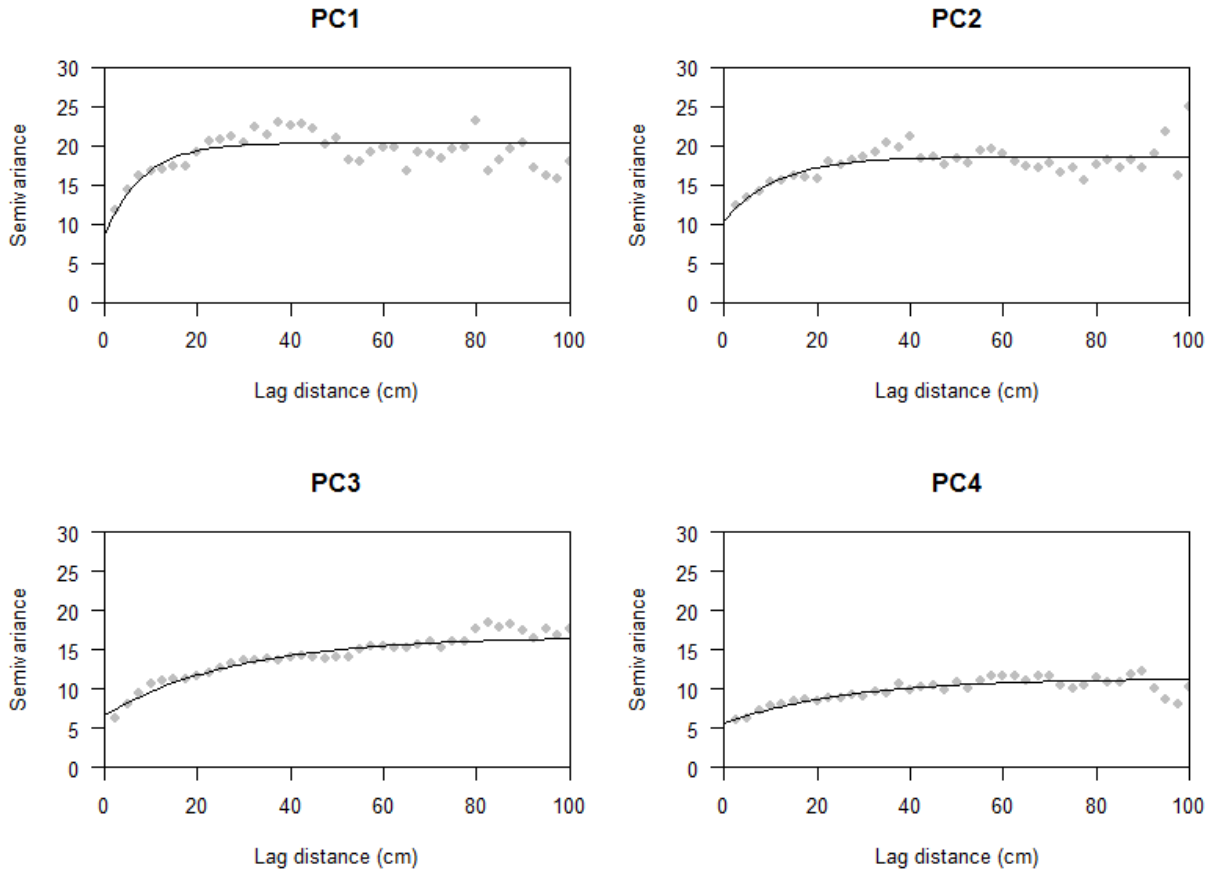


Figure 4.8 Empirical variograms and fitted models for lateral directional variograms. A Matérn function was used to fit the variogram models.

The point of intersection of the lateral sill with the vertical variogram model was found to occur at 2.8, 3.7, 8.4 and 7.8 cm for the first four PCs respectively (Fig. 4.9). The weighted mean of the vertical increment was 4.1 cm.

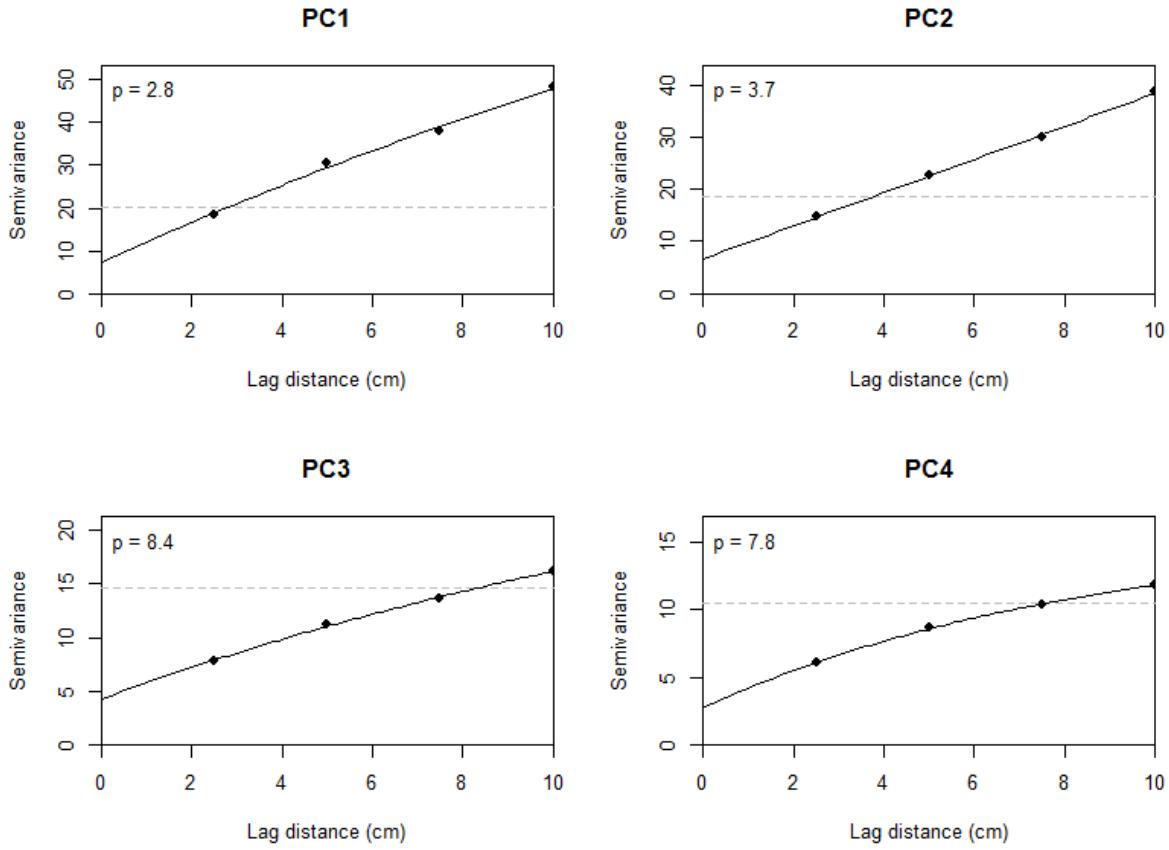


Figure 4.9 Empirical variograms and fitted models for vertical directional variograms. The sill of the equivalent lateral variogram is indicated by the dotted grey line with the point of intersection, p, inscribed in the upper left corner. A Matérn function was used to fit the vertical and lateral variograms.

Given the weighted range of the lateral transects of 38.1 cm, and the vertical increment of 4.1 cm. The sampling design of three vertical transects, at 0, 50 and 100 cm lateral spacing, with observations taken at 2.5 cm increments, would effectively and efficiently characterise profile variability.

4.5 Conclusion

Variation in soil property variance was stable with depth, indicating that a soil profile should be sampled with equal density in the topsoil and subsoil. The variance in soil properties experienced from a 4.1 cm vertical increment was found to be equivalent to the variance experienced over an entire 1 m lateral cross-section for fifteen profiles. This indicates that to fully capture profile variability it must be sampled with a greater density in the vertical dimension, and sampling intensity should be constant in topsoil and subsoil locations. The weighted range of lateral sills was found to be 38.1 cm. In general, sampling much below these increments for the purpose of routine soil description is redundant.

4.6 References

Adhikari, K., Hartemink, A. E., Minasny, B., 2016. Mapping a Profile Wall of a Typic Udipsamments from the Central Sands in Wisconsin, USA, Digital Soil Morphometrics. Springer, pp. 191-206.

Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B., Hong, S. Y., Lagacherie, P., Lelyk, G., McBratney, A. B., McKenzie, N. J., (2014). GlobalSoilMap: toward a fine-resolution global grid of soil properties, Advances in Agronomy. Elsevier, pp. 93-134.

Bathke, G., Cassel, D. (1991). Anisotropic variation of profile characteristics and saturated hydraulic conductivity in an Ultisol landscape. Soil Science Society of America Journal 55(2), 333-339.

Cambardella, C., Moorman, T., Parkin, T., Karlen, D., Novak, J., Turco, R., Konopka, A. (1994). Field-scale variability of soil properties in central Iowa soils. Soil Science Society of America Journal 58(5), 1501-1511.

Cook, S., Corner, R., Groves, P., Grealish, G. (1996). Use of airborne gamma radiometric data for soil mapping. Soil Research 34(1), 183-194.

Cruvinel, P. E., Flocchini, R. G., Artaxo, P., Crestana, S., Herrmann Jr, P. S. (1999). Elemental analysis of agricultural soil samples by particle induced X-ray emission (PIXE) technique. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 150(1), 478-483.

Gee, G. W., and Bauder, J. W. (1986). "Particle-size analysis," in Methods of Soil Analysis: Part 1. Physical and Mineralogical Methods II. A. Klute (Ed.), Soil Science Society of America, Madison.

Grauer-Gray, J. R., Hartemink, A. E., 2016. Variation of soil properties in a Mollisol profile wall, Digital Soil Morphometrics. Springer, Dodrecht.

Hartemink, A. E., Minasny, B. (2014). Towards digital soil morphometrics. Geoderma 230, 305-317.

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

Jenny, H. (1994). Factors of soil formation: a system of quantitative pedology. Dover Publications, New York.

McBratney, A., Webster, R. (1986). Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. European Journal of Soil Science 37(4), 617-639.

Minasny, B., McBratney, A., Brough, D., Jacquier, D. (2011). Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. European Journal of Soil Science 62(5), 728-732.

Oliver, M. A., Webster, R. (1990). Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information System 4(3), 313-332.

Rabenhorst, M., Parikh, S. (2000). Propensity of soils to develop redoximorphic color changes. Soil Science Society of America Journal 64(5), 1904-1910.

Ribeiro, P.J. Jr., Diggle, P.J. (2016). geoR: Analysis of Geostatistical Data. R package version 1.7-5.2. https://CRAN.R-project.org/package=geoR

Potoglou, D., Kanaroglou, P. S. (2005). Carbon monoxide emissions from passenger vehicles: predictive mapping with an application to Hamilton, Canada. Transportation Research Part D: Transport and Environment 10(2), 97-109.

Six, J., Conant, R., Paul, E. A., Paustian, K. (2002). Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. Plant and Soil 241(2), 155-176.

St-Onge, B., Cavayas, F. (1995). Estimating forest stand structure from high resolution imagery using the directional variogram. International Journal of Remote Sensing 16(11), 1999-2021.

Stockmann, U., Cattle, S., Minasny, B., McBratney, A. B. (2016). Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. CATENA 139, 220-231.

5 CHECKS AND MASS BALANCES FOR *IN SITU* QUANTIFICATION OF SOIL MINERAL COMPOSITION

5.1 Abstract

Soil mineral composition impacts soil behaviour but field estimation of mineral composition until now has been nigh on impossible. To investigate the potential of predicting soil mineral composition *in situ*, fifteen soils representing diverse mineral composition from New South Wales, Australia, were scanned with visible near-infrared (VisNIR) and portable X-ray fluorescence (pXRF) spectrometers to a depth of 1 m at 2.5 cm sampling increments. The presence of phyllosilicate and Fe-oxide species was assessed using a pattern-matching algorithm utilising mineral end-member libraries. Rule-based iterative partitioning was then applied on the recorded pXRF elemental compositions based on known stoichiometry of the minerals to give the abundance of kaolinite, smectite, illite, hæmatite, goethite, $CaCO_3$, gypsum and quartz in a sample. This fusion model was able to correctly identify the most abundant mineral in a sample with 72% accuracy, with the remaining 28% assigned to the second most abundant mineral of the sample. The second, third and fourth most abundant minerals were correctly assigned in 51%, 49% and 48% accuracy respectively. Mineral

predictions were stable under variable moisture and surface conditions, as experienced when scanning samples *in situ* and under air-dry and ground condition. Relative changes in mineral composition within a profile and across horizon boundaries were accurately expressed. The model was able to quantify the abundance of quartz with a κ(linear weighting) of 0.67 and $CaCO_3$ with a κ(linear weighting) of 0. 76 (LCCC = 0.96, RMSE = 2.1 g 100 g^{-1}). The dominant phyllosilicate species was identified correctly with 86% accuracy, although accurate quantification of phyllosilicates and Fe-oxides was not achieved. This may be due to variation in the elements involved in isomorphic substitution and charge balancing of these minerals, non-crystalline species in the sample that were not identified from XRD analysis, or dilution of readings through the presence of lattice water, variable organic matter levels. Local calibration will undoubtedly further improve model outcomes.

5.2 Introduction

Soil mineral composition impacts nearly all soil properties and has a large impact on soil behaviour. To gain insight into soil function, the spatial variability of soil minerals must be captured and explored. Laboratory-based X-ray diffraction (XRD) has been the standard in analysing soil minerals. However, preparing samples for XRD analysis is a labour intensive task. Random powder samples must be ground to a fine powder, and clay samples must be isolated, basally oriented and subject to appropriate pretreatments, e.g. Mg/K saturation, ethylene glycolation, and heating to 550°C. Despite the availability of diffractogram interpretation software, expert interpretation is often required, and the reporting mineral composition remains semi-quantitative.

Field portable XRD devices are available. These field portable XRD devices require samples to be relatively dry and also ground before analysis (Sarrazin *et al.*, 2005). Moreover, it can take several hours to analyse complex materials such as soil. Non-destructive, *in situ* XRD devices are also available although they commonly have a reduced range and are limited to angles greater than 20° 2θ (Gianoncelli *et al.*, 2008). This is problematic as most phyllosilicates have primary peaks at smaller angles (<10° 2θ), thus these devices are less equipped to estimate these important soil constituents. Subsequently, portable XRD devices

do not offer a solution for rapid soil assessment and have been mostly limited to archaeological and lithological investigations (e.g. Uda, 2004; Nakai and Abe, 2012; Cannon *et al.*, 2015).

In recent times, there has been a growing interest in using proximal soil sensors to obtain information on soil systems. Two devices – visible near–infrared diffuse reflectance (VisNIR) and portable X-ray fluorescence (pXRF) spectrometers, have shown potential as field diagnostic devices, as they can provide a wealth of information in a timely manner (e.g. Viscarra Rossel *et al.*, 2011; Weindorf *et al.*, 2014).

Many soil minerals have characteristic absorption features in the VisNIR range (Clark *et al.*, 1990). Studies have successfully predicted a range of minerals in air-dry and ground samples. Brown *et al.* (2006) used boosted regression trees to predict the presence of kaolinite and montmorillonite in the clay fraction on a 0-5 ordinal scale, with 96% and 88% of samples falling within one ordinal unit respectively. Mulder *et al.* (2013) parameterised absorption features in the 2.1 – 2.4 μm band of the VisNIR spectrum and predicted the abundance of kaolinite, dioctahedral mica, smectite and calcite using regression tree analysis. The model worked well in the presence of ancillary minerals not accounted for in the training phase with kaolinite, dioctahedral mica and calcite having RMSE values of less than 8%. Meanwhile, Malone *et al.* (2014) used a pattern-matching algorithm based on diagnostic absorbance of mineral end members first demonstrated by Clark *et al.* (2003). This approach predicted the presence of kaolinite, smectite, illite, hæmatite and goethite, which were used to map terrons in the viticultural study area.

Studies predicting minerals *in situ* are limited. Viscarra Rossel *et al.* (2009) investigated absorbance values at diagnostic wavelengths from continuum removed spectra, collected *in situ*, to gain estimates of kaolinite, smectite, illite, hæmatite, goethite, gibbsite, calcite and attapulgite. The authors describe qualitative agreement with observations from XRD diffractograms but noted that quartz could not be estimated from this approach.

One limitation of VisNIR for comprehensive mineral characterisation is a lack of absorbance features for quartz, and also low-Fe feldspars, in the 350-2,500 nm wavelength range (Clark *et al.*, 1990). To date, no attempt has been made to quantify quartz and feldspars from VisNIR spectra, and subsequently no attempt has been made to offer a full description of sample mineralogy. To provide a full description of sample mineralogy, other solutions must be explored.

Applications of pXRF to pedology and soil science are increasing (Weindorf *et al.*, 2014). Zhu *et al.* (2011) demonstrated pXRF's ability to predict soil textural attributes, which is of interest as texture is related to mineral composition (McKenzie *et al.*, 2004). Another study found that after correcting for Ca associated with calcite, pXRF data could predict soil gypsum content within 6% of laboratory values (Weindorf *et al.*, 2009).

There is a clear benefit in combining the two devices. Visible near-infrared diffuse reflectance spectroscopy gives information on sample colour as well as molecular overtones and combination vibrations (Burns and Ciurczak, 2007). Portable X-ray fluorescence spectroscopy can give an estimate of the concentration of elements with atomic number ≥12, and is also relatively stable under varying moisture conditions (Stockmann *et al.*, 2016). Using the data from both devices, information on the elemental composition of a sample is provided via pXRF, and some molecular information is provided via VisNIR. Weindorf *et al.* (2016) showed that combining the two devices produced the best prediction of calcium carbonate equivalent in a diverse set of arid soils from Spain (RPD = 1.74).

This study investigates the potential of combining VisNIR and pXRF data, using pattern-matching and an elemental mass balance, to characterise the full soil mineral composition of soil profiles *in situ*.

5.3 Materials and methods

5.3.1 Site descriptions

Fifteen sites were chosen that exhibited a diverse range of mineralogy from across the state of New South Wales, Australia. For a full description of the fifteen sites see section 2.3.

5.3.2 *In situ* scanning procedures

At each site a soil pit was excavated, and a 1 m x 1 m surface of the pit wall was prepared. The soils were scanned *in situ* using VisNIR and pXRF at 2.5 cm increments in three vertical transects. Justification for this sampling procedure is given in Chapter 4. Horizons were identified, and samples taken for laboratory analysis and for scanning in under air-dried, ground condition. For a full description of scanning procedures see section 2.2.2.

5.3.3 Data-fusion

Mineralogical prediction was based on a data-fusion approach. The VisNIR spectra were utilised to predict clay content, phyllosilicate speciation and Fe-oxide speciation. The results were then moderated using an elemental mass balance from pXRF data (Fig. 5.1). This approach differs from that presented by Jones and McBratney (2016) in that clay content is predicted using VisNIR alone and not in conjunction pXRF data. Gypsum has also been added to the model. The data-fusion approach was implemented in R (R Core Team, 2016).

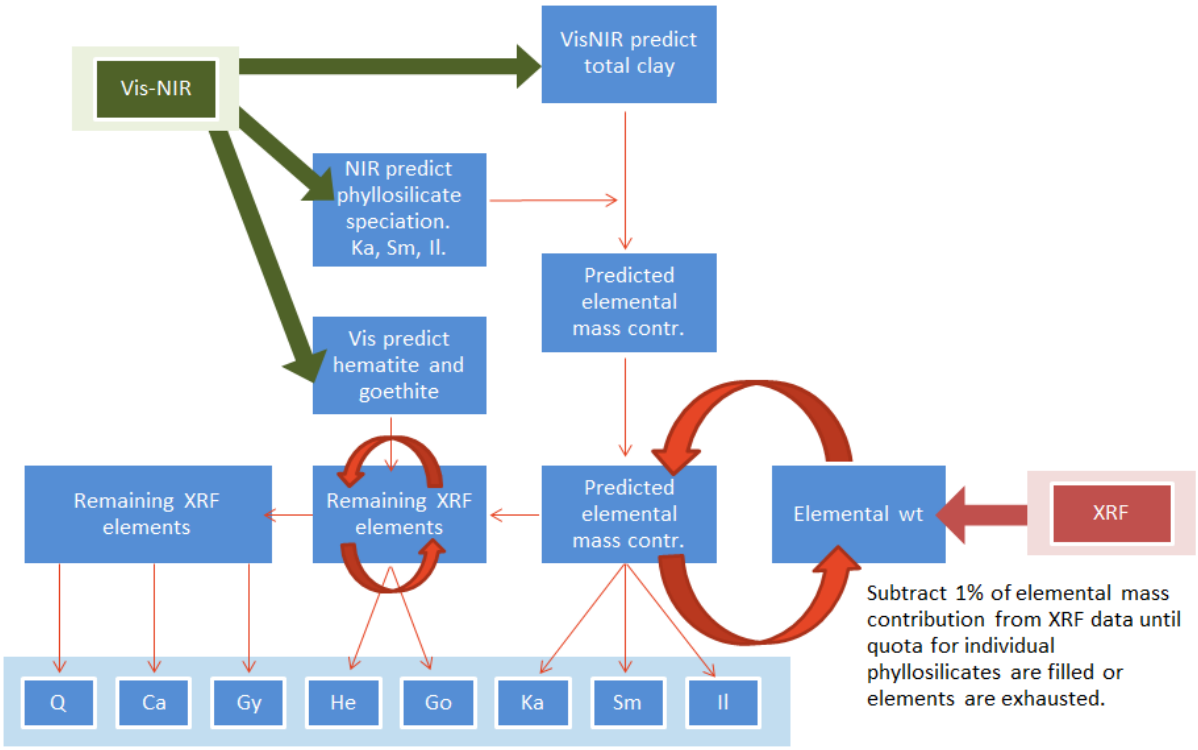


Figure 5.1 The data-fusion approach. VisNIR and pXRF data are input into the model and the mass contributions of individual species to total mineral mass is returned. Ka - kaolinite; Sm - smectite; Il - illite; He - hæmatite; Go - goethite; Gy – gypsum; Ca - $CaCO_3$; and Q - quartz.

Prediction of phyllosilicates (kaolinite, smectite and illite), as well as Fe oxides (hæmatite and goethite), was performed using a shape-fitting algorithm across diagnostic wavelength ranges, to quantify the deviation from convex hull corrected reference mineral spectra. This method is defined in more detail in Malone *et al.* (2014) and previously utilised in the Tetracorder decision-making framework by the U.S. Geological Survey (Clark *et al.*, 2003). The VisNIR spectra were also used to predict total clay concentration using existing Cubist spectral models. Total clay content was then used to scale phyllosilicate speciation ratios to give quantitative predictions of each phyllosilicate species. It should be noted that the soils did not contain any muscovite or biotite, which would influence the prediction of illite, and the clay fraction of these soils consists mainly of secondary minerals (McKenzie *et al.*, 2004)

The average stoichiometry of minerals were used to calculate the proportional element mass contribution of Al, Si, S, K, Ca, Ti and Fe, as well as light elements (LE), which represent constituent elements unquantifiable using pXRF, i.e. $Z \leq 12$ (Table 5.1). Stoichiometric values were derived from the U.S. Geological Survey digital spectral library (Clark *et al.*, 2007) and other sources (Brigatti *et al.*, 2006). The seven elements made up a mean of 99.45% (s.d. = 0.65%) of the mass of total identifiable elements (Z>12) for all *in situ* scans. The required elements for each phyllosilicate species was then tested against the pXRF reported elemental concentrations by iteratively subtracting 1% of the predicted elemental requirement from the available elements. In the event that individual quotas are filled or until a constituent elemental is exhausted, the partitioning of further elements to that species is halted, but the remaining species may continue to draw elements to fill their quota. For example, illite requires contributions from Si, Al and K. In the event that K runs out before the illite quota is filled then the partitioning of elements to illite is halted, but kaolinite may continue to fill its quota if Si and Al are still available.

Once all quotas have been filled or constituent elements exhausted, the moderated values were recorded, and the remaining elements were passed for Fe-oxide prediction in a similar fashion. Residual elements were then used to predict gypsum, $CaCO_3$ and quartz using known stoichiometry of these minerals. Gypsum content was dictated by the least limiting availability of S and Ca. The assumption was made that all remaining Ca was in the form of $CaCO_3$ and that all remaining Si was in the form of tectosilicates, i.e. quartz and feldspars, simplified in this model to quartz. Reported values for individual minerals were divided by the sum of observed mineral mass of a sample to give the proportional mass contribution.

Table 5.1 Average mass-based mineral stoichiometry used in the elemental mass balance.

	Element contribution to total mineral mass (g 100 g^{-1})							
Mineral	LE[†]	Al	Si	S	K	Ca	Ti	Fe
Kaolinite	57.6	20.1	21.2	0.0	0.1	0.0	0.7	0.5
Smectite	58.8	9.8	27.9	0.0	0.3	1.1	0.1	2.1
Illite	54.5	11.8	24.3	0.0	6.3	0.3	0.4	2.6
Hæmatite	30.1	0.0	0.0	0.0	0.0	0.0	0.0	69.9
Goethite	37.1	0.0	0.0	0.0	0.0	0.0	0.0	62.9
$CaCO_3$	60.0	0.0	0.0	0.0	0.0	40.0	0.0	0.0
Gypsum	58.1	0.0	0.0	18.6	0.0	23.3	0.0	0.0
Quartz	53.3	0.0	46.7	0.0	0.0	0.0	0.0	0.0

[†]Light elements (LE) are those with atomic number ≤12, which are unidentifiable by pXRF.

5.3.4 X-ray diffraction

Horizon-based samples were ground to a fine powder, and randomly oriented samples were analysed using monochromatic CuKα radiation at 30 kV and 28.5 mA (GBC MMA diffractometer). The samples were scanned from 4 to 65° 2θ at a speed of 1° 2θ min^{-1} and using a step size of 0.01° 2θ. The clay fraction of samples was isolated using a sedimentation process based on Stoke's Law. Oriented samples from the clay fraction were analysed following four pre-treatments for the identification of phyllosilicate species, i.e. Mg-saturated, Mg-saturated and ethylene glycol solvated, K-saturated, and K-saturated and heated to 550°C (Brown and Brindley, 1980). A randomly oriented sample of the clay fraction was also scanned to aid in the identification of minerals which were masked by the dominant quartz peaks of the bulk sample.

5.3.5 Semi-quantitative XRD interpretation

The presence of kaolinite, smectite, illite, hæmatite, goethite, calcium carbonate, gypsum and quartz were quantified into six ordinal classes representing 0, 0-5,5-20, 20-40, 40-60, 60-80, 80-100% of the total mineral composition. The four most abundant minerals in each sample were identified and used for comparison with fusion model reported values.

5.3.6 Correlation between minerals and pXRF values

Spearman's rank order correlation coefficient (ρ) was used to explore relationships between XRD reported mineral abundance and pXRF reported elemental composition under air-dry, ground condition. Spearman's rank correlation coefficient was chosen over Pearson's correlation coefficient as the XRD ranking scale was not based on a uniform interval and observations were not normally distributed. The Spearman's rank order correlation coefficient describes how well the relationship between two variables can be described using a monotonic function, which is not necessarily linear as is stipulated with Pearson's correlation coefficient. Spearman's rank correlation coefficient is defined as the Pearson correlation coefficient between ranked variables (Eqn 5.1).

Equation 5.1

$$\rho = \frac{\sum_{i=1}^{n}(rx_i - \overline{rx})(ry_i - \overline{ry})}{\sqrt{\sum_{i=1}^{n}(rx_i - \overline{rx})^2}\sqrt{\sum_{i=1}^{n}(ry_i - \overline{ry})^2}}$$

Where:

rx_i , ry_i are the i^{th} ranked variables of rx and ry respectively

$\overline{rx}$, $\overline{ry}$ are the mean values of the ranked variables rx and ry respectively

5.3.7 Evaluation of mineral estimations

In situ mineral estimations were aggregated by horizon and the mean value was compared to XRD estimations of ordinal abundance, by calculating the linear weighted Cohen's Kappa coefficient (Cohen, 1960). This statistic measures the inter-rater agreement between the fusion model and XRD analysis (Eqn 5.2).

Equation 5.2

$$\kappa = 1 - \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} w_{i,j}x_{i,j}}{\sum_{i=1}^{k}\sum_{j=1}^{k} w_{i,j}m_{i,j}}$$

Where:

k is the number of ordinal classes

w, x and m are the weight, observed and expected matrices respectively.

The weight matrix consisted of zeros on the diagonal, cells one off the diagonal are weighted 1, cells two off the diagonal are weighted 2, etc.

5.3.8 Evaluation of carbonate prediction

$CaCO_3$ estimations were further validated using carbonate equivalent values obtained by the rapid titration method (Rayment and Lyons, 2011, pp. 415-7). As these results are continuous in nature, validation metrics such as the Lin's concordance correlation coefficient (LCCC), root-mean-square error (RMSE), and the ratio of performance to deviation (RPD) could be calculated, see section 2.2.6. *In situ* estimations of $CaCO_3$ were aggregated by horizon and mean values compared to laboratory measurements.

The $CaCO_3$ values predicted using the fusion model were modified prior to validation. The $CaCO_3$ contribution to the mineral soil fraction ($[CaCO_3]_{mineral}$) was converted to the $CaCO_3$ contribution of total soil mass ($[CaCO_3]_{soil}$) by incorporating the mass of organic matter (Eqn 5.3).

Equation 5.3

$$[CaCO_3]_{soil} = [CaCO_3]_{mineral} \times \frac{100 - organic\ carbon\ (\text{g } 100\ \text{g}^{-1}) \times 1.72}{100}$$

5.4 Results and discussion

5.4.1 Mineral composition

5.4.1.1 Dominant minerals

The fifteen soils showed diverse mineral composition between sites and frequently between horizons within a profile. Five separate minerals were identified as the most dominant in at least one of the horizons, demonstrating the diversity of soils sampled (Table 5.2). Quartz and kaolinite were the most ubiquitous minerals, both were identified in every sample, although at times only in trace amounts. Quartz was the most dominant mineral in half of the samples, reflecting the parent material and highly weathered nature of many of the profiles. Relationships between mineralogy and soil type were observed. Smectite was the dominant mineral in seventeen samples, corresponding mostly with Vertosol profiles. Fe-oxides were dominant in the Ferrosol. $CaCO_3$ was dominant in two of the subsoil horizons of the Calcarosol. Illite and gypsum were not the most dominant mineral in any soil. Illite was found to be the second most dominant mineral in three horizons. Gypsum was only identified in a single horizon, and thirteen horizons had only three identifiable minerals.

Table 5.2 Count of dominant mineral allocation from XRD analysis of 65 soil horizon samples.

	XRD dominant				
Mineral	I	II	III	IV	Total
Kaolinite	9	32	14	10	65
Smectite	17	4	5	2	28
Illite	0	3	15	21	29
Fe-oxide	5	0	20	14	39
$CaCO_3$	2	1	2	5	10
Gypsum	0	0	1	0	1
Quartz	32	25	8	0	65
Total	65	65	65	52	

Random powder and oriented diffractograms for each sample may be found in Appendices B1 and B2 respectively. The full XRD semi-quantitative composition assessment including auxiliary minerals may be found in Appendix B3.

5.4.1.2 Auxiliary minerals

A number of minerals not directly accounted for in the fusion model were also observed. These were vermiculite, hydroxy-interlayered vermiculite, rutile, anatase and gibbsite. K-feldspar, anorthite and albite were also identified in a number of horizons. However, these feldspars were not found to be dominant in any sample. The low quantity of feldspars in these profiles is to be expected as parent materials were identified as Jurassic sandstones and shales, basaltic alluvium or marl. No parent materials were identified as feldspar-rich granites. Although Site 9 contained trachyte saprolite in the subsoil, feldspars were not identifiable in the soil matrix. Many profiles were also heavily weathered. The inclusion of feldspars to the quartz category in the fusion model is logical, given their similar physical and chemical attributes in soils, and low concentrations observed in these soils. Future efforts should attempt to quantify all auxiliary minerals and individual feldspar species for comprehensive mineral characterisation.

5.4.2 Correlation between mineral composition and pXRF reported elements

Spearman's rank correlation coefficients revealed that mineral abundance was significantly correlated with pXRF reported elemental composition. Each mineral being positively correlated with at least one element, and significant negative correlations were also observed (Table 5.3).

Table 5.3 Spearman's rank-order correlation coefficient between XRD observed ordinal mineral abundance and pXRF reported element composition.

	pXRF elements							
Mineral	LE†	Al	Si	S	K	Ca	Ti	Fe
Kaolinite	-0.16	0.56***	-0.12	0.05	0.01	-0.50***	-0.09	0.33**
Smectite	0.45***	-0.05	-0.29*	-0.10	0.38**	0.78***	-0.13	0.14
Illite	0.05	0.03	0.29*	-0.36***	0.52***	0.09	-0.45***	-0.33**
Hæmatite	0.00	0.62***	-0.57***	0.36**	-0.21	-0.12	0.62***	0.81***
Goethite	-0.36**	0.28*	-0.08	0.35**	-0.47***	-0.60***	0.32**	0.25*
$CaCO_3$	0.38**	-0.07	-0.40***	-0.06	0.23	0.82***	-0.07	0.15
Gypsum	-0.17	-0.11	0.03	0.46***	-0.14	0.19	0.00	-0.04
Quartz	-0.23	-0.54***	0.77***	-0.43***	0.01	-0.44***	-0.19	-0.71***

†Light elements (LE) are those with atomic number ≤12, which are unidentifiable by pXRF.

Significant correlations are indicated below entries at three levels: $p<0.05$*; $p<0.01$**; and $p<0.001$***.

Kaolinite was found to be positively correlated with Al (0.56***). This may be due to two factors. Firstly, kaolinite is a 1:1 layer silicate meaning it has a greater proportion of octahedral alumina sheets compared to 2:1 layer silicates such as illite and smectite. Secondly, kaolinite is the most abundant phyllosilicate in heavily weathered soils, such as Ferrosols, which are also more likely to contain Al in the form of gibbsite or associated with Fe-oxides (Isbell, 1994). This is also supported by a positive correlation of kaolinite with Fe (0.33**).

Illite had a significant positive correlation with K (0.52***). In contrast to other phyllosilicates, illite has a large amount of K ions in the interlayer space of the phyllosilicate to offset the charge associated with isomorphic substitution in the mineral. The strong

correlation here may also be attributable to the lack of significant quantities of other K enriched minerals, such as K-feldspars which were only observed in trace amounts in two horizons.

Smectite had a significant positive correlation with Ca (0.78***). While Ca is not a structural element of smectite it is often found as a hydrated ion in the interlayer space. The correlation between smectite and Ca could also be attributable to the fact that both are easily weatherable and the presence of one indicates favourable conditions for the other. Intuitively, an even stronger correlation was observed between $CaCO_3$ and Ca (0.82***).

Hæmatite had a significant correlation with Fe (0.81***), Al (0.62***) and Ti (0.62***) respectively. This reflects the strong association with Al and Ti oxides in heavily weathered soils (Isbell, 1994). Goethite showed similar but less significant relationships. Goethite was often found in small quantities in sandstone-derived alluvial soils which may have influenced the correlation values and explain why it does not have a significant negative correlation with Si, as observed with hæmatite.

Intuitively, gypsum showed a significant relationship with S (0.46***), although only one sample contained identifiable levels of gypsum. Meanwhile, quartz showed a highly significant positive relationship with Si (0.77***) and a negative or null correlation to all other elements.

These significant relationships between pXRF reported elemental composition and ordinal XRD mineral abundance measures underlie the potential of the device compared to using VisNIR alone. Although the pXRF and VisNIR must be used conjointly to elucidate soil mineralogy as multiple minerals may be correlated with each element, i.e. Al has a significant positive correlation with both kaolinite and hæmatite.

5.4.3 Fusion model predictions

5.4.3.1 High resolution mineral composition estimations

Fusion model predictions gave high resolution estimates of mineral composition within a profile. Mineral composition estimates often matched XRD reported values. Comparing estimated values between two distinct profiles demonstrates the success of the fusion model (Fig. 5.2). Dominant mineral abundance between profiles and relative changes within profiles were well represented. Changes in mineral composition often occurred across horizon boundaries. Fusion model predictions for all sites may be found in Appendix B4.

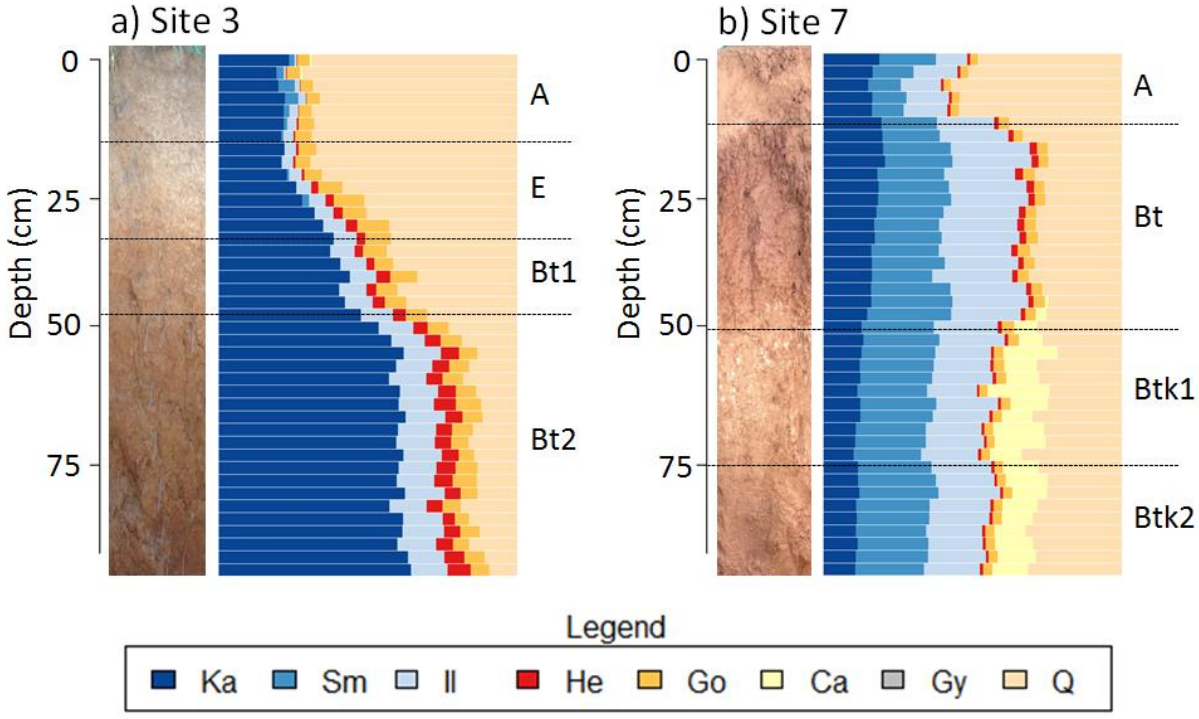


Figure 5.2 Predicted soil mineral contribution to total mineral mass for a) Site 3 and b) Site 7. Ka - kaolinite; Sm - smectite; Il - illite; He - hæmatite; Go - goethite; Ca - $CaCO_3$; Gy – gypsum; and Q - quartz. The average of three vertical transects is presented. Horizon designations are indicated.

X-ray diffraction analysis of Site 3 indicated that the A and E horizons of this profile were dominated by quartz, while kaolinite was the dominant phyllosilicate in the clay fraction and Fe-oxides were identified in trace amounts. In the Bt1 and Bt2 horizons, kaolinite and Fe-

oxides became more dominant, while traces of illite appeared and the overall abundance of quartz decreased. These observations closely matched to the fusion model predictions (Fig. 5.2a).

The A horizon of Site 7 was also observed to be quartz dominated by XRD analysis. Kaolinite, smectite and illite were all observed in the clay fraction, and no Fe-oxides were identified. A texture contrast boundary was observed between the A and Bt. The Bt horizon experienced a decrease in quartz and an increase in all phyllosilicates, as expected. Phyllosilicate content remained constant throughout the B horizons. $CaCO_3$ was observed in Btk1 and Btk2 horizons only. Again, these observations closely matched the fusion model predictions (Fig. 5.2b).

5.4.3.2 Dominant minerals

The fusion model was very effective in identifying the dominant mineral in a sample. After aggregating *in situ* scans by horizon, it was observed that the dominant mineral was correctly assigned with 72% accuracy, while the dominant mineral was incorrectly assigned in 28% of cases to the second most dominant mineral (Table 5.4). Four of the incorrect assignments occurred when quartz was assigned as the dominant mineral in place of kaolinite, however, in each case XRD analysis identified quartz as constituting 20-40% of the mineral phase. One incorrect assignment occurred in the 2Bs horizon of Site 15, where kaolinite was assigned in place of quartz, this may be attributable to the increased Fe and Al in this horizon, incorrectly assigned to kaolinite. Five incorrect assignments occurred in the Ferrosol, where kaolinite and not Fe-oxides, was identified as the dominant mineral. Eight occurred in Vertosols, where XRD showed smectite as the dominant mineral, although smectite was correctly estimated as the dominant phyllosilicate in each instance. This suggests the model could be improved for Vertosols and Ferrosols, which are dominated by smectite and Fe-oxides respectively. The second, third and fourth most dominant mineral was estimated with 51%, 49% and 48% accuracy respectively. When scanning under air-dry, ground condition and with the provision of laboratory-based clay content, the correct allocation was improved to 77%, 60%, 60% and 48% accuracy respectively.

Table 5.4 Contingency table of XRD identified dominant mineral order and those predicted using the approach. The most abundant mineral in a sample is signified with the roman numeral I, the second most with II, et cetera. For thirteen of the samples only three minerals were identified.

Model dominant	XRD dominant			
	I	II	III	IV
I	47	18	0	0
II	18	33	7	4
III	0	7	32	15
IV	0	5	16	25
>IV	0	2	10	8
Total	65	65	65	52

Two of the profiles that gave poor predictions of dominant mineral were in smectite-rich Vertosols. The under-prediction of smectite in these soils may be due to differences in the elements involved in isomorphic substitution and charge balancing (Moore and Reynolds, 1997), which may also impact absorbance features in the VisNIR spectra (Clarke *et al.*, 1990). Analysis of the seven minerals in the USGS library used to construct the smectite mass balance had the following ranges in consituent oxides: 52.0 – 69.6% SiO_2; 15.0 – 20.9% Al_2O_3; and 1.17 – 5.18% Fe_2O_3. In contrast the six minerals used to construct the kaolinite mass balances had much smaller ranges: 42.0 – 47.1% SiO_2; 36.4 – 38.0% Al_2O_3; and 0.26 – 1.61% Fe_2O_3. Smectite was also over-predicted in agriculture topsoil samples. This may be due to enrichment of Ca from plant material, organic matter, fertilisers or lime application. Both of these areas of future improvement for the model.

5.4.3.3 Phyllosilicate prediction by VisNIR and fusion model

VisNIR predictions of phyllosilicate speciation were stable under variable moisture content and surface condition, as observed by scanning *in situ* and also under air-dry, ground

condition in the laboratory. The LCCC values were greater than 0.90 for each phyllosilicate species (Fig. 5.3). This demonstrates that the metrics chosen to estimate phyllosilicate speciation were robust under variable moisture conditions.

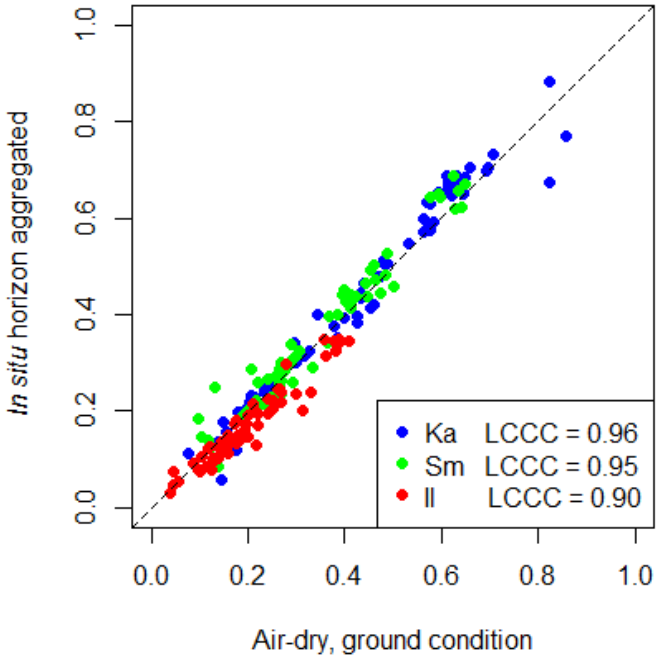


Figure 5.3 Comparison of the proportion of phyllosilicate speciation predicted using VisNIR *in situ* and aggregated to horizon versus horizon-based samples scanned under air-dry and ground condition. Ka – kaolinite; Sm - smectite; Il - illite.

The chosen metrics were also effective at identifying the dominant phyllosilicate. The dominant phyllosilicate was identified correctly using only VisNIR in 85% and 88% of cases *in situ* and in air-dry, ground condition respectively (Table 5.5). However, VisNIR alone was less effective at predicting the ordinal dominance of all three phyllosilicate species in a sample. This was only achieved in 34% of cases in both *in situ* and under air-dry and ground condition in the laboratory. The fusion model outperformed VisNIR alone when estimating the dominance of all three phyllosilicate species. The correct order of phyllosilicates was identified in 63% of cases under both *in situ* and air-dry, ground condition for the fusion

models. However, the estimation of the dominant phyllosilicates in a sample was very effective, the absolute abundance of individual phyllosilicates was poorly estimated.

Table 5.5 Comparison of the ability to predict the dominant phyllosilicates by VisNIR only and using the fusion model.

	VisNIR only		Fusion model	
Phyllosilicate prediction	*In situ*	Air-dry, ground	*In situ*	Air-dry, ground
Dominant	85%	88%	86%	86%
All three	34%	34%	63%	63%

5.4.3.4 $CaCO_3$ and gypsum predictions from the fusion model

The fusion model was effective at predicting $CaCO_3$ equivalent (LCCC = 0.96, RMSE = 2.1 g 100 g^{-1}, RPD = 4.47) (Fig. 5.4). It must be acknowledged that 47 of the 65 samples did not contain observable levels of $CaCO_3$, which may have enhanced validation statistics. The fusion model results compare favourably with other studies that were conducted on air-dry and ground samples. Mulder *et al.* (2013) predicted calcite with an RMSE of 7 g 100 g^{-1} using deconvolution of the 2,100 – 2,400 nm range of the spectrum, combined with regression tree analysis. Weindorf *et al.* (2016) achieved an RPD of 1.74 when attempting to estimate $CaCO_3$ equivalent. This study utilised penalised spline regression and random forest modelling of XRF data and VisNIR spectra. The samples were from an arid region of Spain and $CaCO_3$ equivalent values range from 2.67 to 47.6 g 100 g^{-1}. The same study achieved an RPD of 4.60 for the prediction of gypsum. While the fusion model achieved a perfect correlation with observed gypsum, this was based on only one observation, therefore validation statistics are not offered.

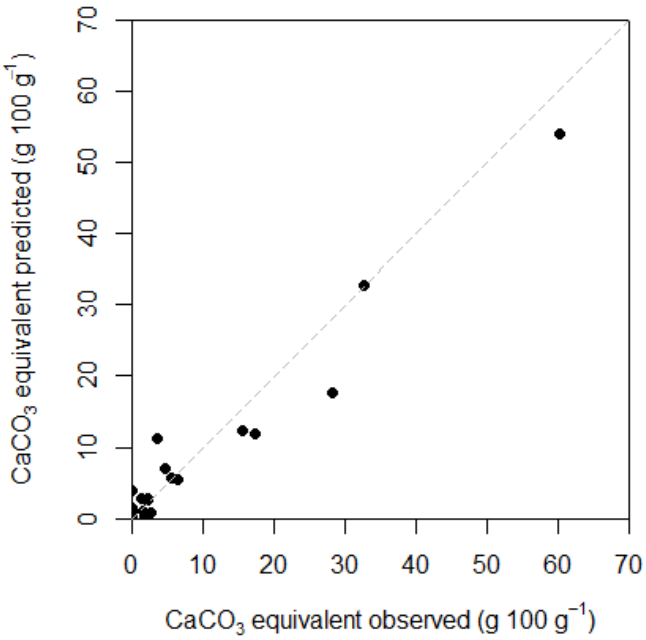


Figure 5.4 $CaCO_3$ equivalent (g 100 g^{-1}) from horizon-based carbonate equivalent data acquired by the rapid titration method, and horizon aggregated *in situ* predictions (LCCC = 0.96, RMSE = 2.1 g 100 g^{-1}).

5.4.3.5 Quartz prediction from the fusion model

Quartz was also estimated with high accuracy, given its κ(linear weighting) of 0.67 (Table 5.6). As quartz has no spectral response in the VisNIR region, the accuracy of this prediction is attributable to the fusion model approach and the strong relationship between quartz and pXRF recorded Si, ρ = 0.77 (Table 5.3).

Table 5.6 Contingency table of semi-quantitative quartz abundance from XRD analysis and the fusion model approach. Grey squares on the diagonal represent a perfect match.

Model estimated	XRD abundance						
	-	tr	x	xx	xxx	xxxx	xxxxx
-	0	0	0	0	0	0	0
tr	0	0	0	0	0	0	0
x	0	2	5	3	0	0	0
xx	0	0	3	16	2	0	0
xxx	0	0	3	3	3	3	0
xxxx	0	0	1	0	2	13	1
xxxxx	0	0	0	0	0	2	3

Key: (-) not present, (tr) trace <5%, (x) 5-20%, (xx) 20-40%, (xxx) 40-60%, (xxxx) 60-80%, (xxxxx) 80-100%.

The accurate estimation of quartz, gypsum and $CaCO_3$ compared to the phyllosilicates and Fe-oxides may be driven by the fact that these minerals have a relatively fixed elemental composition and crystal structure. These minerals are not as heavily influenced by isomorphic substitution as phyllosilicates, or by lack of crystallinity affecting XRD identification as with Fe-oxides.

5.4.4 Model improvements

A number of improvements could be made to the model. The inclusion of gibbsite, Ti-oxides, feldspars and vermiculite must be addressed, as well as any other relevant minerals. Gibbsite and vermiculite have absorbance features in the VisNIR region, which could be incorporated into the fusion model. The lack of absorption features of low Fe feldspars in the VisNIR range will continue to limit their prediction. Distinguishing between hæmatite and goethite

also presented difficulty as identified previously (e.g. Vendrame *et al.*, 2012; Malone *et al.*, 2014). Raman spectroscopy may offer a solution, as it has been demonstrated to provide, under laboratory conditions, a number of absorption peaks related to feldspars and Fe-oxides, as well as quartz and common igneous minerals such as olivine and pyroxene (Freeman *et al.*, 2008). Mid-infrared spectroscopy would also be useful to incorporate as it contains fundamental absorbances of phyllosilicate minerals and calcite (Farmer and Russell, 1964; Janik *et al.*, 1995). Although the mid-infrared spectrum is more susceptible to variable moisture compared to VisNIR, although field portable devices are available (Sorak *et al.*, 2012).

The mass balance is limited in that the pXRF sensor was only able to detect elements heavier than Mg. Other pXRF sensors are available with the capacity to measure Mg in the field, which is an integral component in many of the predicted minerals, and also vermiculite. An alternative to pXRF in the model would be laser-induced breakdown spectroscopy. Laser-induced breakdown spectroscopy is able to provide quantitative estimates of light elements that cannot be quantified using pXRF, including Li (Jantzi and Almirall, 2011). However, laser-induced breakdown spectroscopy is also affected by moisture content, and the amount of sample analysed is in the range of nanograms to picograms, which suggest small heterogeneities in a sample will greatly influence results.

This investigation presented the most comprehensive attempt at estimating the mineral composition of soil *in situ*. The limitation of a comprehensive prediction of mineral abundance is that if one prediction is incorrect, then the accuracy of all other predictions is reduced as well. It is difficult to assess the performance of the fusion model in context with previous studies. Many existing studies give only qualitative description of the strength of the mineral signal and relative changes therein without formal validation (e.g. Viscarra Rossel *et al.*, 2009; Malone *et al.*, 2014). In other instances, validation is only recorded for the clay fraction and not the entire soil (e.g. Brown *et al.*, 2006). Future studies should aim at providing comprehensive mineral estimations, and appropriate validation statistics.

5.5 Conclusion

A new method for *in situ* quantification of soil mineral composition based on VisNIR and pXRF spectroscopy was investigated. This fusion model was centred on pattern-matching of VisNIR spectra to identify the relative abundance of phyllosilicates and Fe-oxides, combined with an elemental mass balance based on pXRF reported elemental composition. Kaolinite, smectite, illite, hæmatite, goethite, $CaCO_3$, gypsum and quartz were estimated, making this model the most comprehensive attempt to estimate soil mineral composition using proximal soil sensors. The fusion model correctly identified the most abundant mineral in a sample with 72% accuracy, with the remainder assigned to the second most abundant mineral of the sample. The second, third and fourth most abundant minerals were correctly assigned with 51%, 49% and 48% accuracy respectively. Prediction of $CaCO_3$, quartz and gypsum were quantitative. The abundance of phyllosilicates was qualitative only, as the model was able to predict the relative dominance of phyllosilicate species, but not their absolute abundance. The model was stable under both *in situ*, as well as air-dry and ground condition, demonstrating it is a viable field technique. Further investigation is required to determine if these predictions can be connected to soil function rather than characterisation.

5.6 References

Brigatti, M. F., Galan, E., Theng, B. K. G. 2006. Structures and Mineralogy of Clay Minerals. In: Bergaya, F., Theng, B.K.G., Lagaly, G. (Eds.), Developments in Clay Science. Elsevier, Amsterdam.

Brown, D. J., Shepherd, K. D., Walsh, M. G., Mays, M. D., Reinsch, T. G. (2006). Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132(3), 273-290.

Brown, G., Brindley, G. (1980). X-ray diffraction procedures for clay mineral identification. In: Brindley, G.W., Brown, G. (Eds.), Crystal structures of clay minerals and their X-ray identification. Mineralogical Society, London.

Burns, D. A., Ciurczak, E. W. (2007). Handbook of near-infrared analysis. CRC Press, Boca Raton, FL.

Cannon, K. M., Mustard, J. F., Salvatore, M. R. (2015). Alteration of immature sedimentary rocks on Earth and Mars: Recording aqueous and surface–atmosphere processes. Earth and Planetary Science Letters 417, 78-86.

Clark, R. N., King, T. V., Klejwa, M., Swayze, G. A., Vergo, N. (1990). High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research: Solid Earth 95(B8), 12653-12680.

Clark, R. N., Swayze, G. A., Livo, K. E., Kokaly, R. F., Sutley, S. J., Dalton, J. B., McDougal, R. R., Gent, C. A. (2003). Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. Journal of Geophysical Research: Planets 108(E12).

Clark, R. N., Swayze, G. A., Wise, R., Livo, K. E., Hoefen, T. M., Kokaly, R. F., Sutley S. J., (2007) USGS Digital Spectral Library splib06a. U.S. Geological Survey, Data Series, 231.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement 20(1), 37-46.

Farmer, V.C. and Russell, J.D. (1964). The infrared spectra of layer silicates. Spectrochimica Acta, 20**,** 1149–1173.

Freeman, John J., et al. Characterization of natural feldspars by Raman spectroscopy for future planetary exploration. (2008). The Canadian Mineralogist 46(6), 1477-1500.

Gianoncelli, A., Castaing, J., Ortega, L., Dooryhee, E., Salomon, J., Walter, P., Hodeau, J. L., Bordet, P. (2008). A portable instrument for in situ determination of the chemical and phase compositions of cultural heritage objects. X-Ray Spectrometry 37(4), 418 423.

Isbell RF (1994) Krasnozems - a profile. Soil Research **32**, 915-929.

Janik, L., Skjemstad, J., Raven, M. (1995). Characterization and analysis of soils using mid infrared partial least-squares .1. Correlations with XRF-determined major-element composition. Soil Research 33(4), 621-636.

Jantzi, S. C., Almirall, J. R. (2011). Characterization and forensic analysis of soil samples using laser-induced breakdown spectroscopy (LIBS). Analytical and Bioanalytical Chemistry 400(10), 3341-3351.

Jones, E.J., McBratney, A.B., 2016. In-situ analysis of soil mineralogy through conjoint use of visible, near-infrared and X-ray fluorescence spectroscopy. In: Hartemink, A.E., Minasny, B. (Eds.), Digital Soil Morphometrics. Springer, Dordrecht.

Malone, B. P., Hughes, P., McBratney, A. B., Minasny, B. (2014). A model for the identification of terrons in the Lower Hunter Valley, Australia. Geoderma Regional 1, 31-47.

McKenzie, N., Jacquier, D., Isbell, R., Brown, K. (2004). Australian soils and landscapes: an illustrated compendium. CSIRO publishing, Melbourne.

Moore, D.M., Reynolds Jr., R.C., 1997. Diffraction and the Identification and Analysis of Clay Minerals, 2nd ed. Oxford University Press, New York.

Mulder, V., Plötze, M., de Bruin, S., Schaepman, M. E., Mavris, C., Kokaly, R. F., Egli, M. (2013). Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1–2.4 μm) and regression tree analysis. Geoderma 207, 279-290.

Nakai, I., Abe, Y. (2012). Portable X-ray powder diffractometer for the analysis of art and archaeological materials. Applied Physics A 106(2), 279-293.

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/.

Sarrazin, P., Blake, D., Feldman, S., Chipera, S., Vaniman, D., Bish, D. (2005). Field deployment of a portable X-ray diffraction/X-ray flourescence instrument on Mars analog terrain. Powder Diffraction 20(02), 128-133.

Sorak, D., Herberholz, L., Iwascek, S., Altinpinar, S., Pfeifer, F., Siesler, H. W. (2012). New developments and applications of handheld Raman, mid-infrared, and near-infrared spectrometers. Applied Spectroscopy Reviews 47(2), 83-115.

Stockmann, U., Jang, H. J., Minasny, B., 2016. The effect of soil moisture and texture on Fe concentration using portable X-ray fluorescence spectrometers. In: Hartemink, A.E., Minasny, B. (Eds.), Digital Soil Morphometrics. Springer, Dordrecht.

Uda, M. (2004). In situ characterization of ancient plaster and pigments on tomb walls in Egypt using energy dispersive X-ray diffraction and fluorescence. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 226(1–2), 75-82.

Vendrame, P., Marchão, R., Brunet, D., Becquer, T. (2012). The potential of NIR spectroscopy to predict soil texture and mineralogy in Cerrado Latosols. European Journal of Soil Science 63(5), 743-753.

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(3–4), 253-266.

Viscarra Rossel, R., Adamchuk, V., Sudduth, K., McKenzie, N., Lobsey, C. (2011). Proximal soil sensing: an effective approach for soil measurements in space and time. Advances in Agronomy 113, 243-291.

Weindorf, D. C., Zhu, Y., Ferrell, R., Rolong, N., Barnett, T., Allen, B., Herrero, J., Hudnall, W. (2009). Evaluation of portable X-ray fluorescence for gypsum quantification in soils. Soil Science 174(10), 556-562.

Weindorf, D. C., Bakr, N., Zhu, Y. (2014). Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications. Advances in Agronomy 128, 1-45.

Weindorf, D., Chakraborty, S., Herrero, J., Li, B., Castañeda, C., Choudhury, A. (2016). Simultaneous assessment of key properties of arid soil by combined PXRF and Vis–NIR data. European Journal of Soil Science 67(2), 173-183.

Zhu, Y., Weindorf, D. C., Zhang, W. (2011). Characterizing soils using a portable X-ray fluorescence spectrometer: 1. Soil texture. Geoderma 167–168, 167-177.

6 ESTIMATING PROFILE CHARACTERISTICS WITH PROXIMAL SENSORS AND A SPECTRAL SOIL INFERENCE SYSTEM

6.1 Abstract

Proximal soil sensors are moving the domain of quantitative soil science from the laboratory into the field. To utilise these sensors effectively platforms must be developed to access sensor information and predict soil properties in near real-time. A framework is presented to estimate soil properties *in situ* from visible near-infrared (VisNIR) diffuse reflectance spectroscopy. These estimates were used to initiate a spectral soil inference system (SPEC-SINFERS) to greatly augment information gain. An automated platform was constructed to link pedotransfer functions in a logical network and predict 29 soil attributes directly from the VisNIR spectra or in conjunction with pedotransfer functions. The platform was implemented on fifteen diverse soil profiles from the state of New South Wales, Australia. At each location, three 1 m vertical transects were scanned using VisNIR at 2.5 cm intervals. Scanning took place under field moist conditions, and observed horizons were sampled for

laboratory analysis. Local versus Global spectral calibration models were tested. Effective characterisation was assessed in terms of organic carbon content, clay, sand, cation exchange capacity, pH, electrical conductivity and bulk density. Local calibration models provided the best estimates for the two profiles within the geographic domain of the calibration library (RMSE values: OC = 0.26 g 100 g^{-1}; clay = 8.7 g 100 g^{-1}; sand = 9.4 g 100 g^{-1}, CEC = 2.0 $cmol_{(+)}$ kg^{-1}; pH (1:5 H_2O) = 0.55; EC = 0.21 dS m^{-1}; BD = 0.11 g cm^{-3}). Transformation using external parameter orthogonalisation improved outcomes *in situ*. However, the best results were achieved when scanning under air-dry and ground condition in the laboratory still provided the best results (RMSE values: OC = 0.13 g 100 g^{-1}; clay = 6.0 g 100 g^{-1}; sand = 7.1 g 100 g^{-1}, CEC = 1.85 $cmol_{(+)}$ kg^{-1}; pH (1:5 H_2O) = 0.48; EC = 0.23 dS m^{-1}; BD = 0.07 g cm^{-3}). Applying models outside of their calibration domain generally doubled RMSE values. Attempts to identify if spectra fit within the spectral domain of the calibration library produced mixed results. In contrast, the size of the prediction interval gave a good indication of model performance. This information may be utilised in the field for improved decision making, including adaptive mapping techniques, management zone delineation and pedogenetic investigations.

6.2 Introduction

The world needs more soil information for use in food (Godfray *et al*., 2010), water (Blanco and Lal, 2010) and soil security (McBratney *et al.*, 2014), climate-change adaptation and mitigation (Lal, 2004), and ecosystem management (Kreuter *et al.*, 2001). Traditional laboratory techniques are time and cost inhibitive to fill this knowledge gap. Provision of sufficient soil information will require data to be gathered more efficiently, and available data to be used more effectively. Devices are available that can provide quantitative soil information in the field (Hartemink and Minasny, 2014). Invariably, these devices are used to gather information in the field, which is then processed into useful soil information off-site. To utilise these devices effectively, systems must be developed to process sensor streams into actionable soil information in near real-time.

One method of using available data more effectively is the employment of pedotransfer functions (PTF) (Bouma, 1989). Pedotransfer functions are empirical equations that build relationships between data to estimate properties which are difficult to obtain, using properties that are easier to obtain. Briggs and McLane (1907) are credited with developing the first recognised PTF in the form of a regression equation to estimate the amount of water retained in the soil when a plant begins to wilt (Eqn 6.1).

Equation 6.1

$$Wilting\ coefficient = 0.01 \times sand + 0.12 \times silt + 0.57 \times clay$$

Soil physical parameters, especially soil hydraulic properties, have received significant attention in PTF development. Saturated hydraulic conductivity, sorptivity and measures of soil moisture at key matric potentials directly, or through construction of soil moisture characteristics, have all received notable attention (e.g. Schaap *et al*., 2001; Cornelis *et al*., 2001). Pedotransfer functions have also proven useful for converting between parallel systems, thus facilitating global data sharing. Padarian *et al.*, (2012) demonstrated how PTFs could be used to convert silt percentages between the international (2 – 20 μm) and USDA (2 – 50 μm) particle size diameter thresholds. Measurements of pH (1:5 H_2O), compared to pH (1:5 $CaCl_2$), have also been reconciled using PTFs (Minasny *et al.*, 2011). Commonly, PTFs take the form of regression equations, although machine-learning techniques such as Cubist models, random forest, neural networks or support vector machines may also be utilised (e.g. Lamorski *et al*., 2008; Twarakavi *et al*., 2009).

Environmental models link soil properties, PTFs and other biotic and/or abiotic factors to model complex systems, such as sediment transport (Morgan *et al.*, 1998) and soil carbon dynamics (Coleman and Jenkinson, 1996). McBratney et al., (2002) proposed to further broaden the use of PTFs, which were commonly focussed on estimating a single property for a single soil type, in a limited geographical region. The authors suggested linking PTFs to form a directional network, or soil inference system (SINFERS). In such a network, the output from one PTF may act as the input for another PTF, and the number of properties

estimated increases dramatically. This first demonstration of SINFERS consisted of a large Excel spread sheet that linked laboratory measured organic carbon, sand, clay and bulk density with a logical sequence of PTFs. This allowed the prediction of a large number of physical and chemical attributes from the limited supply of input properties.

The concept of SINFERS was further extended with the advent of a spectral soil inference system (SPEC-SINFERS) (McBratney *et al*., 2006). This advancement is characterised by SINFERS being initiated with spectroscopic methods, instead of laboratory or field observations. Spectroscopic methods have the advantage of being cost-effective and timelier than laboratory techniques, and they are more objective than field observations. Additionally, a single spectrum can hold information on many soil properties (Viscarra Rossel *et al*., 2010). While spectroscopy can provide a vast amount of information about a soil, two limitations remain: i) not all attributes of interest are spectrally active; and ii) the development of calibrations for all properties may be restricted. Restrictions may be imposed when an existing spectral library is available, however measures of hydraulic conductivity were not obtained at the site or prior to air-drying and grinding the samples. These are the situations where SPEC-SINFERS may be implemented. Tranter *et al*. (2008) demonstrated the potential of SPEC-SINFERS. The authors identified that moisture retention of soil was more accurately estimated by coupling mid-infrared predicted basic soil properties with PTFs, rather than attempting to estimate moisture retention directly from the spectra. This system proved especially effective at lower matric potentials, where texture was the dominant influence on volumetric moisture content. If such a system were to be initiated using proximal soil sensors it would be fully functional in the field.

This study assesses the effectiveness of combining VisNIR readings with pre-calibrated spectral models and SPEC-SINFERS to predict a large number of soil properties *in situ* and in near real-time.

6.3 Materials and methods

To run SPEC-SINFERS, a number of different disciplines and techniques must be brought together. An overview of the approach is shown in Figure 6.1; detailed description of each process follows.

6.3.1 Spectral libraries

Two calibration libraries were tested for this investigation: a Local library based on samples from within a limited geographic region only, and a Global library formed by supplementing the Local library with addition datasets, primarily from within the state of New South Wales (Fig. 6.2).

6.3.1.1 Local library description

The Local library was derived from two soil surveys in a semi-arid, cotton-growing region, centred around the township of Hillston, NSW. Soil samples were taken from all land uses to a depth of 150 cm (Onus *et al.*, 2003; Filippi *et al.*, 2018). There are 1,292 samples in the library, with varying numbers of observations for the properties of interest. The total land area of the study site is 265,000 km^2.

6.3.1.2 Global library description

The Global library was constructed utilising the Hillston library and four other independent datasets:

Dataset 1 consisted of 391 horizon-based samples from agricultural areas in southern New South Wales and Victoria (Geeves *et al.*, 1995; Minasny *et al.*, 2009; Minasny *et al.*, 2011). The samples represented a diverse range of soil types in the study area. Samples located in Victoria are not shown.

Dataset 2 consisted of 464 samples from 100 sites within the Hunter Valley, NSW (Malone *et al.*, 2011). Sites were selected via strata-based sampling. Samples were taken from depth intervals corresponding to the GlobalSoilMap project to a maximum depth of 1 m.

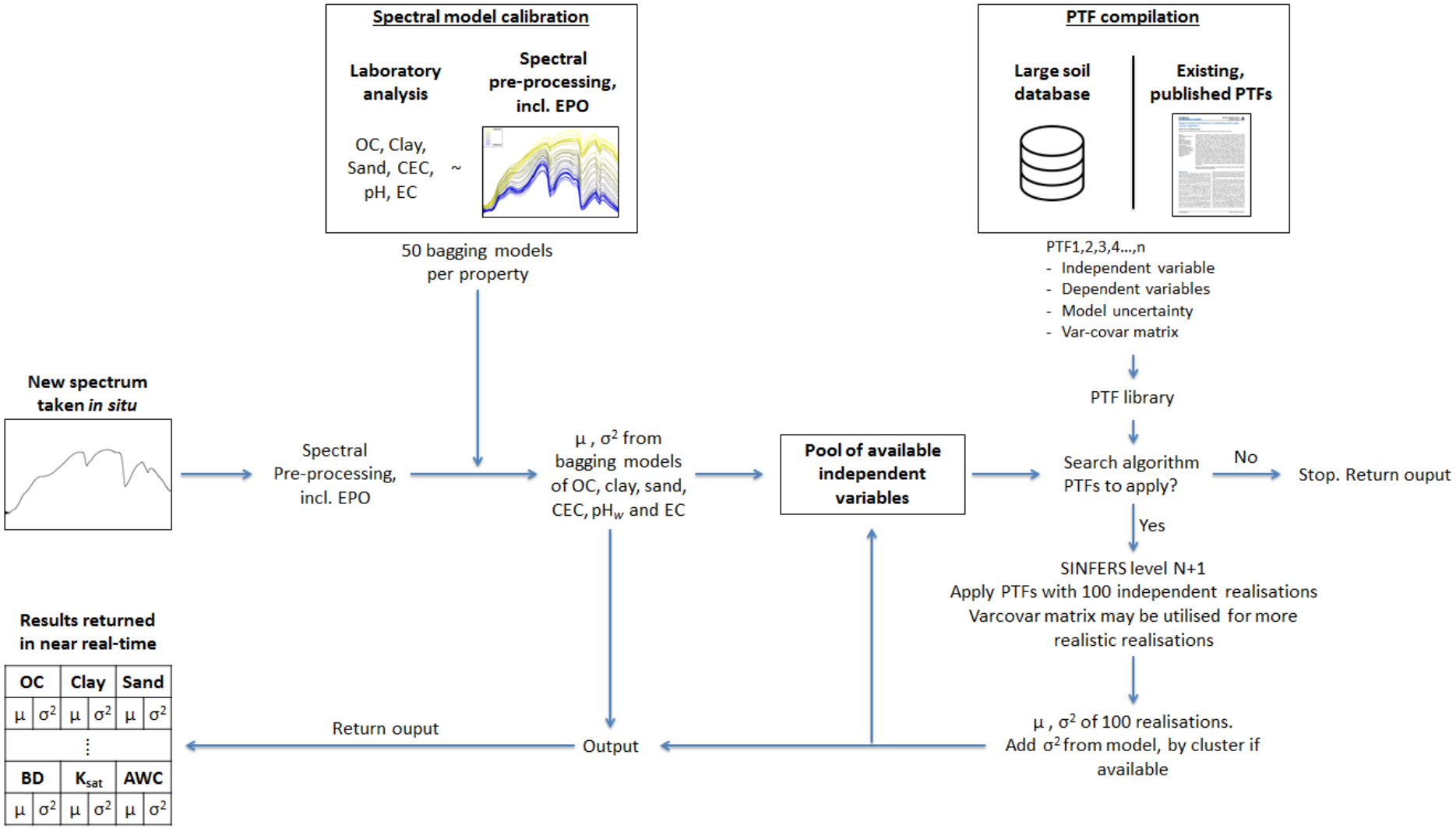


Figure 6.1 Schematic representation of the SPEC-SINFERS approach.

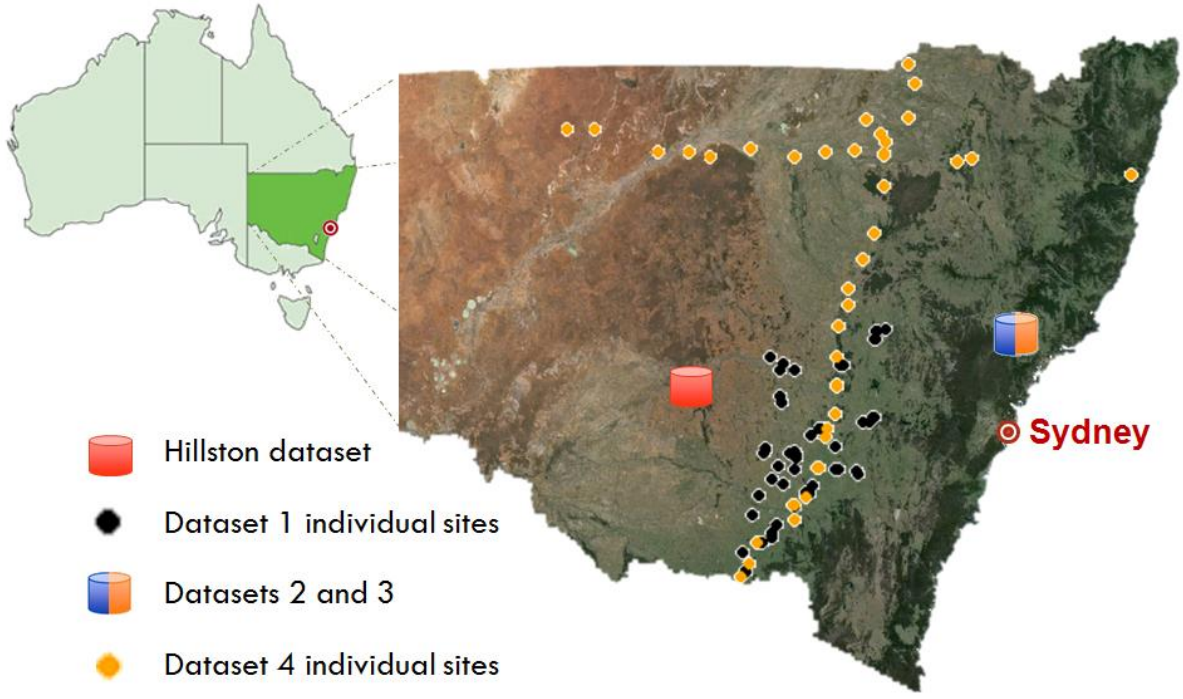


Figure 6.2 Location of individual and intensive sample sites within the state of New South Wales. Points indicate individual sites, drums represent databases of intensively sampled locations. The Hillston dataset serves as the Local library for this research.

Dataset 3 consisted of 200 samples from 56 sites within the Hunter Valley, NSW (Fajardo *et al.*, 2016). Sites were determined in a previous investigation using conditioned Latin hypercube sampling (Minasny and McBratney, 2006). Samples represent 2 cm sections of soil cores.

Dataset 4 consisted of 247 samples derived from two large transects dissecting the state of New South Wales (Pino, 2016). The North-South transect follows a 550 mm mean annual rainfall isohyet, meanwhile the East-West transect experiences a mean annual rainfall gradient of >1,500 mm to <300 mm. Paired sites were sampled to reflect agricultural and natural use.

All samples were scanned with the same AgriSpec^{TM} spectrometer (Analytical Spectral Devices, Boulder, Colorado, USA). Samples were scanned with 3-5 replicates per sample, a baseline white reference reading was taken every 15-21 scans and the corresponding reflectance values of the replicates were averaged. Further description of the scanning for each dataset may be found in Filippi *et al.* (2018), Minasny *et al.* (2011), Malone *et al.* (2011), Fajardo *et al*. (2016) and Pino (2016).

6.3.2 Cubist models

6.3.2.1 Spectral pre-processing and external parameter orthogonalisation

Reflectance spectra of the spectral libraries were subjected to a number of pre-processing techniques: conversion to absorbance; Savitzky-Golay filtering; standard normal variate calculation; cropping to 500 – 2,450 nm; and resampling at 10 nm intervals. Libraries were then either left as pre-processing only (PP) or subject to external parameter orthogonalisation (EPO) to reduce the effects of moisture (Minasny *et al*., 2011). Four libraries were constructed in total, reflecting combinations of the Local versus Global libraries, with PP or transformed using EPO. For a full description of pre-processing techniques and EPO see sections 2.5.5.1 and 2.5.5.2.

6.3.2.2 Producing calibration and validation datasets

Calibration and validation sets were constructed independently for each property, as available observations varied between and within individual spectral libraries (see Table 6.1). Samples were split into 80% calibration and 20% validation sets using conditioned Latin hypercube sampling (Minasny and McBratney, 2006). The first four principal component scores of the pre-processed spectra were used as input for conditioned Latin hypercube sampling, which was implemented in the R platform for statistical computing using the 'clhs' package (Roudier, 2011). The same calibration and validation sets were used for the construction of both PP and EPO models.

6.3.2.3 Producing 50 models via bootstrap aggregating

A bootstrap aggregating (bagging) method (Breiman, 1996) was applied to develop 50 cubist models for each property. Bagging is a special case of model averaging that improves the stability of machine-learning algorithms, by reducing the influence of over fitting. Bagging also allows the construction of measures of uncertainty to accompany the bagging estimate, based on the spread of individual estimates. External validation was achieved by comparing the bagging estimate with laboratory data for the validation set.

6.3.3 Compiling pedotransfer functions

The PTF library was established from published regression equations and some internally developed PTFs. Each variable was given a unique identifying code and units for each variable were fixed to avoid any calculation errors (Appendix C1). For a full list of the PTFs used in this study see Appendix C2.

6.3.3.1 Entering PTF information

Each PTF was recorded with the following information. Only the first field is mandatory.

1. The PTF written in R language

   The PTF must be written in the form of a function executable in R. The function may be a simple regression equation, a model executable using the *predict* function, or any specialised function. Individual PTFs are stored externally and executed using the *source* function when required.

   The system will split a PTF into the following:

   a. A vector containing the dependent variable

      Required to determine if multiple PTFs are available for a dependent variable.

   b. A vector containing each of the independent variables

   Required to identify when a PTF can be executed. If all independent variables are available, then the PTF is executed.

An example of a PTF written in R script is shown in Equation 6.2. This PTF is registered as PTF_043 in Appendix C.

Equation 6.2

```
ptf <- "ph_cacl2 = -0.05 + 0.9 * ph_h2o + 0.14 * log(ec_1_5)"
```

The system will source the PTF and split the character vector at the equals sign to identify the dependent variable *ph_cacl2*, which is the unique identifying code for pH (1:5 $CaCl_2$). The system will then search on the right hand side of the equals sign and identify *ph_h2o* and *ec_1_5* as the independent variables. These unique identifying codes equate to pH (1:5 H_2O) and EC (1:5 H_2O) dS m^{-1} respectively. The model will not identify the *log* function as an independent variable as it was not established as a unique identifier code. Care must be taken to ensure that there is not overlap between unique identifier codes and R functions.

2. Measure of uncertainty

   A measure of uncertainty is required for each PTF. Ideally, this is the variance of the residuals from the validation set of the PTF. If the variance is not provided, then model uncertainty is estimated as the square of the RMSE of the SINFERS estimates. This uncertainty is added to the existing SINFERS variance for future Monte Carlo realisations.

3. Clustered uncertainty

   k-means clustering can be used to further refine the propagation of uncertainties by identifying differences in PTF variance by cluster. Cluster centroids and the correct variance are carried into SINFERS and an extragrade uncertainty is estimated as twice the maximum cluster variance.

4. Variance-covariance table

Provision of a variance-covariance table allows more realistic Monte Carlo realisations to occur which may increase the overall accuracy of SPEC-SINFERS.

Restrictions on the execution of individual PTFs can be implemented based on the range of input values or locality if desired.

6.3.4 Running SPEC-SINFERS

When a new reflectance spectrum is supplied, SPEC-SINFERS is initiated. The spectrum is pre-processed, and an EPO transformation may be applied in accordance with the spectral calibration models to be utilised. The 50 Cubist bagging models are initiated, returning the mean and estimate variance of OC, clay, sand, CEC, pH and EC. These VisNIR predicted properties are then transferred to the pool of available independent variables. The system searches through each PTF's independent variable list and identifies which PTFs have all of their independent variables available in the pool. Each viable PTF is then executed using 100 Monte Carlo realisations; the realisations are based on variance-covariance matrices if provided. If multiple PTFs are available to predict a property, then each realisation is inversely weighted by its uncertainty, as ensemble methods have been demonstrated to outperform single PTFs (Guber *et al*., 2009). The PTF uncertainty is added to the estimate uncertainty to provide a more realistic measure. The SINFERS predicted properties and their associated uncertainty then become available in the pool of independent variables. The search algorithm is initiated again, and the process proceeds until no new PTFs can be called, at which point all estimates and their associated uncertainties are returned.

As the entire process is automated the SPEC-SINFERS network can be rapidly updated or edited (Fig. 6.3).

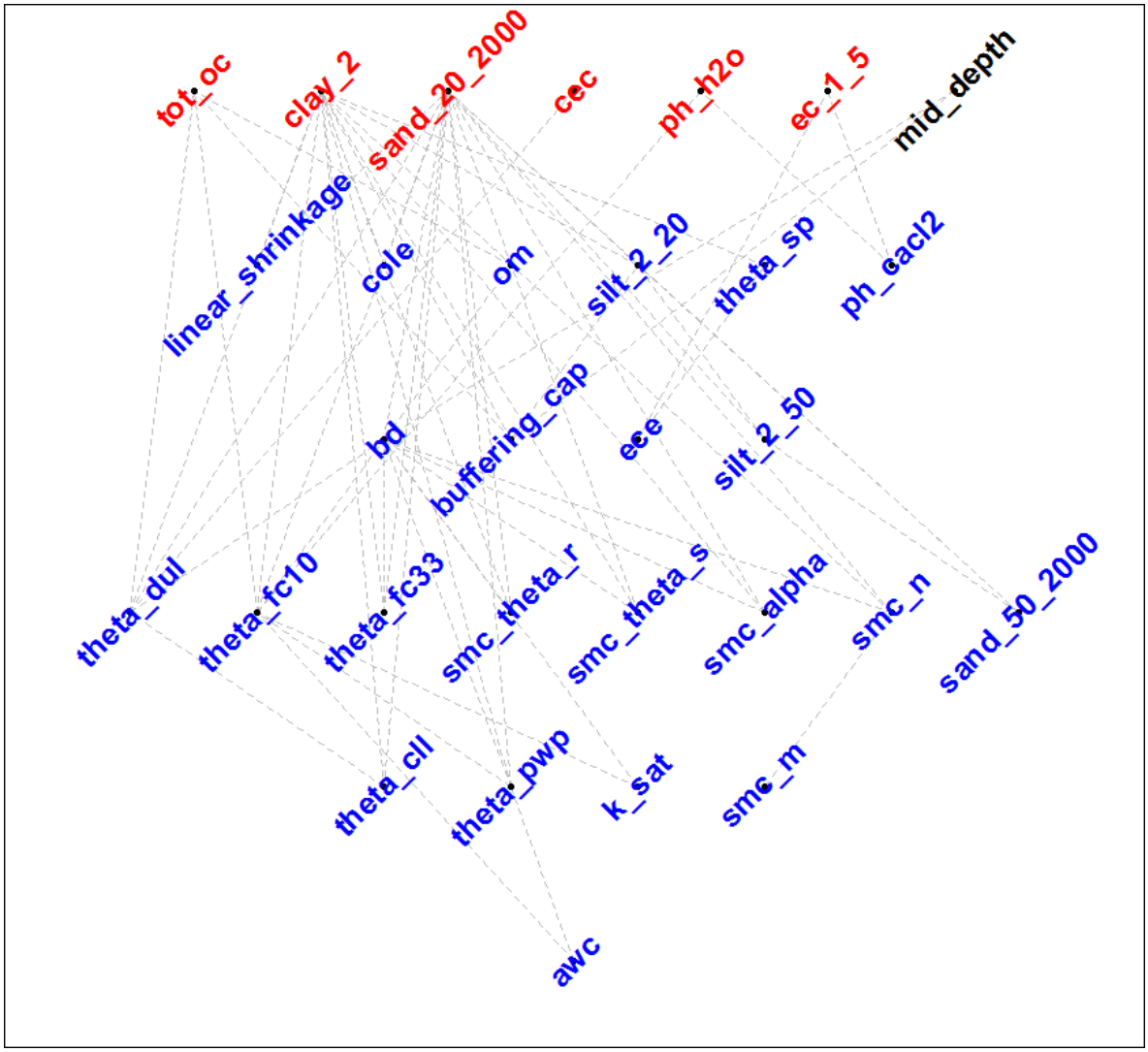


Figure 6.3 A SPEC-SINFERS network. Red properties are predicted using VisNIR in situ. Black properties are provided by user input, e.g. mid-depth of sample. Blue properties are predicted using the SPEC-SINFERS network. Dotted grey lines indicate an input variable to one or more PTFs to predict the property on a lower tier.

Additional functionality has been added to the system for querying the network. When a user clicks on a property, the SPEC-SINFERS pathway to predict this property is highlighted. An example is given in Figure 6.4. By selecting "k_sat" (saturated hydraulic conductivity), a user can observe the network path used to predict k_sat, progressing down each tier from the top to the bottom of the network. In this instance, OC, clay, sand and pH (1:5 H_2O) are predicted from the VisNIR spectra, and sample mid-depth is provided by the user. Organic matter

(OM) is predicted from OC on lower tier. Mid-depth, OM and sand content are then used to predict bulk density on the third tier. Bulk density, OC, clay, sand and pH are then used to predict $\theta_{field\ capacity}$, which is in turn combined with bulk density to predict K_{sat}. The automation of this SPEC-SINFERS network is a major advantage, as constructing such a network manually and updating such a large network as new PTFs became available would be tedious.

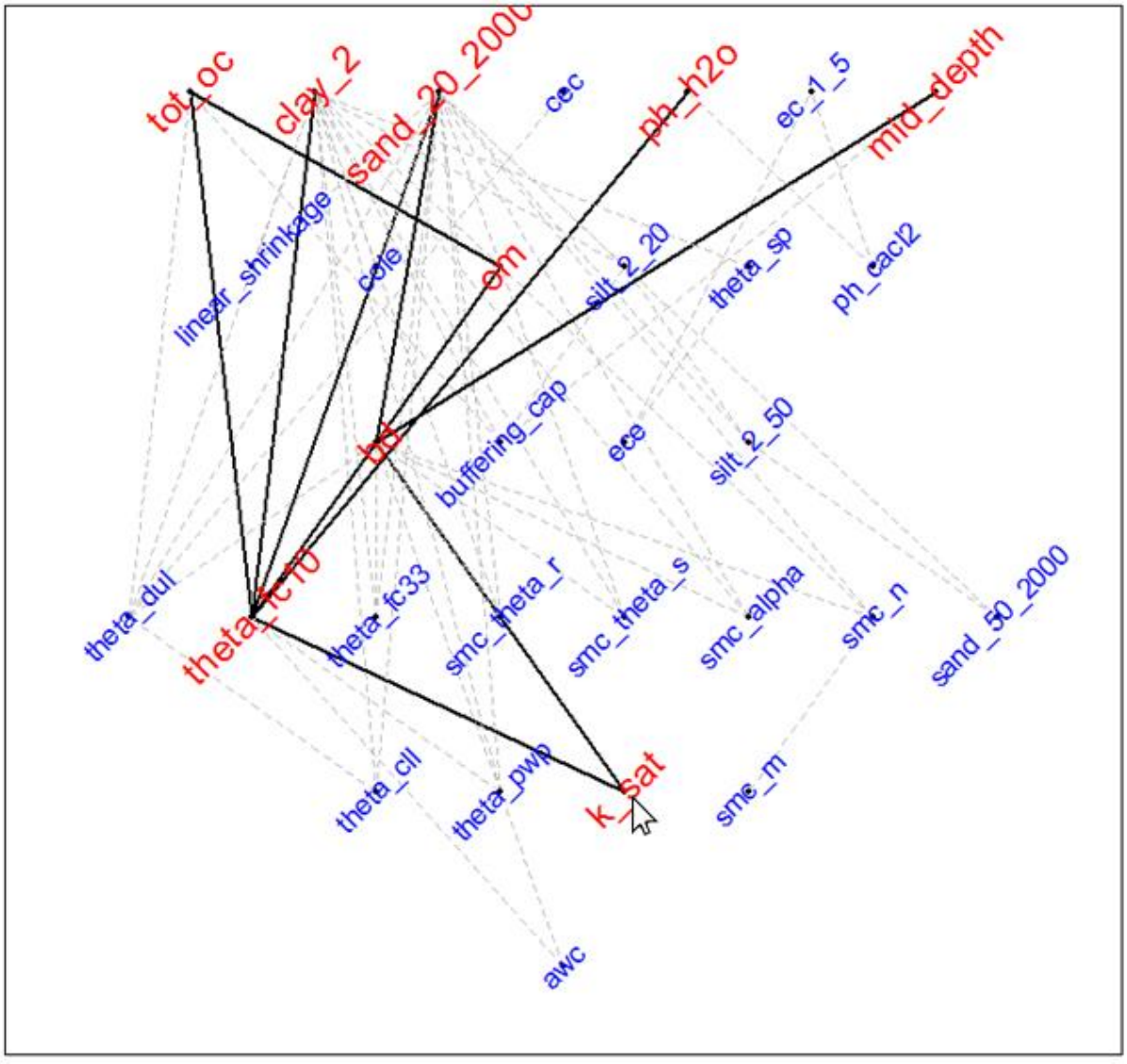


Figure 6.4 Querying the SPEC-SINFERS network. By clicking on "k_sat", the properties used to predict k_sat (saturated hydraulic conductivity), are highlighted in red and the network path is indicated by black lines. The prediction path proceeds from the top of screen to the bottom.

New PTFs and properties may be readily provided and modified using the purpose built graphical user interface. The user interface also facilitates editing of properties, including removing them completely if required. Such a scenario can be readily visualised if clay is removed from the properties predicted using spectral models (Fig. 6.5). The importance of clay in the underlying PTFs is demonstrated. Without a measurement of clay, the only properties that can be estimated are the remaining five properties estimated from the spectral calibration models, as well as organic matter, pH (1:5 $CaCl_2$) and bulk density.

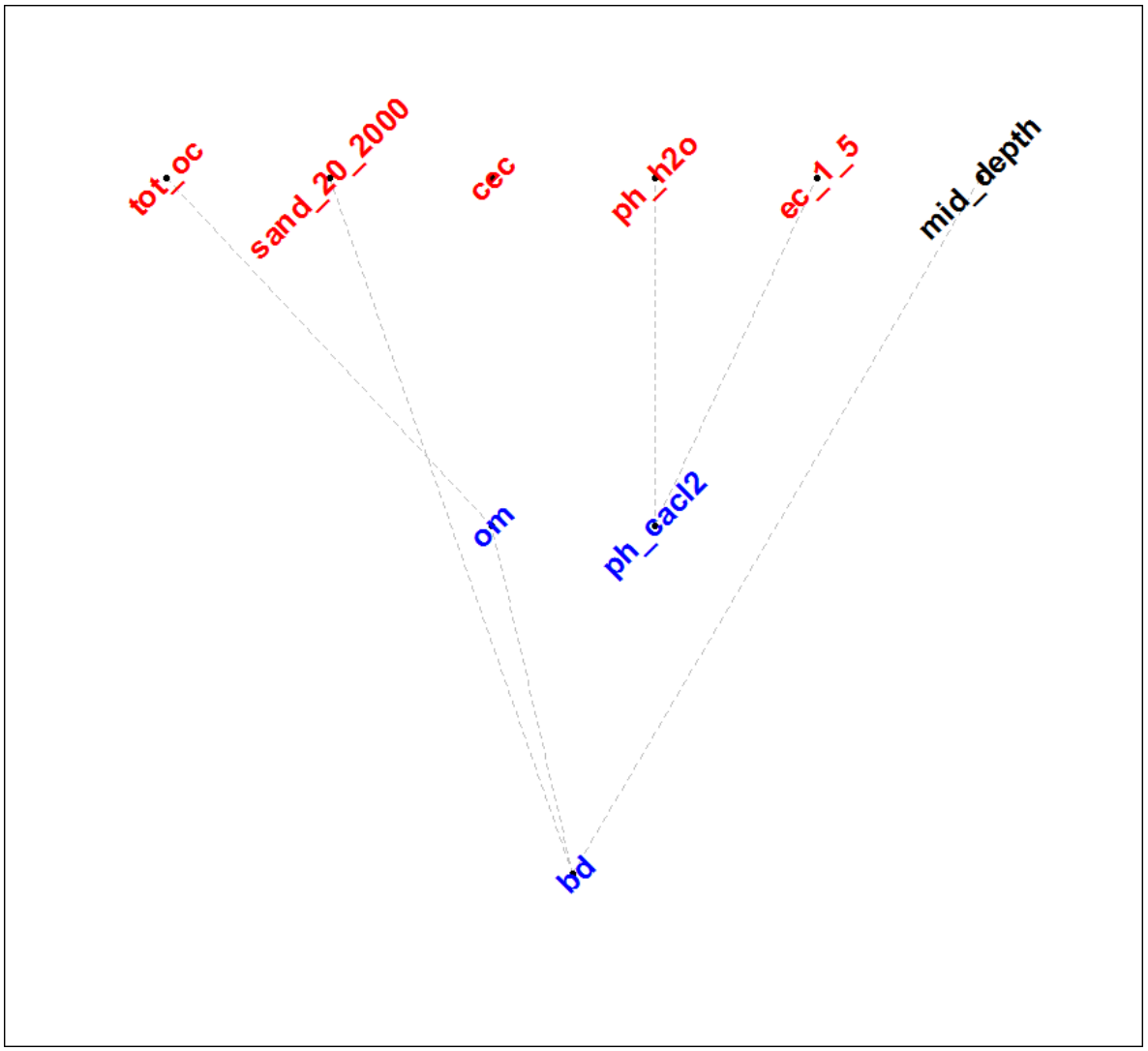


Figure 6.5 The SPEC-SINFERS network without a clay prediction.

6.3.5 Running SPEC-SINFERS *in situ*

6.3.5.1 Site descriptions

Fifteen sites were chosen that exhibited a diverse range of soil properties and climates from across the state of New South Wales, Australia. For a full description of the fifteen sites see section 2.3.

6.3.5.2 *In situ* scanning procedures

At each site, a soil pit was excavated, and a 1 m x 1 m surface of the pit wall was prepared. The soils were scanned *in situ* using VisNIR, in three vertical transects, at 2.5 cm increments. Horizons were identified, and samples taken for laboratory analysis and for scanning in under air-dried, ground condition. For a full description of scanning procedures see section 2.2.2, for justification of the sampling procedure see Chapter 4.

6.3.6 Validation of *in situ* results

In situ SPEC-SINFERS estimates were aggregated by horizon and the mean bagging estimate was compared to laboratory data. Air-dry and ground horizon samples were scanned in the laboratory, and analysed as a point of reference for the performance of *in situ* scans.

6.3.6.1 Convex hull screening

When applying models to a sample it must be assessed if it is represented in the spectral domain of calibration library. Applying models outside of their calibration range can lead to unexpected results. A convex hull was constructed from the first two principal component scores of the calibration library. New points are projected onto the same principal component space. If the new point lies outside the convex hull it is flagged as an outlier, as it does not fall within the spectral domain of the calibration library. Performance of the models on those samples inside and outside of the convex hull is assessed.

6.4 Results and discussion

6.4.1 Model calibration

6.4.1.1 Calibration, validation and pit properties

Analysis of the datasets used for calibration and validation indicate that a diverse range of soils were present in both the Local library (Table 6.1) and Global library (Table 6.2). The semi-arid climate of Hillston is reflected in the low OC found in the Local library. Apart from OC, the ranges of properties of the Local library are relatively large.

6.4.1.2 Performance of calibration models

The bagging approach maintained or improved validation statistics compared to using a single Cubist model for both the Local library (Tables 6.3 and 6.4) and the Global library (Tables 6.5 and 6.6). In each case, the bagging approach was equally or more accurate in terms of larger LCCC and smaller RMSE values. This demonstrates the strength of the bagging approach at producing models that are more robust than a single model. An average decrease in RMSE of 13.3% was observed across all models. A maximum decrease in the RMSE value of 26.4% was observed for sand in the Global library after EPO transformation. Clay in the Hillston library without EPO transformation was the only model that experienced an increase in RMSE using the bagging approach, although the increase was only 2.5%.

There was no discernible difference in performance between models calibrated on the EPO transformed spectra and the PP spectra. This demonstrates that spectral information is preserved in the transformation process. It should be noted that EPO models were developed on dry samples only, the performance the models on field moist spectra is assessed in section 6.4.2.

Table 6.1 Summary of the Local library (Hillston dataset) used for calibration and validation, and of the two soil profiles from the Hillston area sampled *in situ*. OC – organic carbon, CEC – cation exchange capacity, EC – electrical conductivity, BD- bulk density, Cal. – calibration datasets, Val. – validation dataset, Prof. – the two profiles being investigated.

	OC			Clay			Sand			CEC			pH_w			EC			BD
	(g 100 g^{-1})			(g 100 g^{-1})			(g 100 g^{-1})			($cmol_{(+)}$ kg^{-1})			(1:5 H_2O)			(dS m^{-1})			(g cm^{-3})
	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Prof.
Minimum	0.09	0.21	0.00	4.4	9.5	17.7	0.0	19.7	21.0	2.0	2.8	4.8	5.1	5.0	6.4	0.01	0.02	0.08	1.28
1st quantile	0.38	0.43	0.00	46.7	47.0	54.4	31.8	31.9	24.7	18.9	16.0	16.0	7.8	7.8	7.2	0.12	0.13	0.17	1.57
Median	0.58	0.61	0.23	53.3	53.0	58.6	35.8	35.9	32.7	25.7	22.2	16.5	8.5	8.5	8.0	0.22	0.23	0.21	1.67
Mean	0.59	0.65	0.25	48.9	48.7	54.4	40.0	39.1	36.5	24.7	22.8	15.5	8.3	8.3	8.0	0.43	0.40	0.24	1.65
3rd quantile	0.74	0.82	0.38	56.6	55.8	62.3	43.1	41.6	40.0	31.2	30.1	18.0	9.0	9.0	8.9	0.44	0.46	0.28	1.75
Maximum	1.75	1.77	0.66	64.4	63.7	67.0	94.7	88.8	75.9	50.7	53.8	18.0	10.1	9.9	9.6	4.53	4.04	0.53	1.86
Count	318	81	8	307	77	8	308	77	8	417	106	8	1030	261	8	1030	261	8	44

Table 6.2 Summary of the Global library used for calibration and validation, and of the fifteen soil profiles sampled *in situ*. OC – organic carbon, CEC – cation exchange capacity, EC – electrical conductivity, BD- bulk density, Cal. – calibration datasets, Val. – validation dataset, Prof. – the two profiles being investigated.

	OC (g 100 g^{-1})			Clay (g 100 g^{-1})			Sand (g 100 g^{-1})			CEC ($cmol_{(+)}$ kg^{-1})			pH_w (1:5 H_2O)			EC (dS m^{-1})			BD (g cm^{-3})
	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Cal.	Val.	Prof.	Prof.
Minimum	0.00	0.00	0.00	0.0	0.0	4.8	0.0	19.0	14.5	0.4	1.6	0.3	4.4	4.7	4.8	0.01	0.01	0.01	0.64
1st quantile	0.36	0.36	0.18	13.0	11.8	18.4	34.0	33.2	26.4	5.8	5.7	4.3	6.6	6.2	5.8	0.08	0.08	0.03	1.43
Median	0.64	0.68	0.42	29.0	29.8	38.9	43.7	44.3	43.1	10.6	10.7	9.5	8.0	7.5	6.5	0.17	0.18	0.08	1.55
Mean	0.99	0.98	0.75	31.1	30.7	39.8	48.1	49.7	48.9	13.3	13.3	11.2	7.7	7.4	6.8	0.33	0.33	0.16	1.52
3rd quantile	1.16	1.25	0.72	51.2	49.9	58.4	63.0	65.5	68.4	18.2	18.9	17.8	8.8	8.5	7.6	0.35	0.38	0.14	1.67
Maximum	7.22	6.42	6.80	77.0	73.0	80.4	94.7	88.8	94.2	55.1	46.3	31.0	10.1	9.7	9.6	4.53	4.20	3.49	1.87
Count	1171	296	65	1020	260	65	621	155	65	648	166	65	1634	415	65	1789	448	65	330

External validation showed that robust models were made for both the Hillston and Global dataset for most properties. The Hillston library produced LCCC values ≥ 0.80 for all properties. The Global library produced robust models for OC, clay, sand and CEC. However, EC and pH in the Global library only produced LCCC values between 0.65 – 0.72.

Organic carbon, clay, sand and CEC produced similar validation statistics for both the Hillston and Global libraries, reflecting the underlying physical nature of these calibrations. Meanwhile, pH and EC produced noticeably better validation statistics in the Hillston library compared to the Global library. This finding reinforces the fact that these properties are not always spectrally active; rather they may be correlated locally with other spectrally active components. For example, Hillston soils that are higher in the landscape are generally redder, have less clay and have a larger EC as a product of the high evapotranspiration deficit in the area. Meanwhile, the less red, clayey soils found lower in the landscape have relatively smaller EC, as soluble salts are flushed during flood events into the underlying aquifer. An EC model calibrated in such an area would not transfer into a higher rainfall location. Under a high rainfall scenario, a redder soil with less clay would typically experience greater internal drainage rates and, subsequently, smaller EC than a corresponding less red and higher in the landscape clay soil.

While it was stated above that pH and EC are generally not spectrally active, some pH and EC influencing factors are spectrally active. E.g. $CaCO_3$ is spectrally active and will only be found in soils in alkaline pH conditions, meanwhile gypsum is also spectrally active, and will produce a high EC reading.

Table 6.3 External validation of PP models for the Local library

		R^2	LCCC	MSE	RMSE	Bias	MSE_c	$RMSE_c$	RPD	RPIQ
OC	Single	0.54	0.73	0.04	0.20	0.00	0.04	0.20	1.44	0.88
	50 models	0.70	0.80	0.03	0.16	0.01	0.03	0.16	1.82	1.12
Clay	Single	0.80	0.86	31.33	5.60	1.08	30.15	5.49	2.19	1.07
	50 models	0.78	0.86	32.98	5.74	0.63	32.57	5.71	2.13	1.04
Sand	Single	0.68	0.82	53.70	7.33	0.78	53.09	7.29	1.68	0.54
	50 models	0.76	0.86	36.63	6.05	0.87	35.87	5.99	2.04	0.65
CEC	Single	0.80	0.84	6.38	2.52	-0.50	6.12	2.47	2.20	1.58
	50 models	0.89	0.87	4.63	2.15	-0.55	4.33	2.08	2.59	1.86
pH_w	Single	0.69	0.82	0.25	0.50	-0.04	0.25	0.50	1.78	1.34
	50 models	0.76	0.86	0.19	0.43	-0.05	0.19	0.43	2.05	1.54
EC	Single	0.58	0.75	0.12	0.34	-0.02	0.12	0.34	1.52	0.27
	50 models	0.70	0.82	0.08	0.28	0.00	0.08	0.28	1.85	0.33

Table 6.4 External validation of EPO models for the Local library

		R^2	LCCC	MSE	RMSE	Bias	MSE_c	$RMSE_c$	RPD	RPIQ
OC	Single	0.47	0.67	0.05	0.22	-0.02	0.05	0.21	1.37	0.84
	50 models	0.67	0.79	0.03	0.17	0.01	0.03	0.17	1.76	1.08
Clay	Single	0.76	0.84	37.18	6.10	0.95	36.27	6.02	2.01	0.98
	50 models	0.79	0.86	31.38	5.60	0.71	30.87	5.56	2.18	1.07
Sand	Single	0.62	0.77	57.58	7.59	0.47	57.35	7.57	1.62	0.52
	50 models	0.79	0.87	31.21	5.59	0.75	30.64	5.54	2.21	0.71
CEC	Single	0.78	0.86	6.67	2.58	-0.25	6.60	2.57	2.16	1.55
	50 models	0.85	0.86	5.03	2.24	-0.21	4.99	2.23	2.48	1.78
pH_w	Single	0.68	0.81	0.26	0.51	-0.02	0.25	0.50	1.77	1.33
	50 models	0.74	0.84	0.21	0.46	-0.03	0.21	0.46	1.95	1.46
EC	Single	0.67	0.81	0.09	0.30	-0.04	0.09	0.30	1.73	0.31
	50 models	0.69	0.80	0.08	0.29	-0.01	0.08	0.29	1.79	0.32

Table 6.5 External validation of PP models for the Global library

		R^2	LCCC	MSE	RMSE	Bias	MSE_c	$RMSE_c$	RPD	RPIQ
OC	Single	0.58	0.75	0.40	0.64	0.02	0.40	0.63	1.49	0.39
	50 models	0.68	0.81	0.29	0.54	0.07	0.29	0.53	1.76	0.46
Clay	Single	0.70	0.83	127.34	11.28	-0.48	127.11	11.27	1.80	1.38
	50 models	0.77	0.86	96.59	9.83	-0.50	96.35	9.82	2.07	1.59
Sand	Single	0.81	0.89	56.98	7.55	1.19	55.55	7.45	2.30	1.35
	50 models	0.83	0.90	50.56	7.11	1.10	49.34	7.02	2.44	1.43
CEC	Single	0.58	0.75	40.42	6.36	-0.71	39.93	6.32	1.54	0.78
	50 models	0.68	0.80	30.39	5.51	-0.13	30.37	5.51	1.78	0.90
pH_w	Single	0.42	0.63	1.11	1.05	0.04	1.11	1.05	1.30	1.51
	50 models	0.50	0.65	0.94	0.97	0.03	0.94	0.97	1.41	1.64
EC	Single	0.42	0.65	0.15	0.39	-0.01	0.15	0.39	1.15	0.24
	50 models	0.51	0.71	0.11	0.32	0.00	0.11	0.32	1.37	0.28

Table 6.6 External validation of EPO models for the Global library

		R^2	LCCC	MSE	RMSE	Bias	MSE_c	$RMSE_c$	RPD	RPIQ
OC	Single	0.65	0.80	0.32	0.57	0.01	0.32	0.57	1.66	0.44
	50 models	0.70	0.82	0.27	0.52	0.08	0.27	0.52	1.81	0.48
Clay	Single	0.67	0.81	139.15	11.80	-0.74	138.61	11.77	1.73	1.32
	50 models	0.76	0.85	99.60	9.98	-0.02	99.60	9.98	2.04	1.56
Sand	Single	0.75	0.85	77.16	8.78	1.05	76.05	8.72	1.97	1.16
	50 models	0.83	0.89	53.40	7.31	1.19	51.98	7.21	2.37	1.39
CEC	Single	0.62	0.78	48.46	6.96	0.61	48.09	6.94	1.55	1.42
	50 models	0.71	0.83	33.49	5.79	0.30	33.40	5.78	1.87	1.71
pH_w	Single	0.42	0.63	1.14	1.07	0.04	1.14	1.07	1.29	1.50
	50 models	0.52	0.67	0.91	0.95	0.04	0.90	0.95	1.44	1.68
EC	Single	0.44	0.66	0.13	0.35	-0.02	0.13	0.35	1.25	0.26
	50 models	0.54	0.72	0.09	0.31	-0.01	0.09	0.31	1.45	0.30

6.4.2 Application in the field

6.4.2.1 Applying the Local models to Hillston sites

The RMSE values of air-dry and ground spectra were comparable to those observed in the validation set, indicating that the two sites were representative of the Local library. Two exceptions were the maximum clay value of two profiles was slightly larger than observed in the Local library, and the minimum OC value was slighlty less (Table 6.1).

Air-dry and ground samples scanned in the laboratory produced the best validation metrics. The EPO transformation improved results from *in situ* scans compared to PP spectra; however even EPO transformed predictions were not as accurate as those scanned in the laboratory. This reinforces investigations showing that EPO can reduce the negative influence of soil moisture on predictions (Minasny *et al*., 2011; Ge *et al*., 2014). The improved performance of laboratory scanned soil may be due to the removal of moisture, but also due to homogenisation of the sample and the benefit of scanning the exact sample that is transferred for laboratory analysis.

Table 6.7 Hillston models applied to Hillston sites

		R^2	LCCC	MSE	RMSE	Bias	MSE_c	$RMSE_c$	RPD	RPIQ
OC	Laboratory	0.97	0.76	0.02	0.13	0.12	0.00	0.04	1.96	1.76
	In situ EPO	0.86	0.54	0.07	0.26	0.25	0.01	0.08	0.99	0.89
	In situ PP	0.82	0.55	0.07	0.26	0.24	0.01	0.10	0.98	0.88
Clay	Laboratory	0.96	0.80	35.94	6.00	-5.10	9.92	3.15	2.60	0.69
	In situ EPO	0.85	0.71	75.79	8.71	-6.52	33.26	5.77	1.79	0.47
	In situ PP	0.63	0.64	116.26	10.78	-6.93	68.22	8.26	1.44	0.38
Sand	Laboratory	0.79	0.79	50.46	7.10	0.37	50.32	7.09	2.53	1.12
	In situ EPO	0.67	0.71	87.73	9.37	2.44	81.75	9.04	1.92	0.85
	In situ PP	0.39	0.58	174.70	13.22	3.94	159.21	12.62	1.36	0.60
CEC	Laboratory	0.89	0.80	3.42	1.85	1.28	1.79	1.34	2.38	0.27
	In situ EPO	0.84	0.77	4.16	2.04	1.36	2.30	1.52	2.16	0.25
	In situ PP	0.44	0.56	16.44	4.05	2.36	10.86	3.30	1.09	0.12
pH_w	Laboratory	0.86	0.77	0.23	0.48	0.19	0.19	0.44	2.47	1.67
	In situ EPO	0.94	0.72	0.30	0.55	-0.19	0.26	0.51	2.18	1.48
	In situ PP	0.69	0.73	0.34	0.58	-0.08	0.33	0.58	2.04	1.38
EC	Laboratory	0.60	0.42	0.05	0.23	0.19	0.02	0.14	0.59	0.15
	In situ EPO	0.83	0.53	0.04	0.21	0.13	0.03	0.17	0.65	0.16
	In situ PP	0.90	0.35	0.18	0.42	0.25	0.11	0.33	0.33	0.08

Visual observation of the bagged estimate values and prediction intervals demonstrate that SPEC-SINFERS worked well *in situ* with the local calibrated models. Bagged estimates aligned well with horizon-based laboratory values and the 95% prediction interval gave a good indication of the accuracy of the estimate (Figs 6.6 and 6.7).

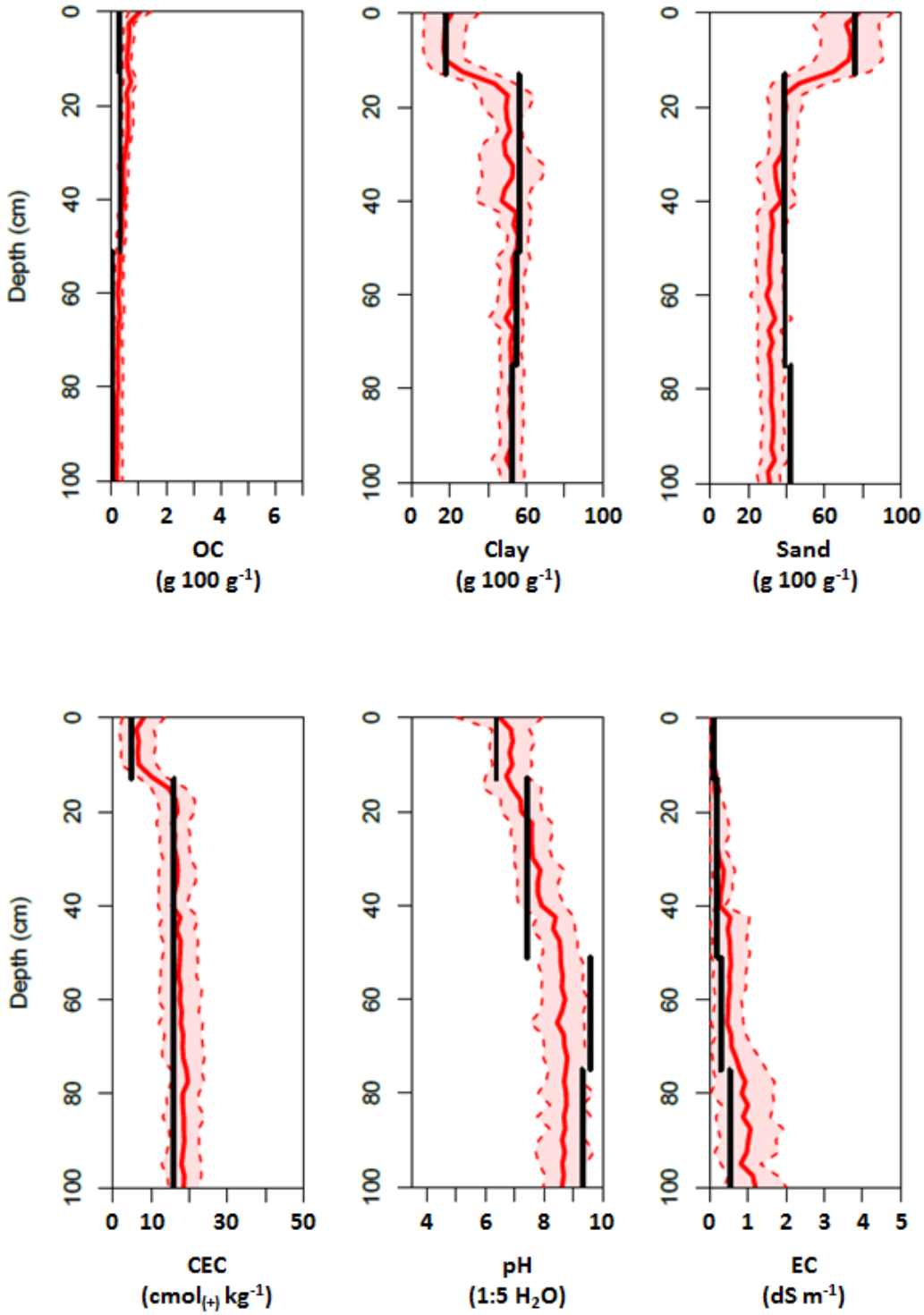


Figure 6.6 Spectrally derived properties for Site 7. Solid red lines indicate mean bagged estimates at 2.5 cm increments; black lines indicate horizon-based laboratory values; broken red lines indicate the prediction interval; the area inside the prediction interval is shaded. OC – organic carbon, CEC – cation exchange capacity, EC – electrical conductivity.

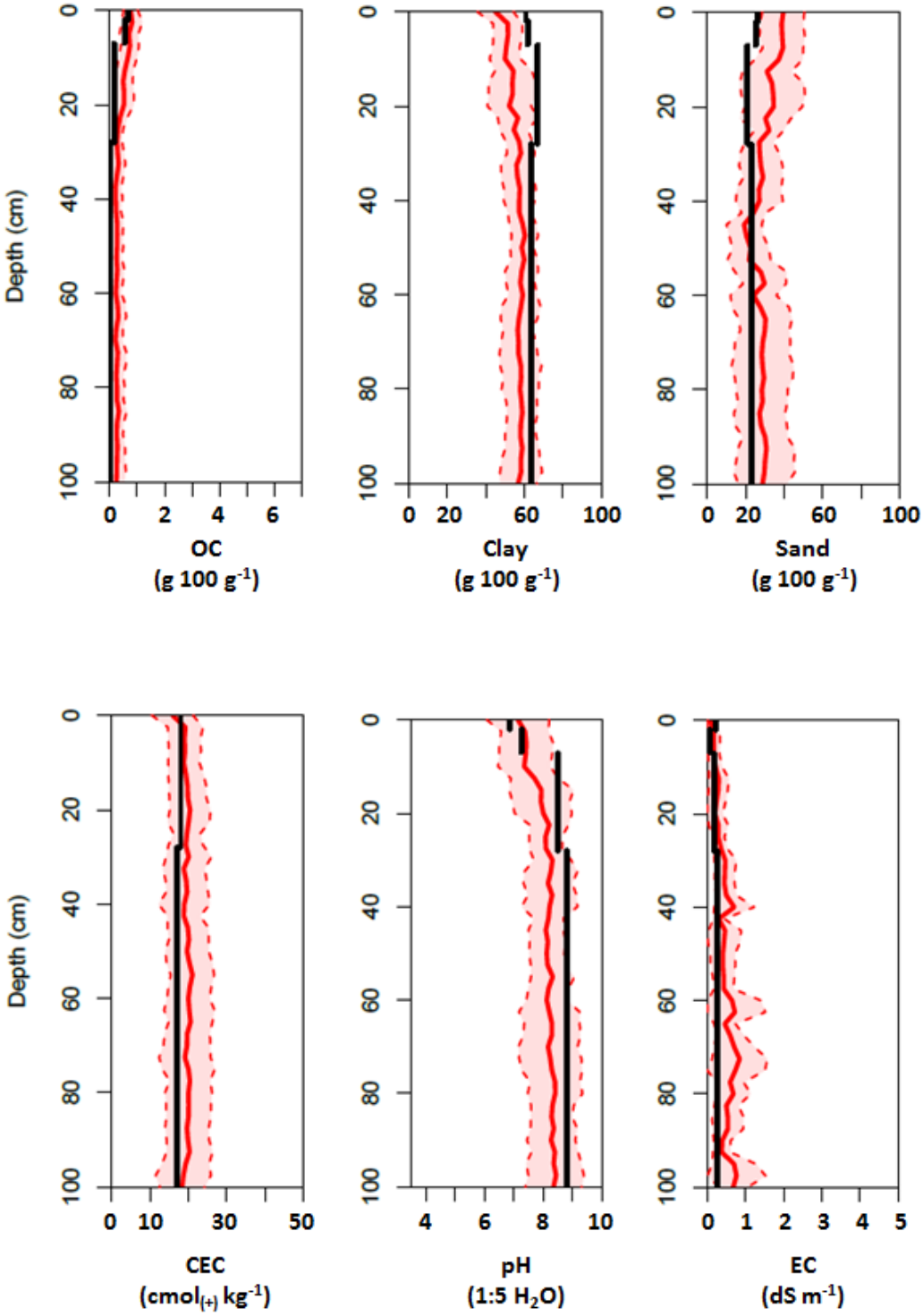


Figure 6.7 Spectrally derived properties for Site 8. Solid red lines indicate mean bagged estimates at 2.5 cm increments; black lines indicate horizon-based laboratory values; broken red lines indicate the prediction interval; the area inside the prediction interval is shaded. OC – organic carbon, CEC – cation exchange capacity, EC – electrical conductivity.

6.4.2.2 SINFERS results for Hillston pits

The SPEC-SIFNERS approach allowed the prediction of 29 properties *in situ*, in near real-time (Fig. 6.8). More properties could be estimated if appropriate PTFs were supplied.

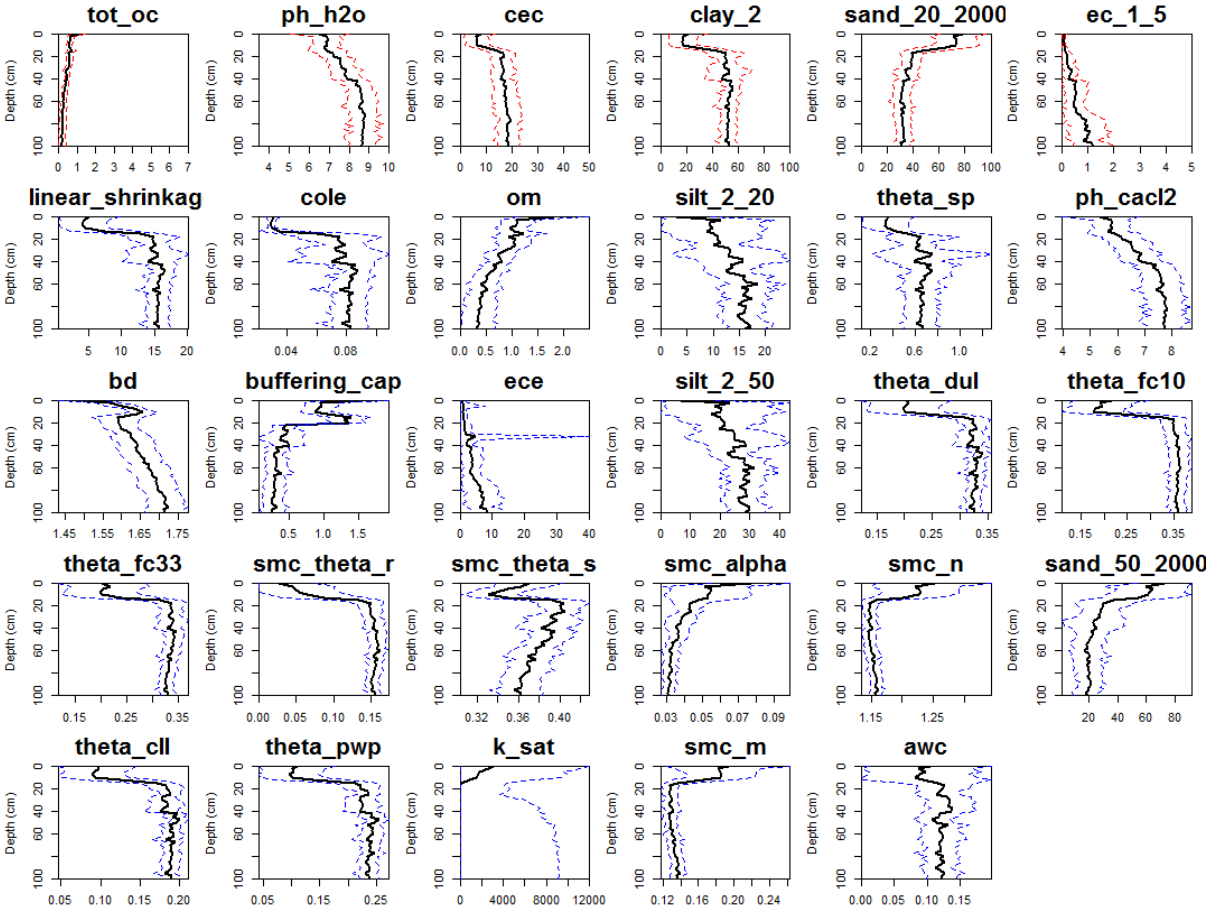


Figure 6.8 All SPEC-SINFERS results for Site 7. Black lines indicate bagging estimate means; dotted lines indicate the prediction interval for spectrally derived properties (red) and SINFERS derived properties (blue).

As SPEC-SINFERS relies on using estimates of soil properties as inputs to estimate further soils properties, the correct propagation of both input and model uncertainties is paramount. Each estimated value should be accompanied with a measure of its uncertainty (McBratney *et al*., 2006). Bootstrap aggregating techniques provide a useful method to provide a bagged estimate and prediction interval from a spectral library. Meanwhile, the Monte Carlo method to can be used to draw realistic simulations of input variables from a probability distribution to provide a measure of uncertainty from deterministic PTFs.

Bulk density was predicted with an RMSE of 0.11 g cm^{-3} *in situ* (Table 6.8, Fig. 6.9). This is comparable with the standard deviation of paired BD samples of 0.10 g cm^{-3}. As the standard deviation and the RMSE are similar metrics for unbiased predictions, this indicates that SPEC-SINFERS predicted BD was accurately predicted. Askari et al., (2015) were able to predict BD directly from VisNIR spectra with an RMSE of 0.9 g cm^{-3}, although for this study five subplots were established at each site and the same site may be present in the calibration and validation sets. This shows the potential of using SPEC-SINFERS when the production of a new calibration set is restricted. Such an example is given in the case of the Local library. An extensive spectral library is available for many properties, yet bulk density was not measured. To create a library including bulk density would require resampling all sites. This is an example where the SPEC-SINFERS approach is useful.

Table 6.8 Local models and SPEC-SINFERS applied to Hillston sites

		R^2	LCCC	MSE	RMSE	Bias	MSE_c	$RMSE_c$	RPD	RPIQ
BD	Laboratory	0.83	0.73	0.00	0.07	0.01	0.00	0.07	2.36	0.79
	In situ EPO	0.39	0.47	0.01	0.11	-0.04	0.01	0.10	1.22	1.22
	In situ PP	0.33	0.42	0.01	0.11	-0.04	0.01	0.11	1.17	1.18

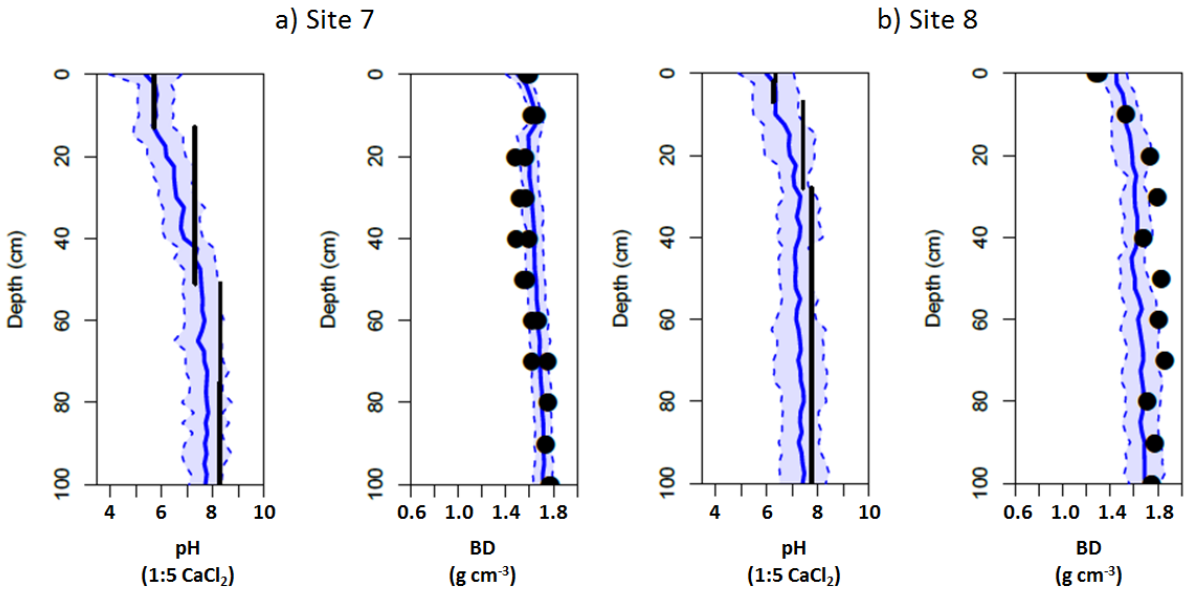


Figure 6.9 SPEC-SINFERS and observed values of pH (1:5 $CaCl_2$) and bulk density for a) Red Chromosol (Site 7), and b) Grey Vertosol (Site 8). Solid blue lines indicate mean bagged estimates at 2.5 cm increments; black lines indicate horizon-based laboratory values; filled black circles represent bulk density measurements at 10 cm depth increments; broken blue lines indicate the prediction interval; the area inside the prediction interval is shaded.

6.4.2.3 Enrichment of profile characterisation

The prediction of such a large number of properties facilitates an enriched analysis of a profile. Combining the parameters estimated for the soil moisture characteristic enables reconstruction of the curve for each sample (Fig. 6.10a). Combining information held in the soil moisture characteristic with BD measures can give a greater understanding of pore space relations in a sample or an entire profile (Fig 6.10b). For this profile SPEC-SINFERS is indicating that the topsoil does not hold much plant available water. Meanwhile, the subsoil does not contain 10% porosity at field capacity, indicating that root respiration may be impaired.

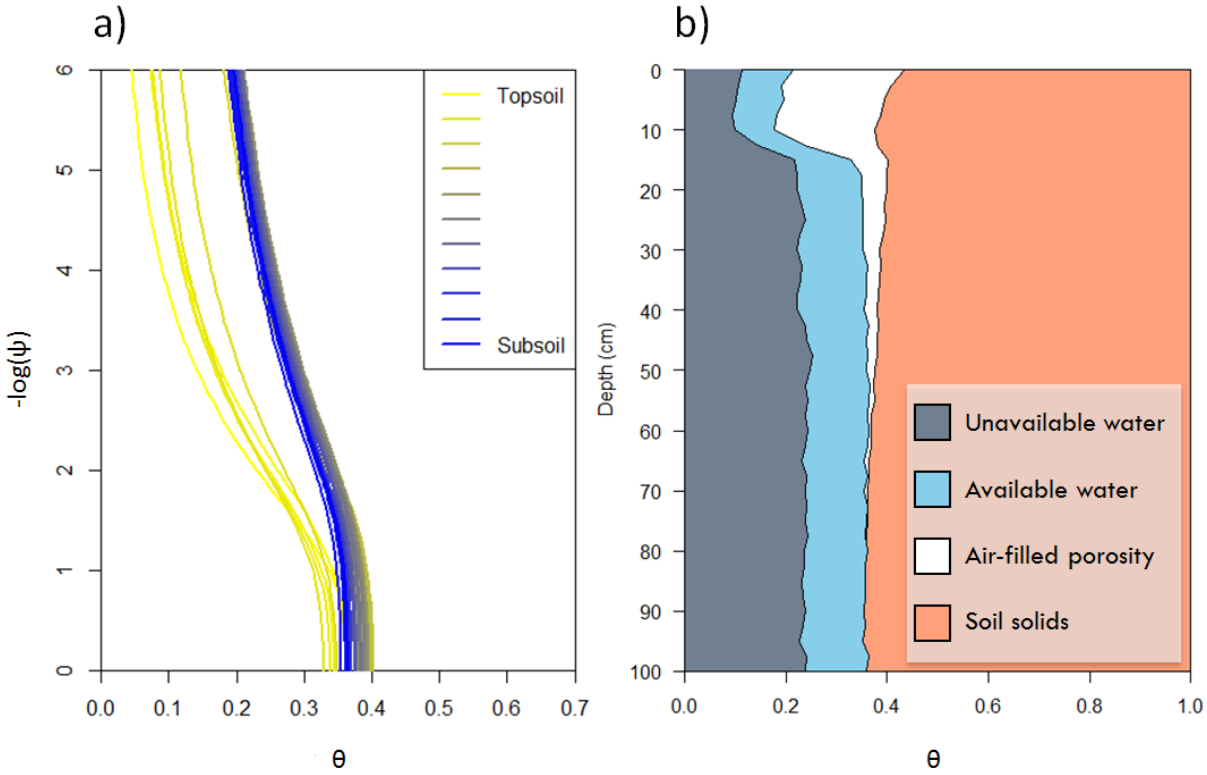


Figure 6.10 Fine-scale investigation of a) the soil moisture characteristic, and b) plant available water, of a Red Chromosol (Site 7) derived from SPEC-SINFERS results.

6.4.2.4 Applying the Global models at Hillston

Supplementing the Local library with the four other available spectral libraries in the Global library generally produced poorer results, than the Local library alone. Increases in RMSE were seen for OC clay, CEC and pH (Table 6.9). In the case of CEC and pH the increases in RMSE represented more than 100% of the Locally calibrated model. Improvements were seen for sand and EC.

Table 6.9 Relative changed in RMSE using Global models compared to Hillston only models.

	OC	Clay	Sand	CEC	pH	EC
Change in RMSE (%)	7.7	29.4	-5.0	153.4	132.7	-38.1

These findings reflect those of Wetterlind and Stenberg (2010), who found that a local dataset, even with as few as 25 samples, produced better results than a larger regional dataset, and that spiking the regional dataset with the 25 local samples only improved outcomes sometimes. While spiking a Global library with local samples may improve outcomes (Guerrero *et al*., 2010), it appears if a sufficiently large local library is available, it is better to use it directly.

Sankey *et al*., (2008) also found mixed results when observing model improvements between local and global calibration sets for three diverse study sites. A local + global model produced the best results for OC at all sites. However, purely local calibration (n=210) produced the best results at one site for the prediction of clay content. Global models alone were never observed to achieve the best validation statistics. With an ever increasing number and sharing of spectral libraries across the globe, more research is needed to determine when it is useful to use external spectral libraries, and when they in fact result in poorer outcomes.

6.4.2.5 Applying the Hillston models to all sites

Applying the Hillston models to all fifteen pits produced poor results (Table 6.10). Observed RMSE values were approximately double those of the Hillston pits. Reinforcing the notion that models can be extremely unstable when applied outside of their calibration range, as has also previously been observed with mid-infrared models (Minasny *et al*., 2009). There was insufficient representation of soil diversity in the Hillston library to extrapolate it to all sites.

Table 6.10 Hillston models applied to all sites.

		R^2	LCCC	MSE	RMSE	Bias	MSE_c	$RMSE_c$	RPD	RPIQ
OC	Laboratory	0.39	0.39	0.82	0.91	-0.14	0.80	0.90	1.21	0.26
	In situ EPO	0.18	0.28	0.97	0.98	0.00	0.97	0.98	1.12	0.24
	In situ PP	0.34	0.36	0.85	0.92	-0.10	0.84	0.92	1.19	0.26
Clay	Laboratory	0.70	0.66	231.62	15.22	-8.53	158.79	12.60	1.42	1.35
	In situ EPO	0.43	0.50	349.06	18.68	-9.37	261.23	16.16	1.16	1.10
	In situ PP	0.46	0.57	276.53	16.63	-5.22	249.26	15.79	1.30	1.23
Sand	Laboratory	0.69	0.78	197.45	14.05	4.93	173.15	13.16	1.70	1.19
	In situ EPO	0.32	0.47	476.06	21.82	9.72	381.65	19.54	1.09	0.77
	In situ PP	0.41	0.61	338.29	18.39	1.97	334.43	18.29	1.30	0.91
CEC	Laboratory	0.65	0.76	22.30	4.72	-0.73	21.77	4.67	1.69	1.09
	In situ EPO	0.43	0.58	35.74	5.98	-0.63	35.34	5.95	1.34	0.86
	In situ PP	0.21	0.42	64.07	8.00	2.88	55.78	7.47	1.00	0.64
pH_w	Laboratory	0.40	0.60	1.29	1.14	0.36	1.17	1.08	1.13	0.67
	In situ EPO	0.50	0.67	0.88	0.94	0.26	0.81	0.90	1.37	0.81
	In situ PP	0.08	0.18	3.32	1.82	1.24	1.79	1.34	0.70	0.42
EC	Laboratory	0.73	0.64	0.26	0.51	0.35	0.14	0.38	0.85	0.08
	In situ EPO	0.48	0.61	0.17	0.41	0.23	0.12	0.34	1.06	0.10
	In situ PP	0.31	0.33	0.73	0.86	0.59	0.38	0.62	0.51	0.05

6.4.2.6 Applying the Global models to all sites

Validation statistics were improved when using the Global library compared to the Hillston library (Table 6.11). However, the models were not effective enough to describe the great diversity of soils measured. The maximum value of clay seen in the fifteen soil profiles

exceeded the maximum clay of the calibration set, while all properties except pH and EC had some observations outside the range of the validation set (Table 6.2). This represents one of the difficulties in applying calibrations outside of their geographical domain. Ongoing maintentance and improvement of spectral libraries is required to reduce any negative influences from this.

Table 6.11 Global models applied to all sites.

		R^2	LCCC	MSE	RMSE	Bias	MSE_c	$RMSE_c$	RPD	RPIQ
OC	Laboratory	0.68	0.77	0.50	0.71	0.20	0.46	0.68	1.71	0.37
	In situ EPO	0.60	0.70	0.70	0.84	0.37	0.57	0.75	1.44	0.31
	In situ PP	0.55	0.61	0.72	0.85	0.11	0.71	0.84	1.43	0.31
Clay	Laboratory	0.57	0.64	213.66	14.62	1.09	212.47	14.58	1.48	1.40
	In situ EPO	0.53	0.62	225.55	15.02	-0.04	225.55	15.02	1.44	1.37
	In situ PP	0.09	0.25	428.28	20.69	2.53	421.87	20.54	1.04	0.99
Sand	Laboratory	0.76	0.79	163.34	12.78	-2.59	156.62	12.51	1.87	1.31
	In situ EPO	0.52	0.66	268.23	16.38	0.64	267.82	16.37	1.46	1.02
	In situ PP	0.28	0.41	514.47	22.68	-10.61	401.87	20.05	1.05	0.74
CEC	Laboratory	0.81	0.76	47.56	6.90	4.56	26.75	5.17	1.16	0.75
	In situ EPO	0.76	0.78	29.55	5.44	3.73	15.62	3.95	1.47	0.95
	In situ PP	0.78	0.69	35.41	5.95	4.95	10.87	3.30	1.15	1.44
pH_w	Laboratory	0.46	0.53	1.16	1.08	0.54	0.87	0.93	1.19	0.70
	In situ EPO	0.44	0.47	1.51	1.23	0.78	0.90	0.95	1.04	0.62
	In situ PP	0.24	0.29	2.04	1.43	0.91	1.21	1.10	0.90	0.53
EC	Laboratory	0.92	0.89	0.05	0.22	0.12	0.03	0.18	2.00	0.19
	In situ EPO	0.70	0.75	0.07	0.26	0.07	0.06	0.25	1.70	0.16
	In situ PP	0.45	0.55	0.21	0.45	0.29	0.13	0.35	0.96	0.09

Geographic location is a good indicator of model performance. Visual examination of results for individual sites indicated that some sites outside of the geographic domain of the libraries also performed well. This raises the question of whether it is possible to predict the accuracy of a prediction based on intrinsic information held within the spectra of the calibration set and the unknown spectrum.

6.4.3 Predicting model performance from intrinsic spectral information

Attempts to discriminate spectra based on inherent spectral information, such as whether new spectra lie within the convex hull of the library, proved ineffective with the restriction increasing the RMSE just as often as it decreased it. Other attempts to discriminate, by applying the convex hull analysis over four dimensions, reducing the size of the convex hull until it contained 90% of the original library, or applying a Euclidean nearest neighbour distance metric also proved ineffective. Based on direct analysis of a received unknown spectrum, it could not be determined if the models would be effective. Differences in spectra are too subtle to be discriminated using this approach.

A more promising result was observed when considering the size of the prediction interval produced for a scan. Profiles that were not well-represented in the calibration library generally had larger prediction intervals, such as the Ferrosol represented at Site 14 (Fig. 6.9). Prediction intervals for Site 14 generally covered more than 50% of the prediction domain. This is a result of the bagging approach, which gives many realisations of a model and this produces instability in the prediction of spectra not well-represented in the calibration library, resulting in larger prediction intervals.

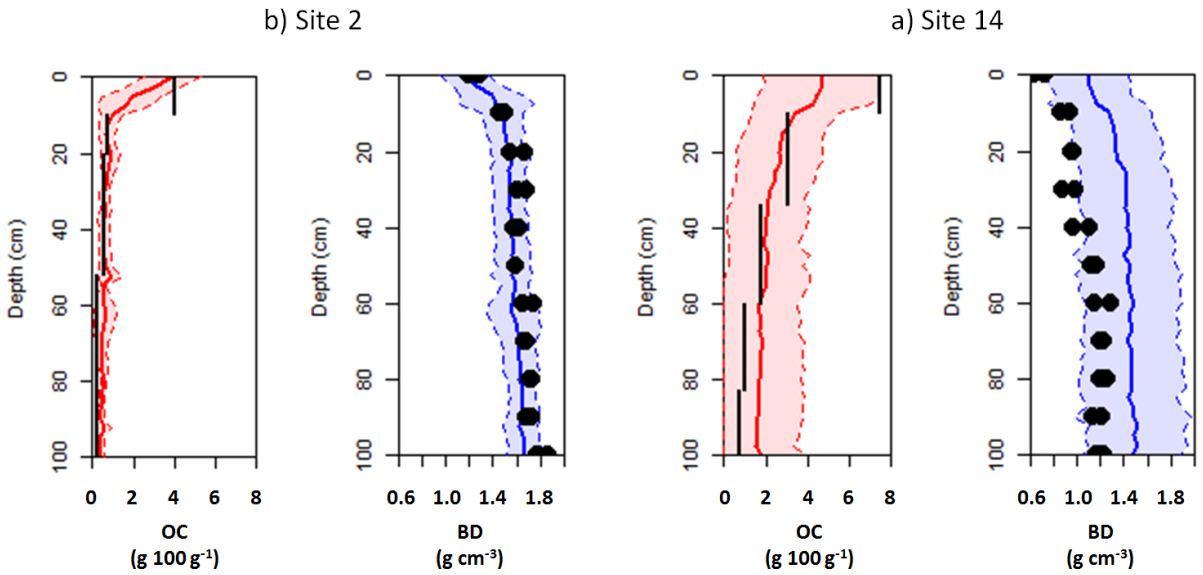


Figure 6.11 Comparison of organic carbon and bulk density estimates between: a) Red Sodosol (Site 2), a relatively well represented profile in the calibration library; and b) Red Ferrosol (Site 14), a profile not a well-represented in the calibration set.

By plotting the quantiles of prediction interval width, it was observed that the residual value generally increases with the size of the prediction interval (Fig. 6.12). An important finding was that prediction interval width was not determined by the magnitude of the bagging prediction, otherwise its utility would be reduced. Observing the clay content prediction of Site 7 (Fig. 6.6b), a wider prediction interval occurs in the topsoil, which has a smaller clay content. The prediction interval remains wide for the Bt horizon before reducing in the Btk1 and Btk2 horizons. This demonstrates that the width of the prediction interval is affecting by more than the magnitude of the prediction.

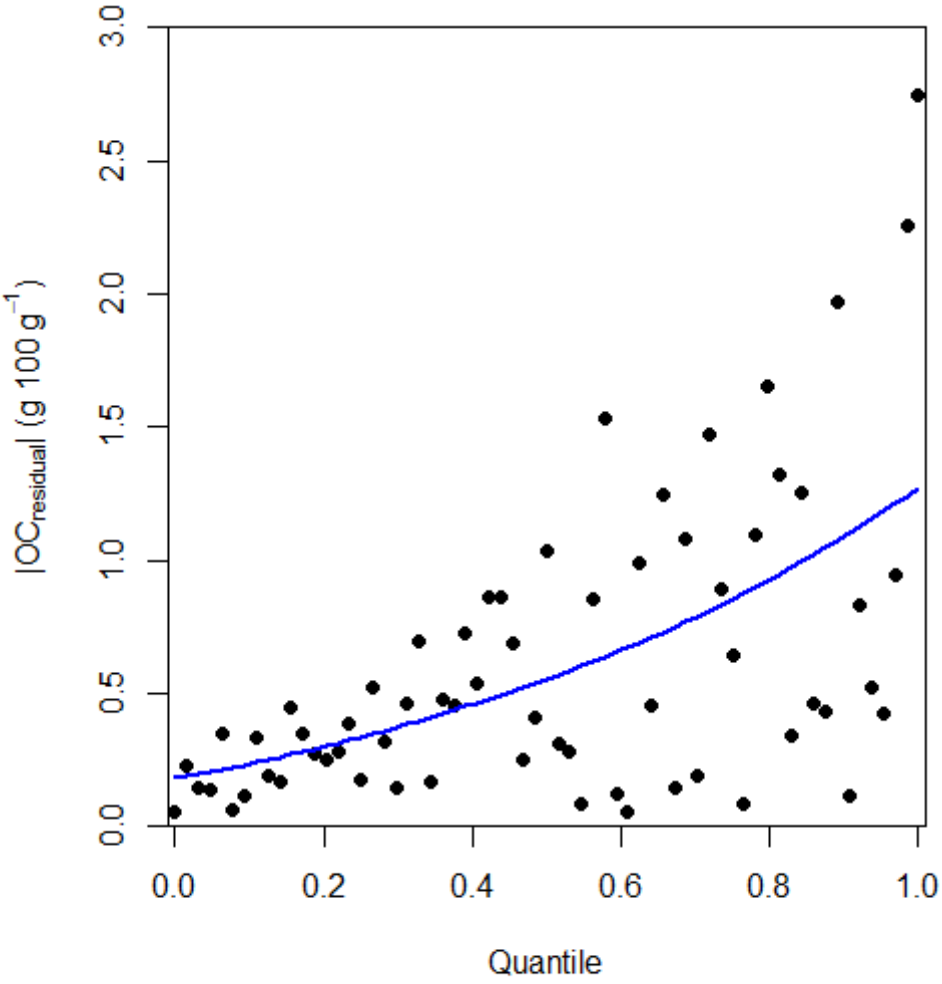


Figure 6.12 Plot of quantiles of the prediction interval of the 50 cubist model estimates and the organic carbon residual for *in situ*, EPO transformed spectra of the Global library. A locally-weighted nonlinear trend line is plotted in blue.

6.5 Conclusion

- The functionality of SPEC-SINFERS in the field was demonstrated.
- Application of bagging improved the Cubist models and allowed the construction of realistic prediction intervals.
- The EPO transformation improved field estimates, although RMSE values remained larger than those analysed in air-dried, ground condition in the laboratory.

- SPEC-SINFERS BD was predicted with an RMSE of 0.11 g cm^{-3} *in situ*, following EPO transformation and using a local calibration library. This was very close to the standard deviation between paired BD observations of 0.10 g cm^{-3}.
- The Global library generally produced poorer results, compared to the Local library for the Hillston sites. This suggests that a local calibration is preferable when a library of sufficient size and representation of local soils is available.
- Pre-screening spectra to determine if they are represented in the calibration dataset by using convex hulls of the first two principal components of the calibration set produced disappointing results. A better indication was obtained by observing the size of the prediction interval. More research is needed to identify the accuracy of a prediction based on intrinsic information held within the spectra of the calibration library and an unknown spectrum.

6.6 References

Askari, M. S., Cui, J., O'Rourke, S. M., Holden, N. M. (2015). Evaluation of soil structural quality using VIS–NIR spectra. Soil and Tillage Research 146(Part A), 108-117.

Bouma, J. (1989). Using soil survey data for quantitative land evaluation. Advances in soil science 9(1989), 177-213.

Breiman, L. (1996). Bagging predictors. Machine-learning 24(2), 123-140.

Briggs, L. J., and J. W. McLane, 1907. The moisture equivalents of soils, Technical Report 45, U.S. Department of Agriculture, Bureau of Soils, Washington D.C.

Coleman, K., Jenkinson, D., 1996. RothC-26.3-A Model for the turnover of carbon in soil, Evaluation of soil organic matter models. Springer, pp. 237-246.

Cornelis, W. M., Ronsyn, J., Van Meirvenne, M., Hartmann, R. (2001). Evaluation of pedotransfer functions for predicting the soil moisture retention curve. Soil Science Society of America Journal 65(3), 638-648.

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.

Filippi P., Cattle S.R., Bishop T.F.A., Pringle M.J., Odeh I.O.A. (2018). Digital soil monitoring of top- and sub-soil pH with bivariate linear mixedmodels. Geoderma, 322, 149-162.

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.

Geeves, G.W., Cresswell, H.P., Murphy, B.W., Gessler, P.E., Chartres, C.J., Little, I.P., Bowman, G.M. (1995). The Physical, Chemical and Morphological Properties of Soils in the Wheat-belt of Southern NSW and Northern Victoria. NSW Department of Conservation and Land Management / CSIRO Division of Soils Occasional Report, CSIRO, Australia.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. Science 327(5967), 812-818.

Guber, A., Pachepsky, Y. A., Van Genuchten, M. T., Simunek, J., Jacques, D., Nemes, A., Nicholson, T., Cady, R. (2009). Multimodel simulation of water flow in a field soil using pedotransfer functions. Vadose Zone Journal 8(1), 1-10.

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.

Hartemink, A. E., Minasny, B. (2014). Towards digital soil morphometrics. Geoderma 230, 305-317.

Kreuter, U. P., Harris, H. G., Matlock, M. D., Lacey, R. E. (2001). Change in ecosystem service values in the San Antonio area, Texas. Ecological Economics 39(3), 333-346.

Lal, R. (2004). Soil carbon sequestration to mitigate climate change. Geoderma 123(1), 1-22.

Lamorski, K., Pachepsky, Y., Sławiński, C., Walczak, R. (2008). Using support vector machines to develop pedotransfer functions for water retention of soils in Poland. Soil Science Society of America Journal 72(5), 1243-1247.

Malone, B. P., de Gruijter, J. J., McBratney, A. B., Minasny, B., Brus, D. J. (2011). Using additional criteria for measuring the quality of predictions and their uncertainties in a digital soil mapping framework. Soil Science Society of America Journal 75(3), 1032-1043.

McBratney, A. B., Minasny, B., Cattle, S. R., Vervoort, R. W. (2002). From pedotransfer functions to soil inference systems. Geoderma 109(1), 41-73.

McBratney, A. B., Minasny, B., Viscarra Rossel, R. (2006). Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. Geoderma 136(1–2), 272-278.

McBratney, A., Field, D. J., Koch, A. (2014). The dimensions of soil security. Geoderma 213, 203-213.

Minasny, B., McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences 32(9), 1378-1388.

Minasny, B., Tranter, G., McBratney, A. B., Brough, D. M., Murphy, B. W. (2009). Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. Geoderma 153(1), 155-162.

Minasny, B., McBratney, A. B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L., Joalland, S. (2011). Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma 167–168(0), 118-124.

Minasny, B., McBratney, A., Brough, D., Jacquier, D. (2011). Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. European Journal of Soil Science 62(5), 728-732.

Morgan, R. P. C., Quinton, J. N., Smith, R. E., Govers, G., Poesen, J. W. A., Auerswald, K., Chisci, G., Torri, D., Styczen, M. E. (1998). The European Soil Erosion Model (EUROSEM): a dynamic approach for predicting sediment transport from fields and small catchments. Earth Surface Processes and Landforms 23(6), 527-544.

Onus A., Cattle S., Odeh I. (2003). How do Lachlan Valley cotton soils compare to northern NSW? Australian Cottongrower 24, 28–30.

Padarian, J., Minasny, B., McBratney, A. (2012). Using genetic programming to transform from Australian to USDA/FAO soil particle-size classification system. Soil Research 50(6), 443-446.

Pino, V. (2016). Soil Microbial Diversity Across Different Agroecological Zones in New South Wales, The University of Sydney.

Roudier, P. (2011) clhs: An R package for conditioned Latin hypercube sampling.

Sankey, J. B., Brown, D. J., Bernard, M. L., Lawrence, R. L. (2008). Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. Geoderma 148(2), 149-158.

Schaap, M. G., Leij, F. J., Van Genuchten, M. T. (2001). ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. Journal of Hydrology 251(3), 163-176.

Tranter, G., Minasny, B., McBratney, A. B., Rossel, R. A. V., Murphy, B. W. (2008). Comparing Spectral Soil Inference Systems and Mid-Infrared Spectroscopic Predictions of Soil Moisture Retention. Soil Science Society of America Journal 72(5), 1394-1400.

Tranter, G., Minasny, B., McBratney, A.B., Murphy, B., McKenzie, N.J., Grundy, M. and Brough, D. (2007). Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use and Management, 23(4), pp.437-443.

Twarakavi, N. K., Šimůnek, J., Schaap, M. (2009). Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. Soil Science Society of America Journal 73(5), 1443-1452.

Viscarra Rossel, R. A., McBratney, A. B., Minasny, B. (2010). Proximal soil sensing. Springer Science & Business Media.

Wetterlind, J., Stenberg, B. (2010). Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. European Journal of Soil Science 61(6), 823-843.

7 CONCLUSIONS

This thesis investigated *in situ* characterisation of soil using proximal sensors. In the four research chapters, methods to remove the negative effects of soil moisture were assessed, and novel sampling techniques were proposed. In total, 37 soil properties were estimated with visible near-infrared diffuse reflectance (VisNIR) spectra and portable X-ray fluorescence (pXRF) reported elements. This included characterising soil mineral composition in terms of eight minerals in Chapter 5, and estimation of 29 soil properties with SPEC-SINFERS in Chapter 6. All results are available in the field. Limitations and further improvements to the approaches, and potential future research directions are outlined below.

7.1 Research summary

Chapter 3 demonstrated the negative effects of soil moisture on VisNIR spectra. The moisture effect was so great that spectra from air-dry subsoil samples were more similar to field condition topsoil samples. This moisture effect could be reduced using external parameter orthogonalisation (EPO). Clustering of EPO transformed spectra produced classes that resembled observed soil horizons. Allocation of samples to classes was relatively stable under variable moisture condition following EPO transformation. Although this investigation was performed on a soil profile, the same methodology could be extended to mapping soil units in two or three dimensions.

Chapter 4 proposed the use of VisNIR spectra as a proxy for soil properties to characterise the vertical and lateral components of soil profile variability. A sampling increment of 4.1 cm in the vertical dimension was found to experience equivalent variation in soil properties as an entire 1 m lateral cross-section. As a general case, sampling at this resolution, or some multiple thereof, will provide efficient characterisation of profile variability.

Chapter 5 presented a data fusion approach to characterise soil mineral composition *in situ*. This was the first comprehensive attempt to estimate soil mineralogy *in situ*. Results were promising, with the dominant mineral of a horizon predicted with 72% accuracy. The fusion model increased the ordinal prediction of phyllosilicates from 34% to 63% accuracy. Future research is required to quantify smectite. Clark *et al.* (1990) describe the elements involved in isomorphic substitution inducing changes in VisNIR spectra. This may provide a solution; if the chemical structure of smectite can be estimated using VisNIR, then the elemental requirements for the mass balance may be adjusted accordingly. Otherwise, local calibration will be required. The significant correlations between mineral abundance and pXRF reported values demonstrated the strength of combining VisNIR and pXRF, through their combined capacity to provide both molecular and elemental characterisation. For comprehensive mineral estimation, both are required. Whether this takes the form of a mass balance, is in the case presented, or machine learning directly from the two data streams, or some derivative thereof.

Chapter 6 demonstrated a field functional SPEC-SINFERS system. A single VisNIR spectrum was used to predict 29 soil properties, and their associated uncertainties, in near real-time. When locally calibrated models were used, effective characterisation was achieved, as demonstrated by small RMSE values. More work is needed to reduce RMSE values of field reported spectra, to those of air-dry and ground samples. An interesting point raised in this chapter was the ability to assess model performance based on intrinsic information held within an unknown spectrum, irrespective of whether the sample was derived from the geographic domain of the calibration library. Investigations based on constructing a convex hull around the first two principal components of the calibration library provided mixed

results. The width of prediction intervals offered some indication of model performance. However, further research is needed to resolve this issue entirely.

7.2 Further techniques

7.2.1 Removing the moisture effect entirely

Observation of transformed and untransformed spectra under field moist and air-dry condition, demonstrated that EPO was effective at reducing the negative effects of sample moisture. Prediction outcomes were also improved for SPEC-SINFERS predictions by utilising EPO transformation over using calibrations based on air-dry, ground models *in situ*. Although both were constructed from air-dry and ground samples. Future research should be directed at removing it completely. Direct standardisation also improved results. More research is required on both methods and other approaches to further reduce the negative moisture effect.

7.2.2 Prediction of model performance

Good results were obtained when using local calibration models. Does this mean spectral libraries must be constructed for each region? Prediction of model performance based on the presence of an unknown spectrum within the spectral domain of the calibration library, defined as the convex hull of the first two principal components, did not improve model outcomes. Perhaps the convex hull could be extended into further principal component dimensions. Presence within the geographical domain of the calibration library was the best indicator of model performance, although width of the prediction interval also gave a good indication. More work is required, to determine when and where models will work well and limit the need for further calibration libraries.

7.2.3 Pre-screening input based on pXRF or VisNIR

Araújo *et al*., (2014) found that clustering spectra and constructing class based models improved validation statistics for clay and OM prediction, when using a large library of 7,172

spectra from Brazil. RMSE values of 8.9% and 0.48% were observed, representing a decrease in RMSE values of 21 and 15% respectively. The authors observed that the classes contained spectra with more uniform mineralogy, regardless of geographical origin. Undoubtedly a similar improvement would be achieved in Australia, where mineralogy is more diverse than the oxide and kaolinite dominated tropical soils. Soils may be split into classes based on oxide dominance, carbonate dominance, or phyllosilicate speciation. Elemental composition from pXRF could also be used to cluster data. Other authors have suggested identifying the most similar spectra from within a library, and constructing individual models for each unknown spectrum (Cheng-Wen *et al*., 2001). This could be facilitated using global VisNIR spectral libraries (e.g. Viscarra Rossel *et al*., 2016), or a global calibration repository of soil data covering the entire electromagnetic spectrum (Hartemink and Minasny, 2014).

7.2.4 Spatial disaggregation

Historically, for efficiency purposes, soil samples have been grouped into horizons or depth intervals, air-dried and ground before being analysed. This results in a homogenisation of spatial variability. We can now analyse soils *in situ* and capture this vertical spatial variability. Spatially disaggregating this variability can allow us to better understand soil forming processes and evaluate a soil's edaphic, environmental or geotechnical engineering potential.

7.3 Missing devices

As defined in Chapter 1, the holy grail of soil sensing would be a non-invasive, non-destructive sensor that could estimate all properties of interest from above the soil surface. Some insights are provided from above the soil surface by electromagnetic induction, γ-radiometrics, and other less commonly used devices such as ground-penetrating radar, inelastic neutron scattering and acoustic sounding. However, due to the large number of properties simultaneously affecting the reading, they cannot be diagnostic in themselves. For example, electromagnetic induction readings are affected by salt content and type, soil moisture, clay content and type, mineralogy, depth to bedrock, and temperature. Often these

survey products are able to determine where soils vary at a farm or paddock scale, but not how the soils vary. Until comprehensive, non-invasive sensors exist, invasive soil sampling and soil sensors will be required. New field portable devices should be investigated.

7.3.1 Laser-induced breakdown spectroscopy

Laser-induced breakdown spectroscopy (LIBS) offers an alternative to pXRF for the provision of elemental composition. A number of benefits are expressed. Laser-induced breakdown spectroscopy is able to provide quantitative information of all elements, whereas pXRF is restricted to elements heavy than Na, i.e. $Z \geq 12$. As LIBS operates in the near-infrared region of the electromagnetic spectrum, two further benefits are provided: i) it has the potential to be operated through a fibre optic cable, and may be incorporated into a push probe system (e.g. Theriault, *et al*., 1998); and ii) it does not produce ionising radiation, thus reducing safety concerns for operators. Limitations of LIBS include greater detection limits, small support size (nanograms to picograms), requiring multiple scans for effective characterisation. However, LIBS can also fire multiple shots, and thereby penetrate into a target. It has become the method of choice for space exploration, as it has the capacity to blast through layers of Martian dust (Sallé *et al*., 2005). Would it be possible to identify exchangeable versus structural Na in minerals, and directly observe soil sodicity? Or characterise clay and oxide coatings on sand particles?

7.3.2 Mid-infrared spectroscopy

Mid-infrared spectroscopy is a potential alternative to VisNIR, for molecular characterisation of soil samples. The MIR section of the electromagnetic spectrum, 2500–25,000 nm, contains fundamental peaks of many constituents of interest (e.g. Farmer and Russell, 1964; Janik *et al.*, 1995), in contrast to VisNIR which contains combination and overtones (McCarty *et al*., 2002). Comparative laboratory-based studies have shown that MIR can outperform the Vis and NIR sections of the electromagnetic spectrum for the prediction of OC, pH, lime requirement, CEC, clay, silt and sand contents, P and EC (e.g. McCarty *et al*., 2002; Viscarra Rossel *et al*., 2006). Field portable devices are available (Sorak *et al.*, 2012); however, MIR is more susceptible to variable moisture and surface preparation then VisNIR, limiting field

applications. Given the potential for improved models using MIR, more research is needed to determine if these limitations are applied for all soil types and moisture potentials, or if methods to remove the moisture and sample preparation limitations of field portable MIR could be overcome.

7.3.3 Raman spectroscopy

Raman spectroscopy is another technique that can offer insight into molecular characterisation, especially for mineral constituents. Raman spectroscopy provides a number of absorption peaks related to feldspars and Fe-oxides, as well as quartz and common igneous minerals such as olivine and pyroxene (Freeman *et al.*, 2008). Many of these minerals are particularly difficult to characterise with VisNIR. Raman spectroscopy has also been used to analyse organic components (Haumaier and Zech, 1995) and soil contamination (Frost and Kloprogge, 2003). Combination of Raman spectroscopy and LIBS has also been suggested for improved predictions (e.g. Marquardt *et al*., 1998).

7.4 Future directions

Proximal soil sensors have often been used to gather data in the field, but not process and use this information in the field. This mentality needs to change. New tools must be developed with a focus of converting these information streams in to something useable. Whether that takes the form of directing further sampling sites (Horta *et al.*, 2014), informing pedogenetic theories, or agronomic decision-making processes.

7.4.1 Soil constraint management

Practical applications of proximal soil sensors need to be given greater relevance to promote adoption. An example could be supporting remediation of hydrophobic soils. Hydrophobic soils occupy more than 5 million hectares of western and southern Australia (Roper, 2004). Soil hydrophobicity normally occurs in soils with a clay content of less than 5 g 100 g^{-1}. To remediate such soils, clay is sourced from on-site pits and incorporated by spreading. Yield

increases of 40-50% have been observed (GRDC, 2016). Ideal clay sources for spreading are yellow, kaolinitic type, with a clay content of ~30 g 100 g^{-1} and low EC, CO_3 and B. However, a recent analysis of 82 clay pits identified a ranges in: clay content of 10-75 g 100 g^{-1}; CEC of 3-28 $cmol_{(+)}$ kg^{-1}; pH (1:5 $CaCl_2$) of 4.7-8.7; and EC of 0.0-2.3 dS m^{-1} (GRDC, 2016). Applying a standard rate of clay from such a wide range of sources can produce negligible, or even negative effects. The use of proximal soil sensors could be beneficial here to: i) identify soils with a clay content of less than 5 g 100 g^{-1}; and ii) analyse clay sources and adjust application rates during spreading. A similar approach could also be imagined for liming soils by measuring the purity of lime being applied, as well as the pH and buffering capacity of soil continuously across a paddock.

7.4.2 Adaptive mapping techniques to support precision agriculture

All crop management decisions today are made based on incomplete knowledge of soil properties and variability. Soil properties vary continuously across a paddock and this variability may have large implications for crop selection, establishment, growth and development. To facilitate true site-specific crop management this data gap must be filled. A system must be developed that will collate survey data to identify locations to be further investigated using ambulatory proximal soil sensors, such as VisNIR and pXRF. This combined approach will allow fine-scale mapping of multiple soil properties, and sampling methodologies could be adapted in real-time to minimise mapping uncertainties (Fig. 7.1).

Soil condition must be optimised for crop growth and to also secure soil for production into the future (McBratney *et al.*, 2014). Kravchenko and Bullock (2000) investigated the spatial distribution of four soil properties and found that even this small number of properties could explain an average of 30% of the yield variation in soybean and corn crops. Grain quality measures are similarly influenced by soil properties (Adams *et al.*, 2004; Stewart *et al.*, 2002). Undoubtedly, more soil data collected efficiently and in a cost-effective manner will result in better decision-making and more profitable farming systems.

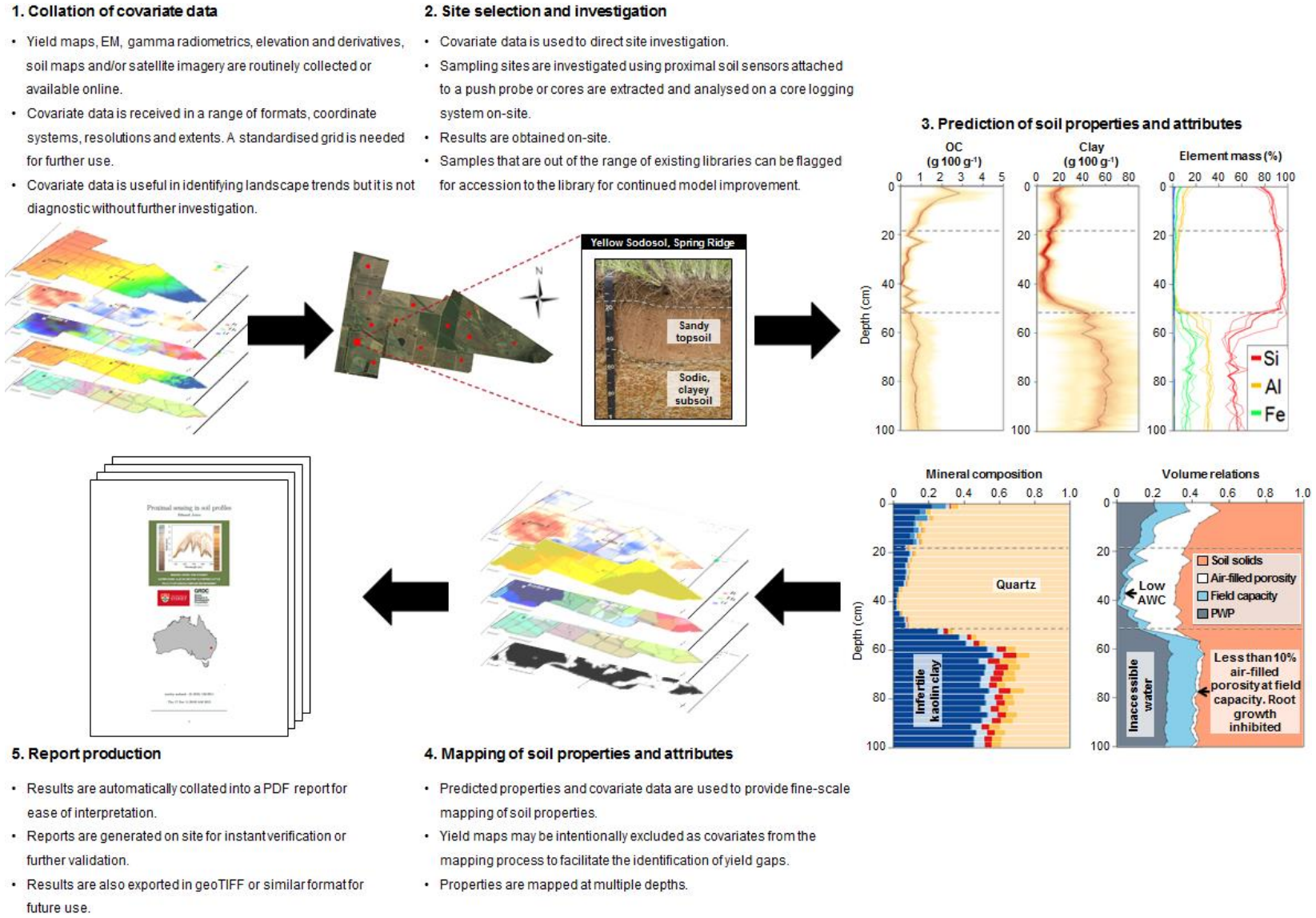


Figure 7.1 Adaptive sampling technique to map, and report on, soil attributes.

We must also use these predicted soil properties in a more purposeful manner. If we assess these soil properties concurrently and combine them with the vast amount of information stored in pedotransfer functions and expert knowledge in the form of a soil inference system (McBratney *et al.*, 2002), we can greatly increase the agronomic relevance of the predictions, such as mapping available water or pH buffering capacity. Connecting these pools of information is vital to boost the benefit of these predicted properties by transforming them into actionable soil attributes that can inform agronomic decision-making processes, and identify potential soil constraints and inherent soil cropping potential.

7.5 References

Adams, M. L., Zhao, F. J., McGrath, S. P., Nicholson, F. A., Chambers, B. J. (2004). Predicting cadmium concentrations in wheat and barley grain using soil properties. Journal of Environmental Quality 33(2), 532-541.

Araújo, S., Wetterlind, J., Demattê, J., Stenberg, B. (2014). Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. European Journal of Soil Science 65(5), 718-729.

Chang, C.-W., Laird, D. A., Mausbach, M. J., Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. Soil Science Society of America Journal 65(2), 480-490.

Clark, R. N., Swayze, G. A., Livo, K. E., Kokaly, R. F., Sutley, S. J., Dalton, J. B., McDougal, R. R., Gent, C. A. (2003). Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. Journal of Geophysical Research: Planets 108(E12).

Farmer, V.C. and Russell, J.D. (1964). The infrared spectra of layer silicates. Spectrochimica Acta, 20**,** 1149–1173.

Freeman, John J., et al. Characterization of natural feldspars by Raman spectroscopy for future planetary exploration. (2008). The Canadian Mineralogist 46(6), 1477-1500.

Frost, R. L., Kloprogge, J. T. (2003). Raman spectroscopy of some complex arsenate minerals—implications for soil remediation. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 59(12), 2797-2804.

Grains Research and Development Corporation (GRDC) "Tests urged before claying water – repellent soils". *GroundCover™ Supplement Issue (123)*. [Accessed – 01/01/2017] Available online: https://grdc.com.au/resources-and-publications/groundcover/ground-cover-supplements/ground-cover-issue-123-more-profit-from-crop-nutrition-2/tests-urged-before-claying-water-repellent-soils

Haumaier, L., Zech, W. (1995). Black carbon—possible source of highly aromatic components of soil humic acids. Organic Geochemistry 23(3), 191-196.

Hartemink, A. E., Minasny, B. (2014). Towards digital soil morphometrics. Geoderma 230, 305-317.

Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T., McBratney, A., Pallasser, R., Pozza, L. (2015). Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. Geoderma 241, 180-209.

Janik, L., Skjemstad, J., Raven, M. (1995). Characterization and analysis of soils using mid infrared partial least-squares .1. Correlations with XRF-determined major-element composition. Soil Research 33(4), 621-636.

Kravchenko, A. N., Bullock, D. G. (2000). Correlation of corn and soybean grain yield with topography and soil properties. Agronomy Journal 92(1), 75-83.

Marquardt, B. J., Stratis, D. N., Cremers, D. A., Angel, S. M. (1998). Novel probe for laser-induced breakdown spectroscopy and Raman measurements using an imaging optical fiber. Applied Spectroscopy 52(9), 1148-1153.

McBratney, A. B., Minasny, B., Cattle, S. R., Vervoort, R. W. (2002). From pedotransfer functions to soil inference systems. Geoderma 109(1), 41-73.

McBratney, A., Field, D. J., Koch, A. (2014). The dimensions of soil security. Geoderma 213, 203-213.

McCarty, G. W., Reeves, J. B., Reeves, V. B., Follett, R. F., Kimble, J. M. (2002). Mid-Infrared and Near-Infrared Diffuse Reflectance Spectroscopy for Soil Carbon Measurement. Soil Science Society of America Journal 66(2), 640-646.

Roper, M. M. (2004). The isolation and characterisation of bacteria with the potential to degrade waxes that cause water repellency in sandy soils. Soil Research 42(4), 427-434.

Sallé, B., Cremers, D. A., Maurice, S., Wiens, R. C., Fichet, P. (2005). Evaluation of a compact spectrograph for in-situ and stand-off laser-induced breakdown spectroscopy analyses of geological samples on Mars missions. Spectrochimica Acta Part B: Atomic Spectroscopy 60(6), 805-815.

Sorak, D., Herberholz, L., Iwascek, S., Altinpinar, S., Pfeifer, F., Siesler, H. W. (2012). New developments and applications of handheld Raman, mid-infrared, and near-infrared spectrometers. Applied Spectroscopy Reviews 47(2), 83-115.

Stewart, C. M., McBratney, A. B., Skerritt, J. H. (2002). Site-specific durum wheat quality and its relationship to soil properties in a single field in northern New South Wales. Precision Agriculture 3(2), 155-168.

Theriault, G. A., Bodensteiner, S., Lieberman, S. H. (1998). A real-time fiber-optic LIBS probe for the in situ delineation of metals in soils. Field Analytical Chemistry & Technology 2(2), 117-125.

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131(1–2), 59-75.

Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C. B., Knadel, M., Morrás, H. J. M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E. M. R., Sanborn, P., Sellitto, V. M., Sudduth, K. A., Rawlins, B. G., Walter, C., Winowiecki, L. A., Hong, S. Y., Ji, W. (2016). A global spectral library to characterize the world's soil. Earth-Science Reviews 155(Supplement C), 198-230.

Appendix A

A1 Correction factor for new pXRF detector and calibrations.

A change in detector technology and calibrations was implemented halfway through the sampling process for the pXRF. This upgrade used less power; however it also changed the calibrations and estimates of standard elemental composition. This change was identified through our continued monitoring of NIST and other standard materials, particularly NIST2709, NIST2711a and a SiO_2 standard. Analysis of 134 samples using the initial configuration and 126 samples between the updated detector exemplifies this change (Fig. 1). Both devices showed a high degree of precision given the range of conditions the scans were obtained in, e.g. lab v field, temperature (ambient and operating), altitude and pressure. Although the results did not always match exactly with standard reported values, and there was a change in accuracy with the new configuration. The differences between the two detector configurations appeared to be linear and as the initial detector setting showed a greater accuracy to the reported standard values it was decided that values given using the new detector should be corrected back to the previous setting. This also brought the new values into correction with our existing XRF library and allowed us to apply our established models more readily. Linear correction factors were calculated from the median reported concentration with the old and new detector configuration for each element. These correction factors were applied to reported results from the new detector for all analyses (Fig. 2)

Appendix

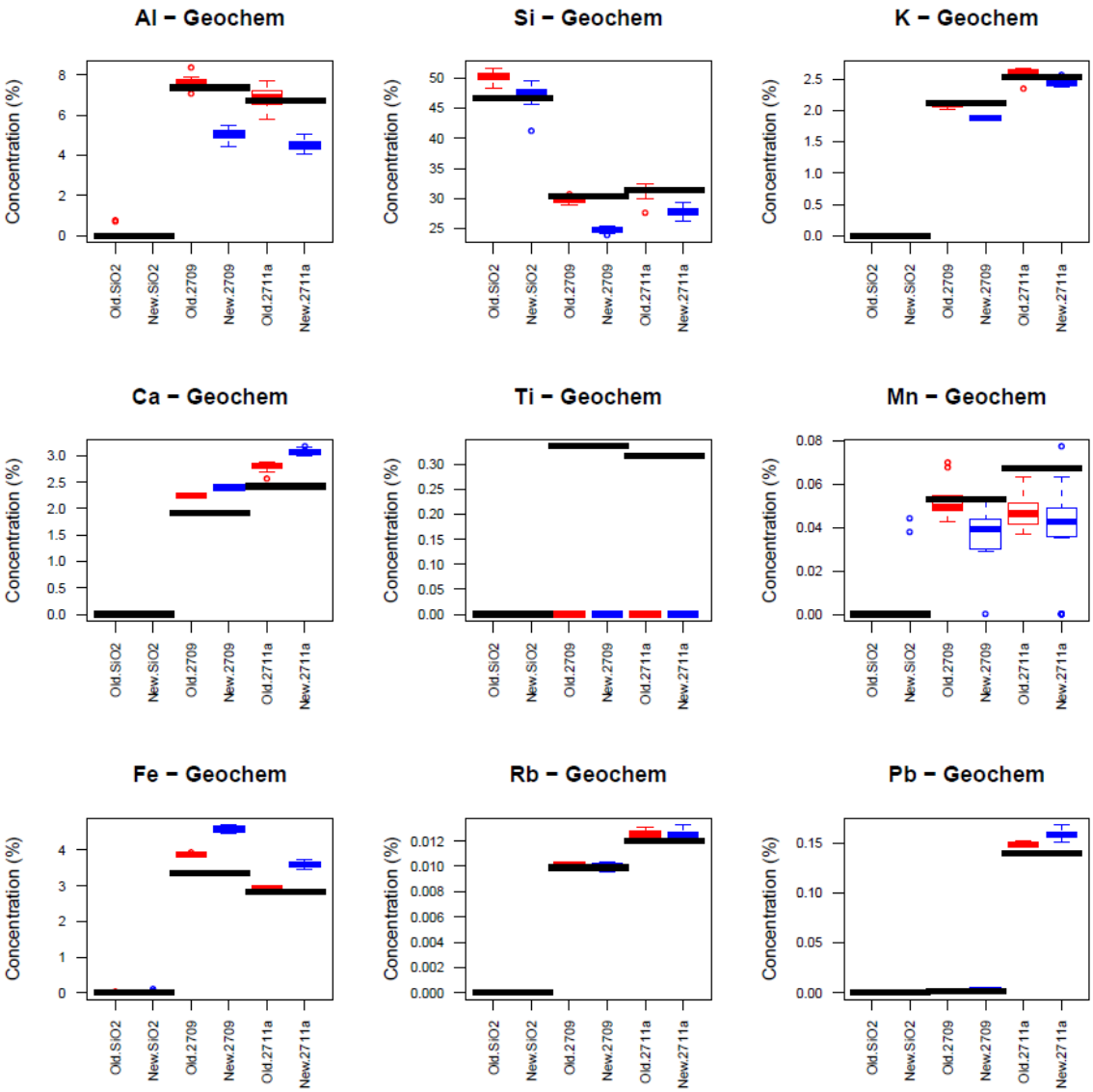


Figure 1 Boxplots of pXRF standards before and after the detector change. The old detector configuration is shown in red, the new detector configuration is shown in blue. Horizontal black bars indicate reported standard concentrations.

Appendix

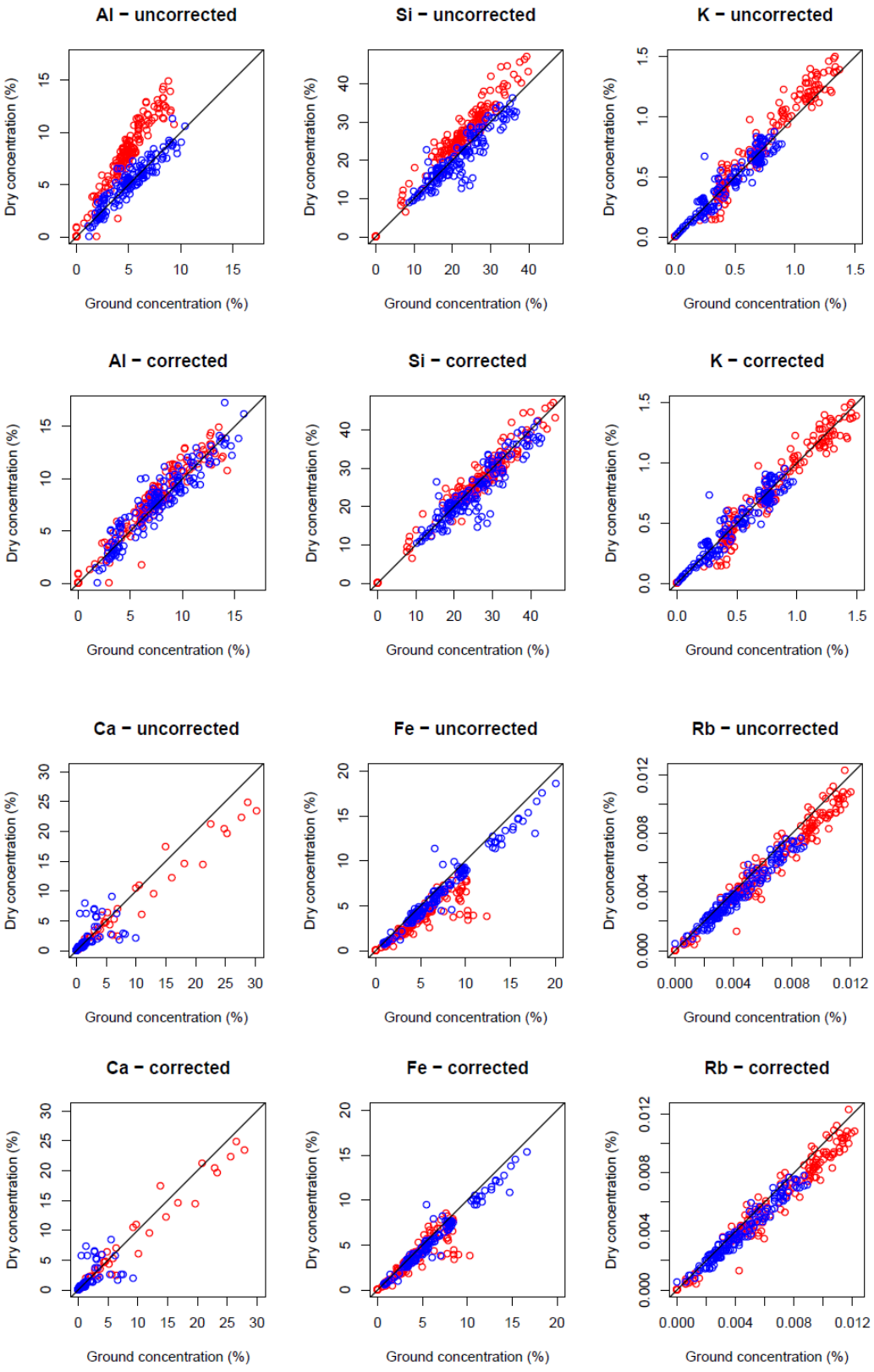


Figure 2 Comparison of pXRF reported values using the old detector (red) in dry condition and the new detector in ground condition, before and after a linear correction factor was applied. Samples scanned with the new detector only (blue) are included for comparison.

A2 Stability of geochemical ratios under variable soil moisture

The stability of geochemical ratios over direct use of pXRF reported elemental concentrations was demonstrated by comparing the correlation between samples scanned under field moist and air-dry condition. When pXRF values were used directly, moisture in the sample resulted in a decrease in observed elements (Fig. 3). Geochemical ratios reduced this effect; comparable results were obtained under variable soil moisture conditions (Fig. 4)

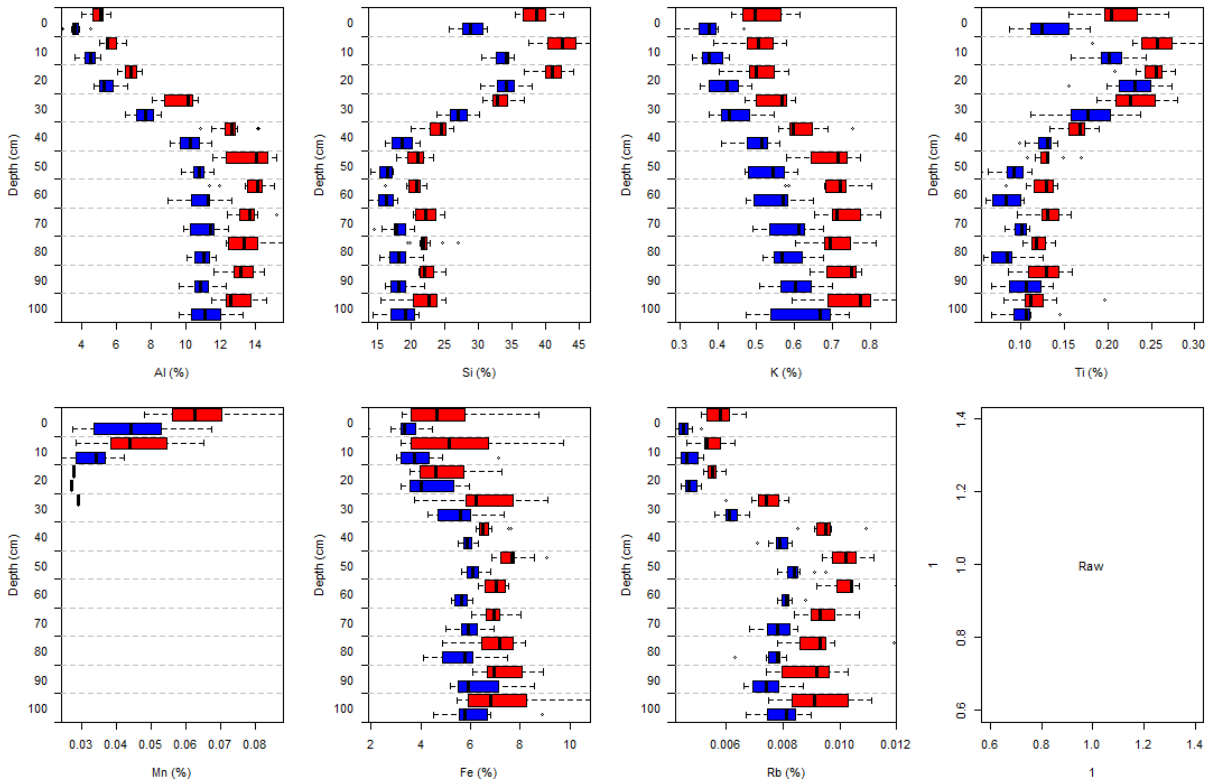


Figure 3 Boxplots of pXRF reported values under field moist (blue) and air-dry (red) scenarios. Samples were derived from the intensively sampled pit mentioned in Chapter 3).

Appendix

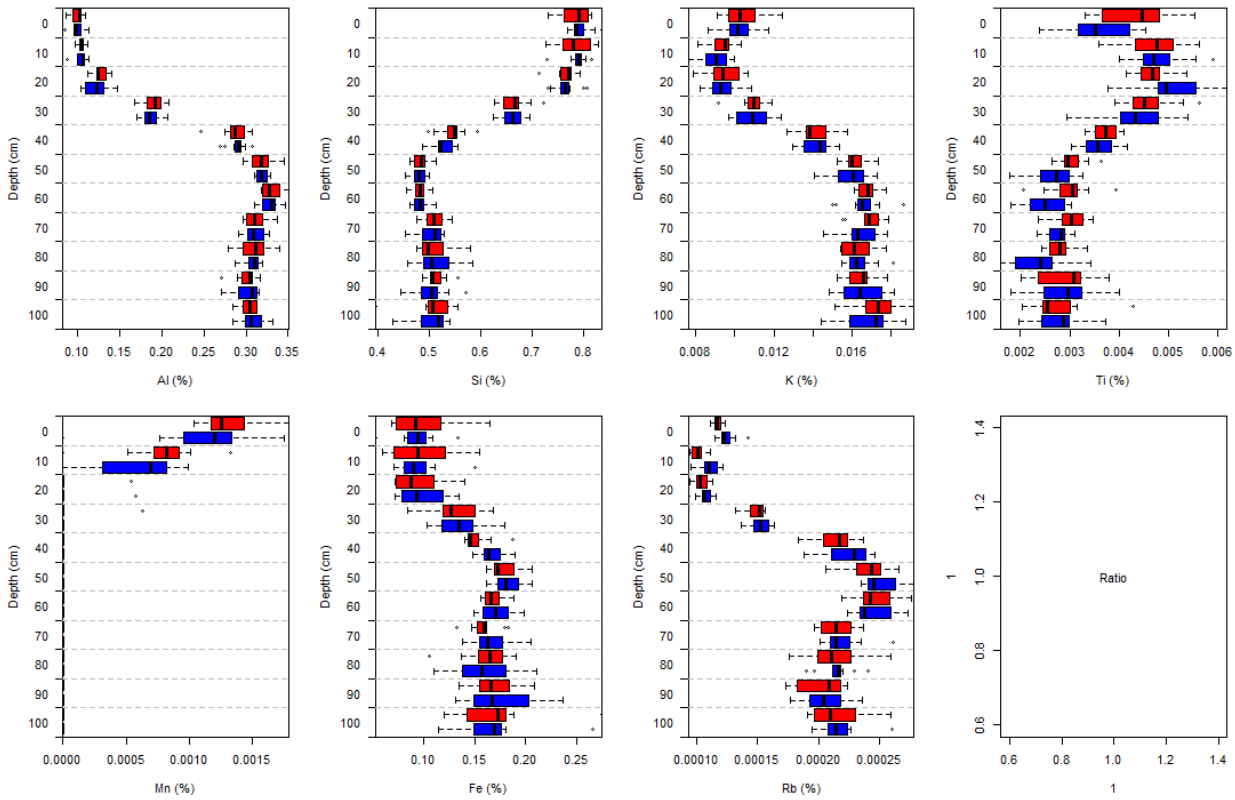


Figure 4 Boxplots of ratios of pXRF reported values under field moist (blue) and air-dry (red) scenarios. Samples were derived from the intensively sampled pit mentioned in Chapter 3). Ratios were obtained as outlined in section 2.2.5.3.

Appendix B

B1 XRD random powder scans of bulk samples and clay isolates

Notable peaks are indicated: Q – quartz; Ca –calcite; P – phyllosilicates; Gy – gypsum.

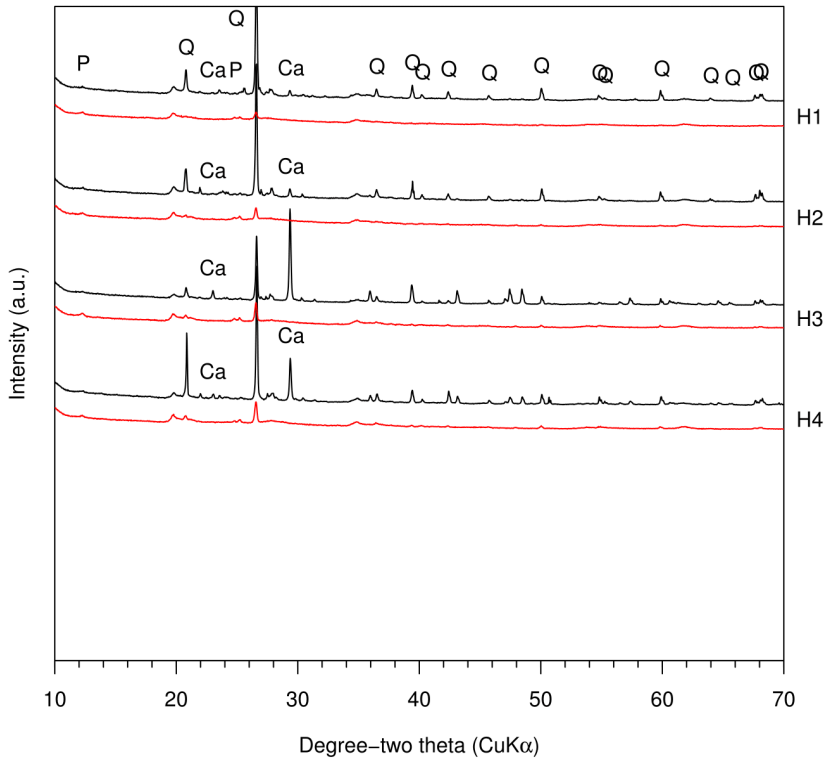


Figure 5 Random powder XRD scan bulk sample and clay isolate from Site 1.

Appendix

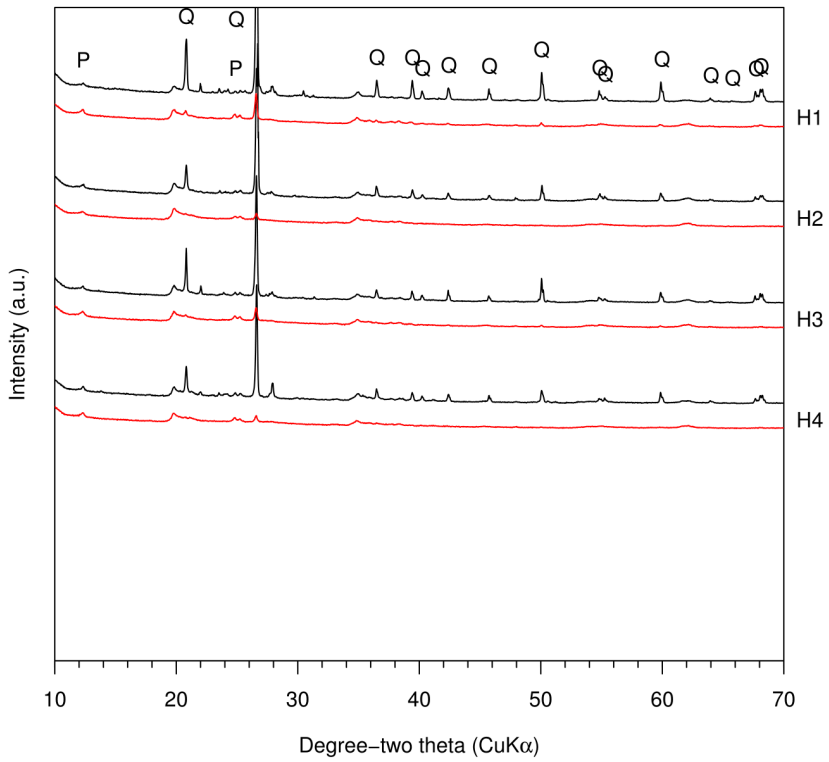


Figure 6 Random powder XRD scan bulk sample and clay isolate from Site 2.

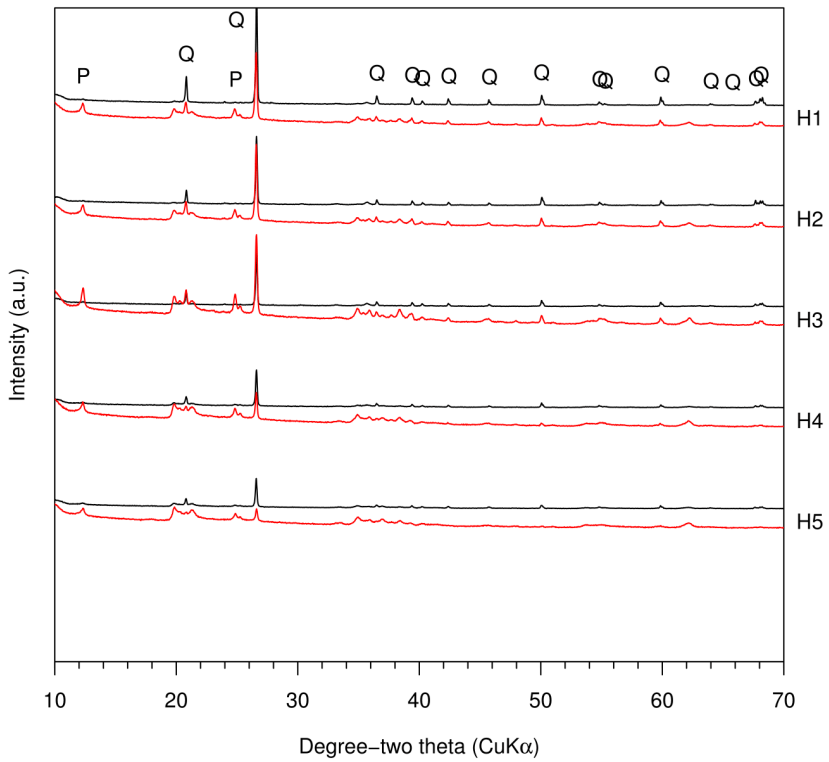


Figure 7 Random powder XRD scan bulk sample and clay isolate from Site 3.

Appendix

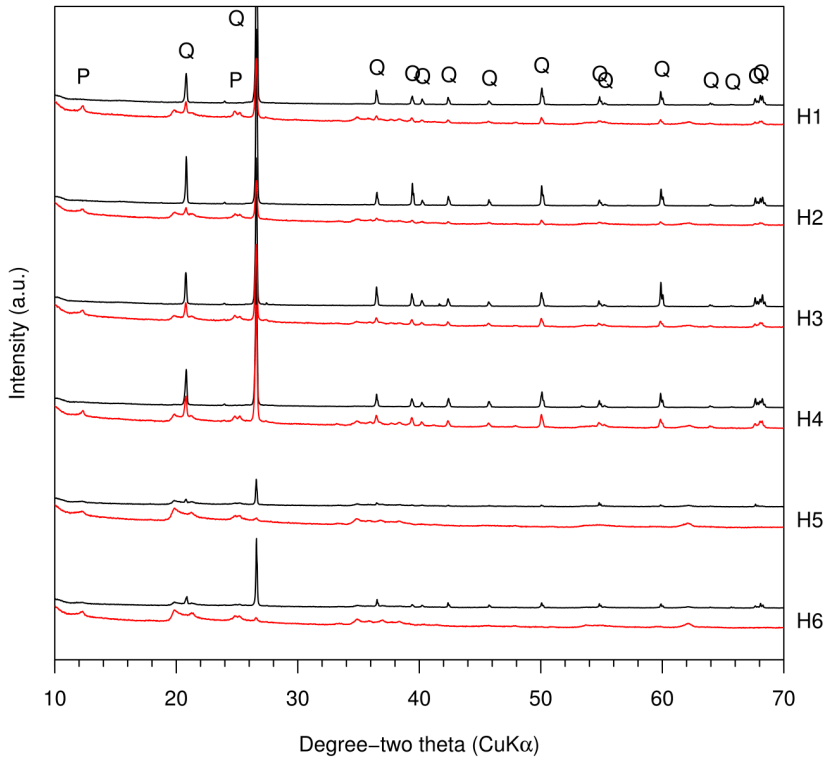


Figure 8 Random powder XRD scan bulk sample and clay isolate from Site 4.

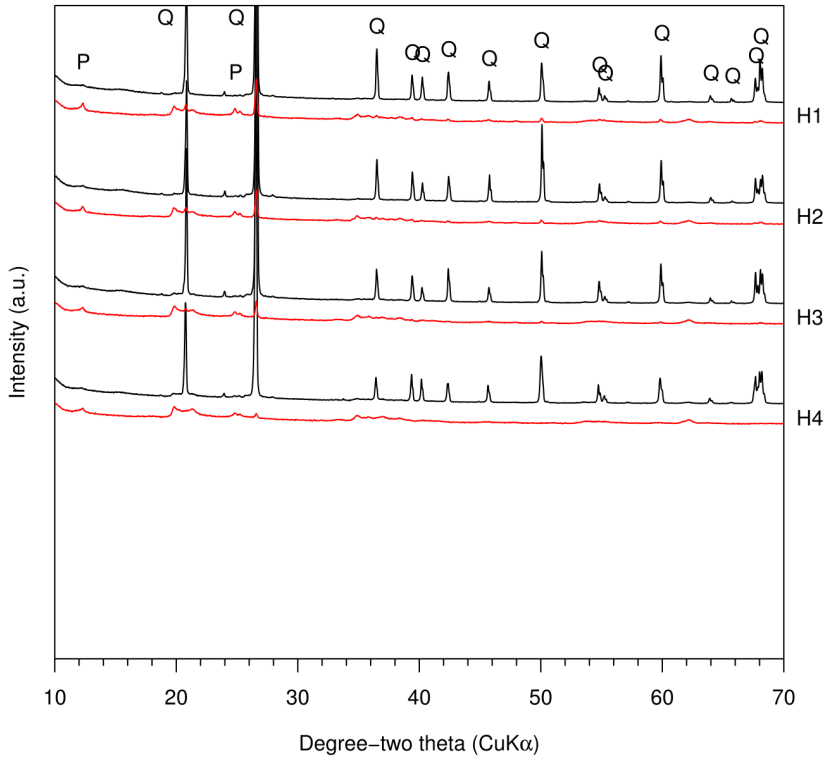


Figure 9 Random powder XRD scan bulk sample and clay isolate from Site 5.

Appendix

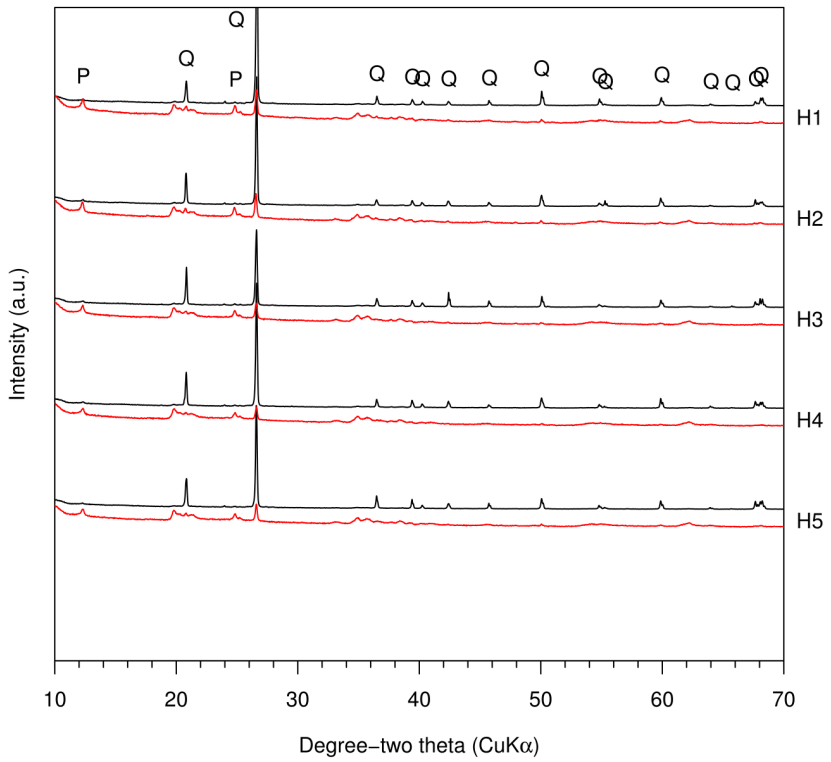


Figure 10 Random powder XRD scan bulk sample and clay isolate from Site 6.

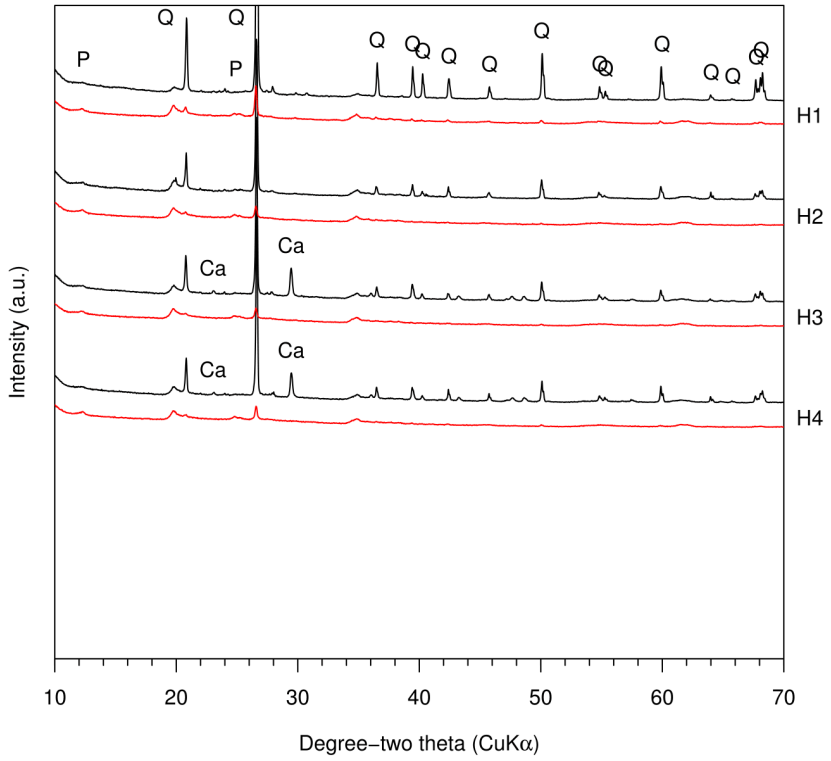


Figure 11 Random powder XRD scan bulk sample and clay isolate from Site 7.

Appendix

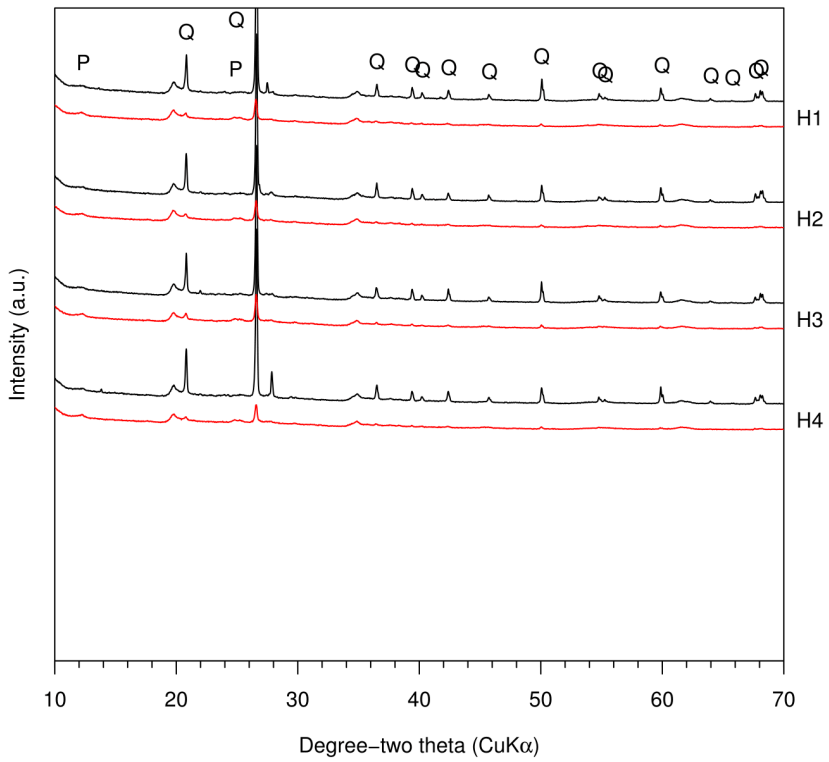


Figure 12 Random powder XRD scan bulk sample and clay isolate from Site 8.

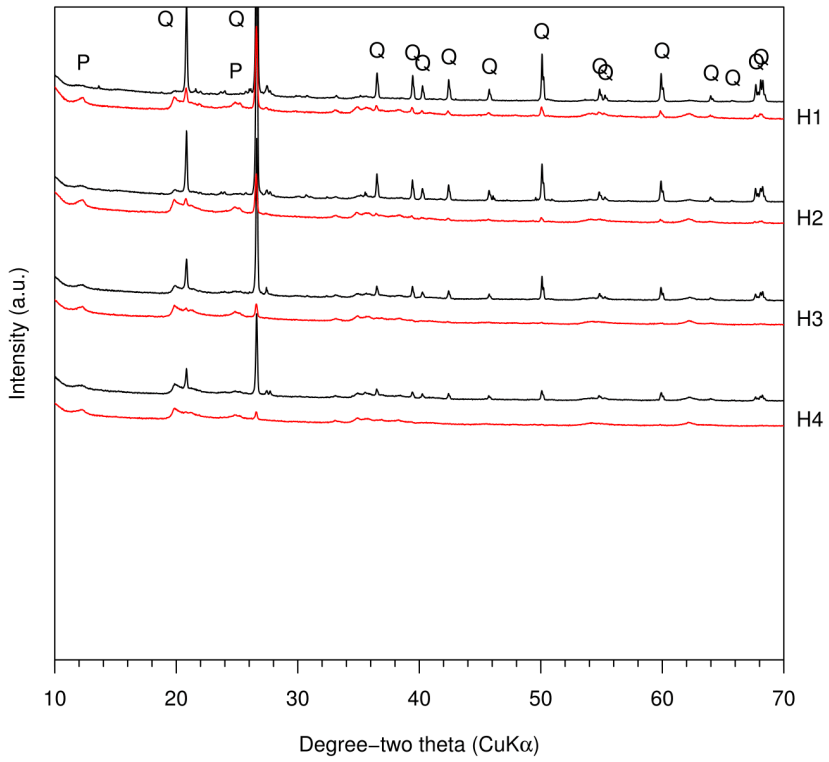


Figure 13 Random powder XRD scan bulk sample and clay isolate from Site 9.

Appendix

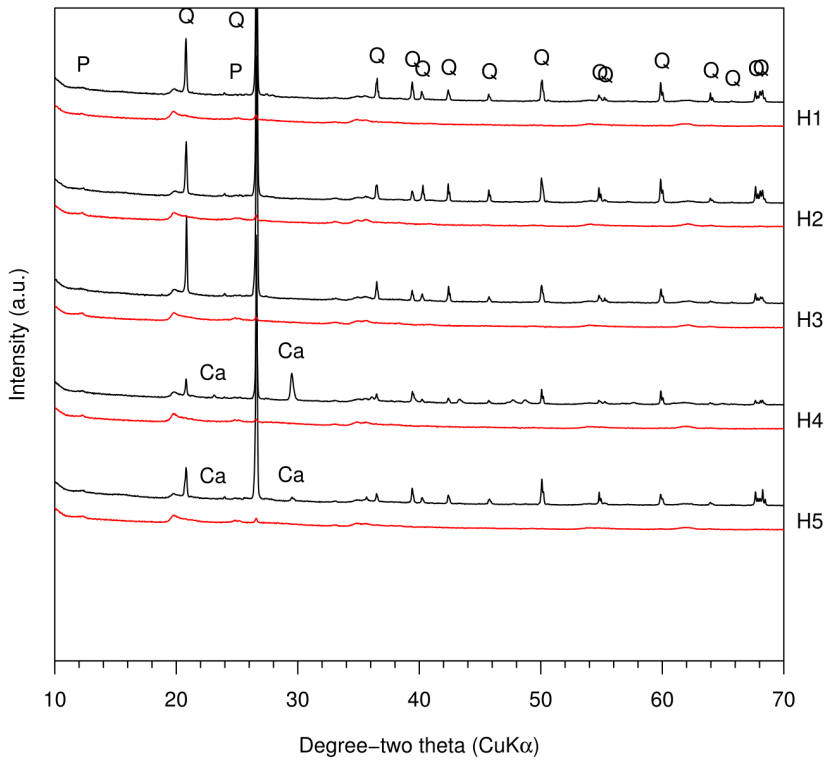


Figure 14 Random powder XRD scan bulk sample and clay isolate from Site 10.

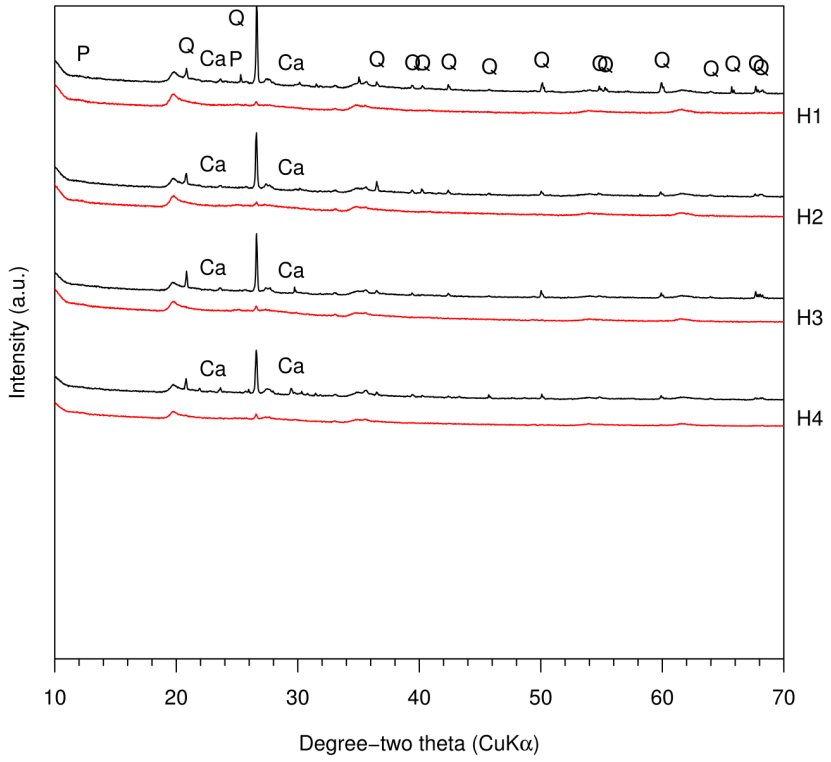


Figure 15 Random powder XRD scan bulk sample and clay isolate from Site 11.

Appendix

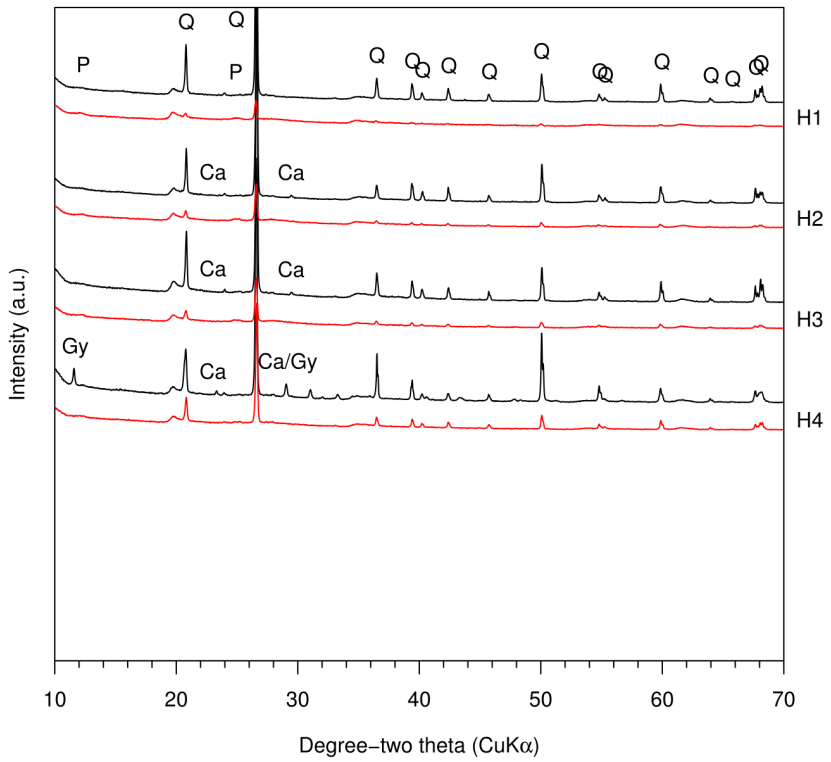


Figure 16 Random powder XRD scan bulk sample and clay isolate from Site 12.

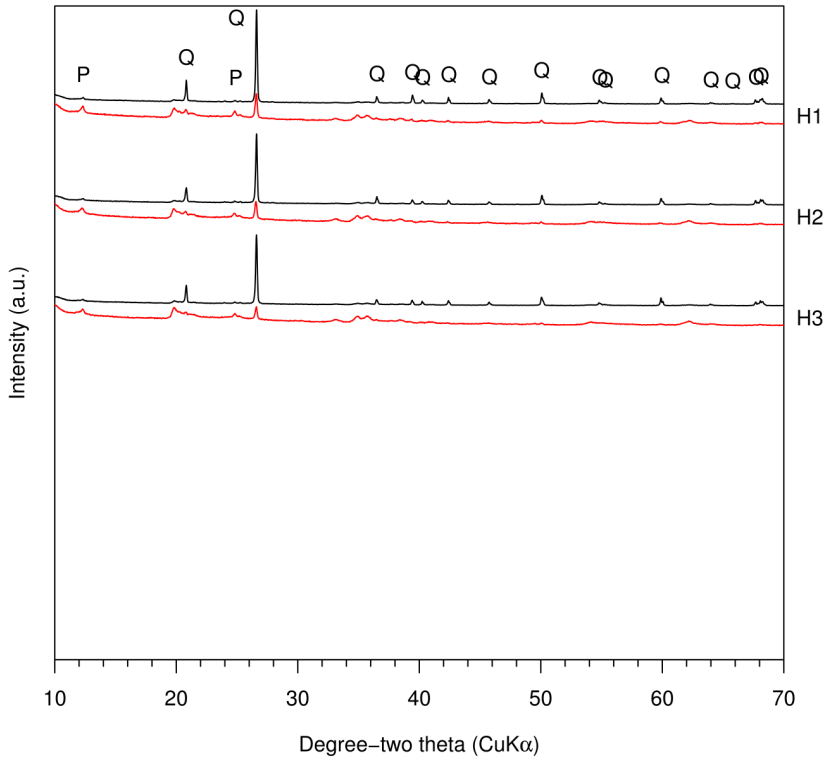


Figure 17 Random powder XRD scan bulk sample and clay isolate from Site 13.

Appendix

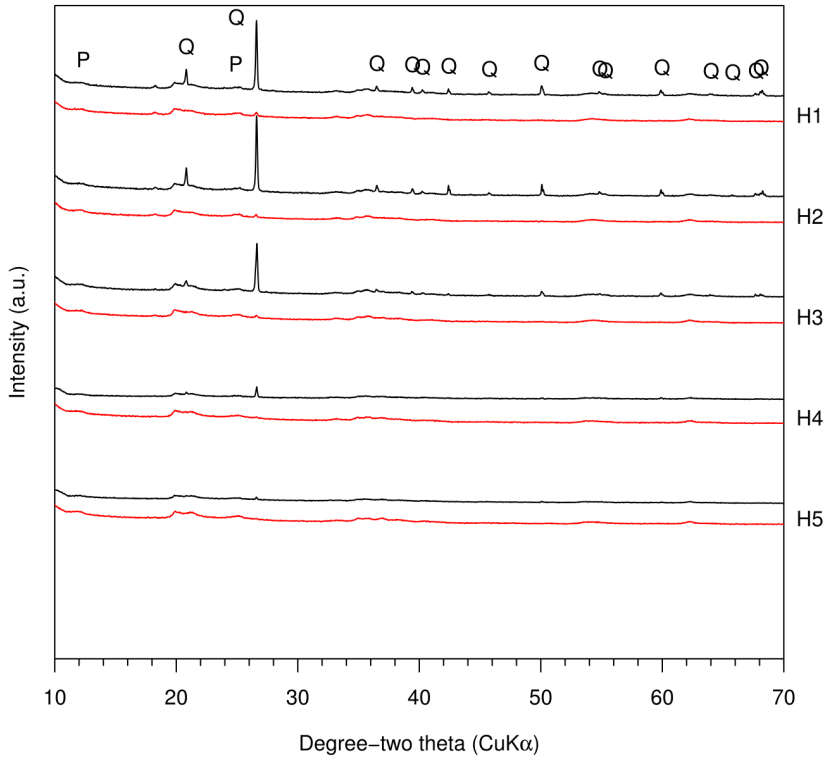


Figure 18 Random powder XRD scan bulk sample and clay isolate from Site 14.

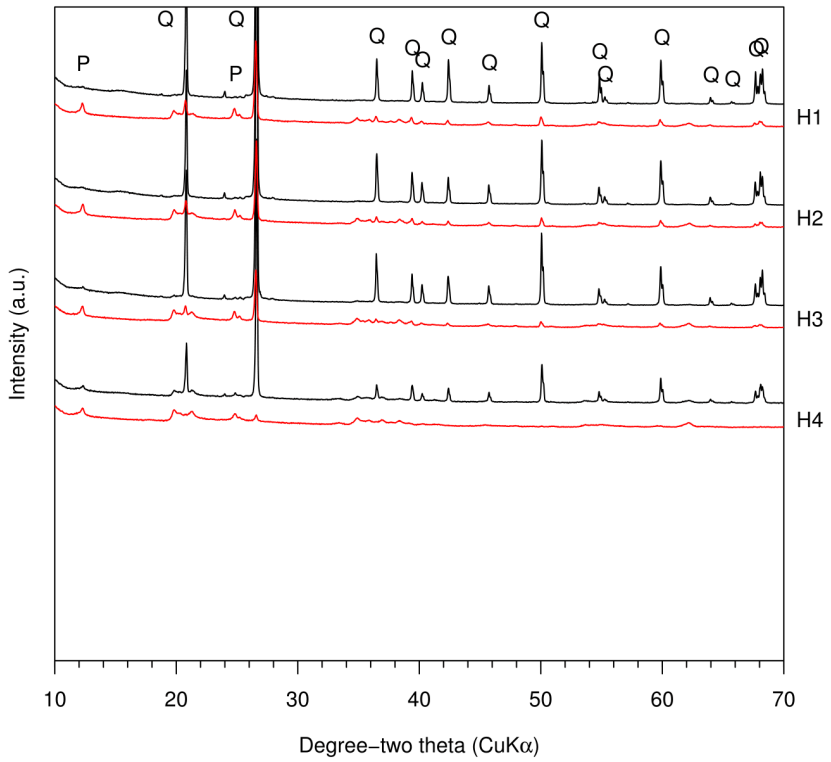


Figure 19 Random powder XRD scan bulk sample and clay isolate from Site 15.

B2 XRD oriented scans of clay isolates

XRD oriented scans of clay isolates following various pretreatments: Mg-saturated - black; Mg-saturated and ethylene glycolated - red; K-saturated - green; and K-saturated and heated to 550°C – blue.

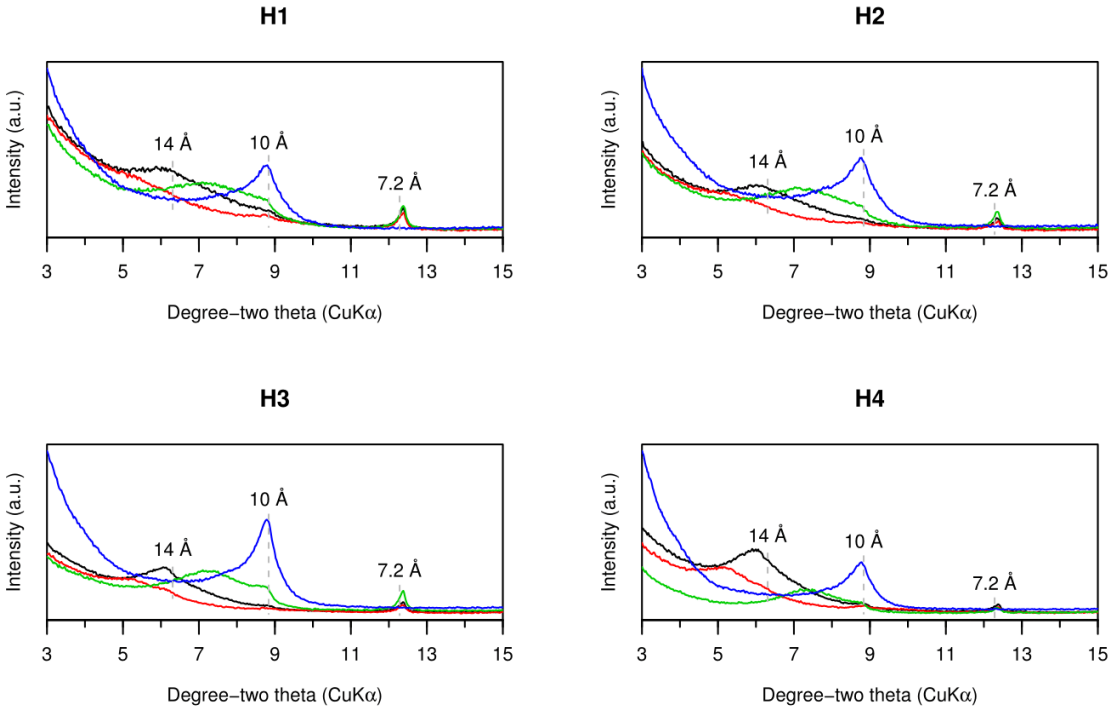


Figure 20 Oriented XRD scan following various pretreatments for Site 1.

Appendix

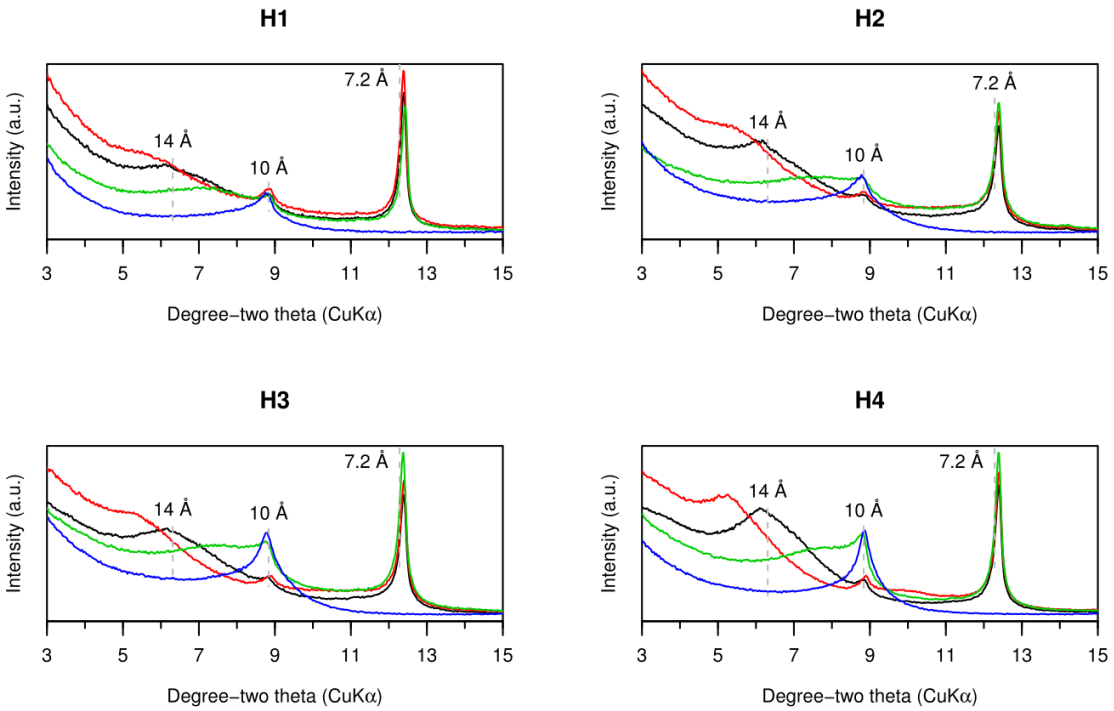


Figure 21 Oriented XRD scan following various pretreatments for Site 2.

Appendix

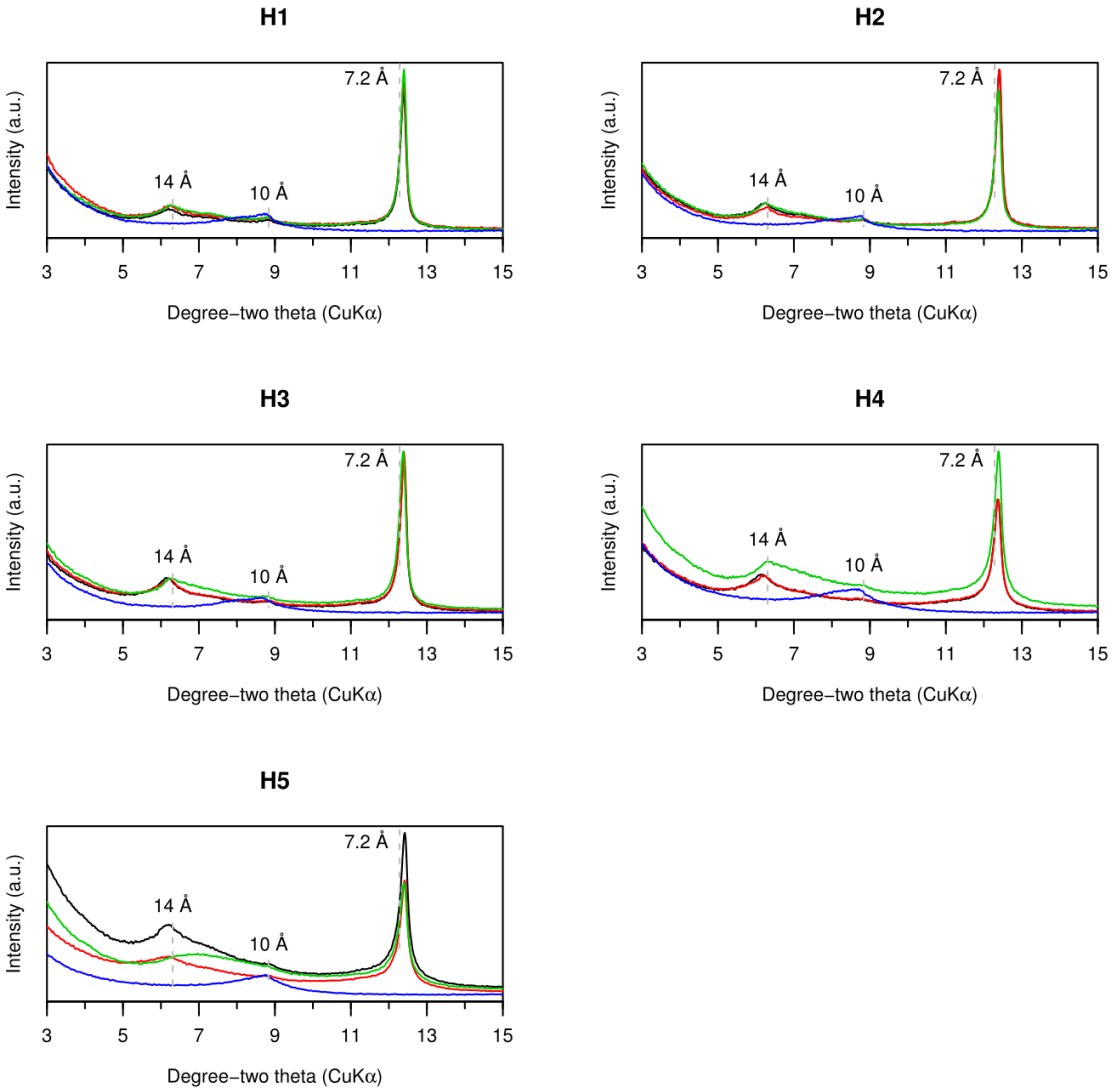


Figure 22 Oriented XRD scan following various pretreatments for Site 3.

Appendix

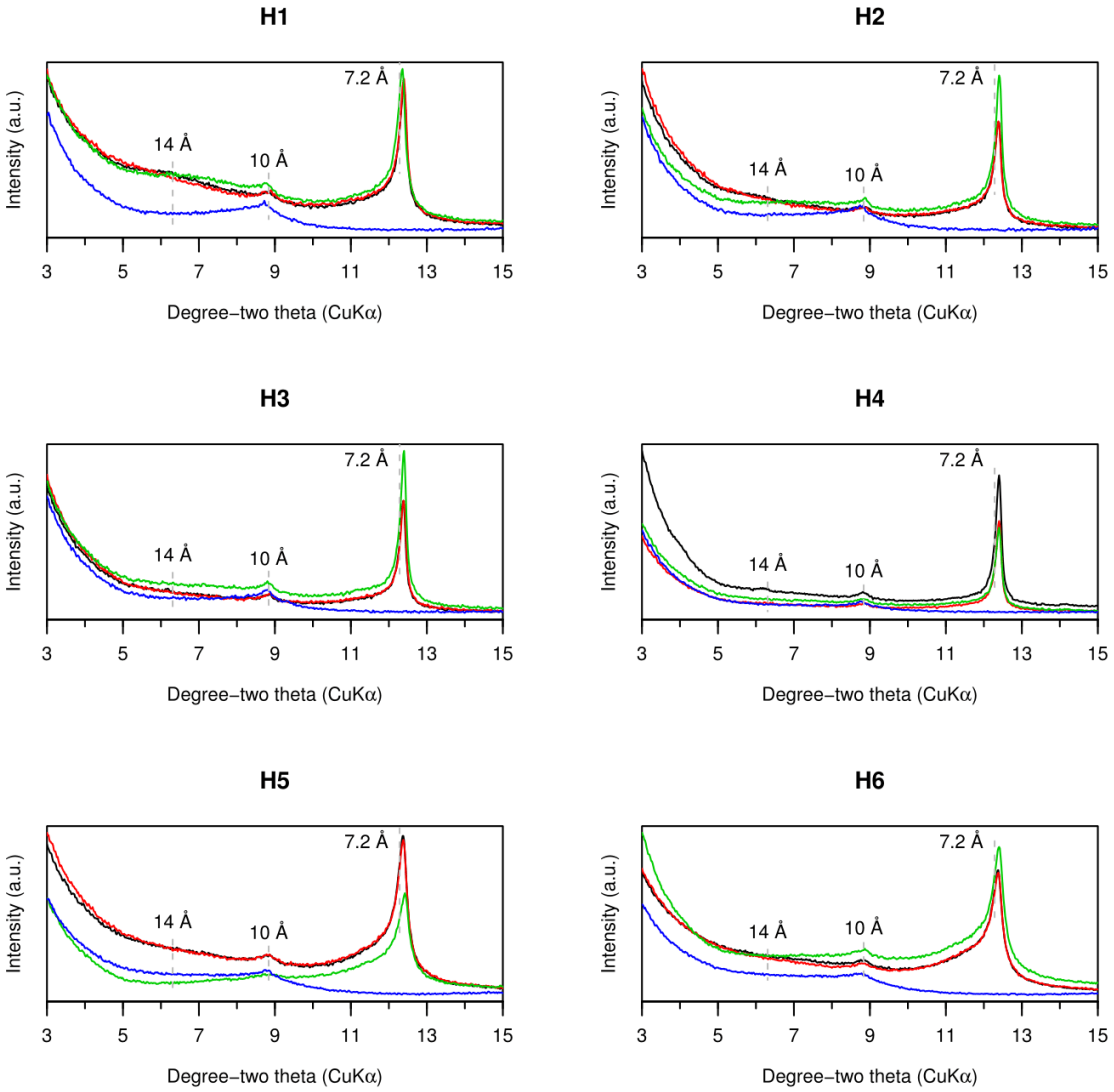


Figure 23 Oriented XRD scan following various pretreatments for Site 4.

Appendix

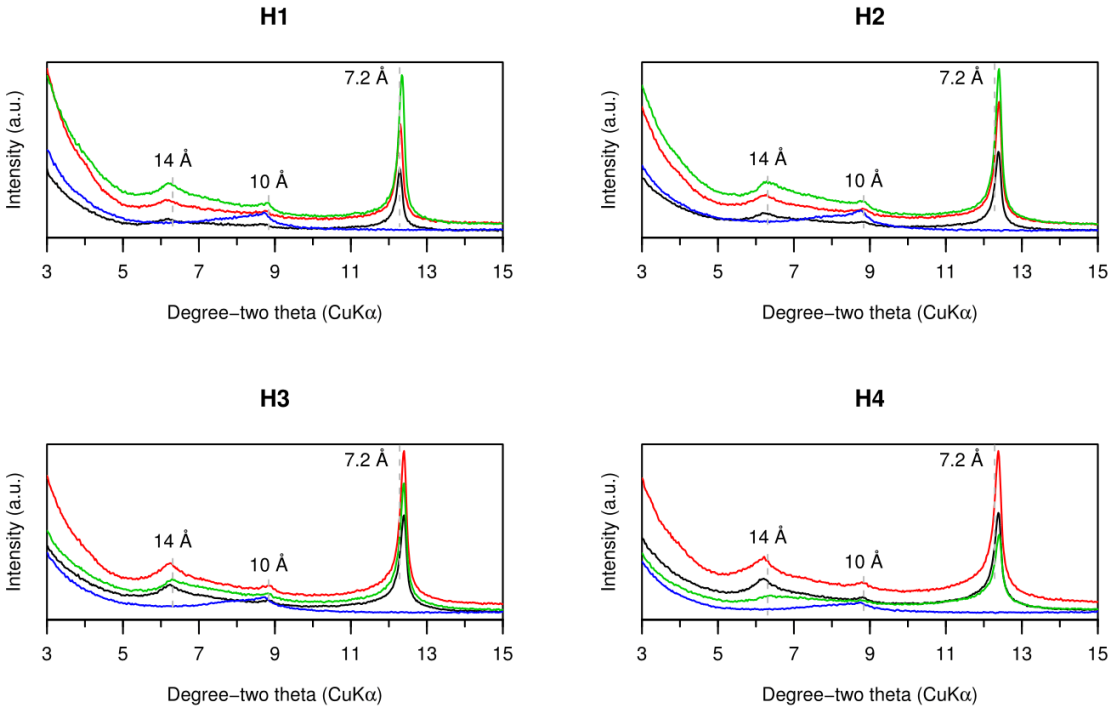


Figure 24 Oriented XRD scan following various pretreatments for Site 5.

Appendix

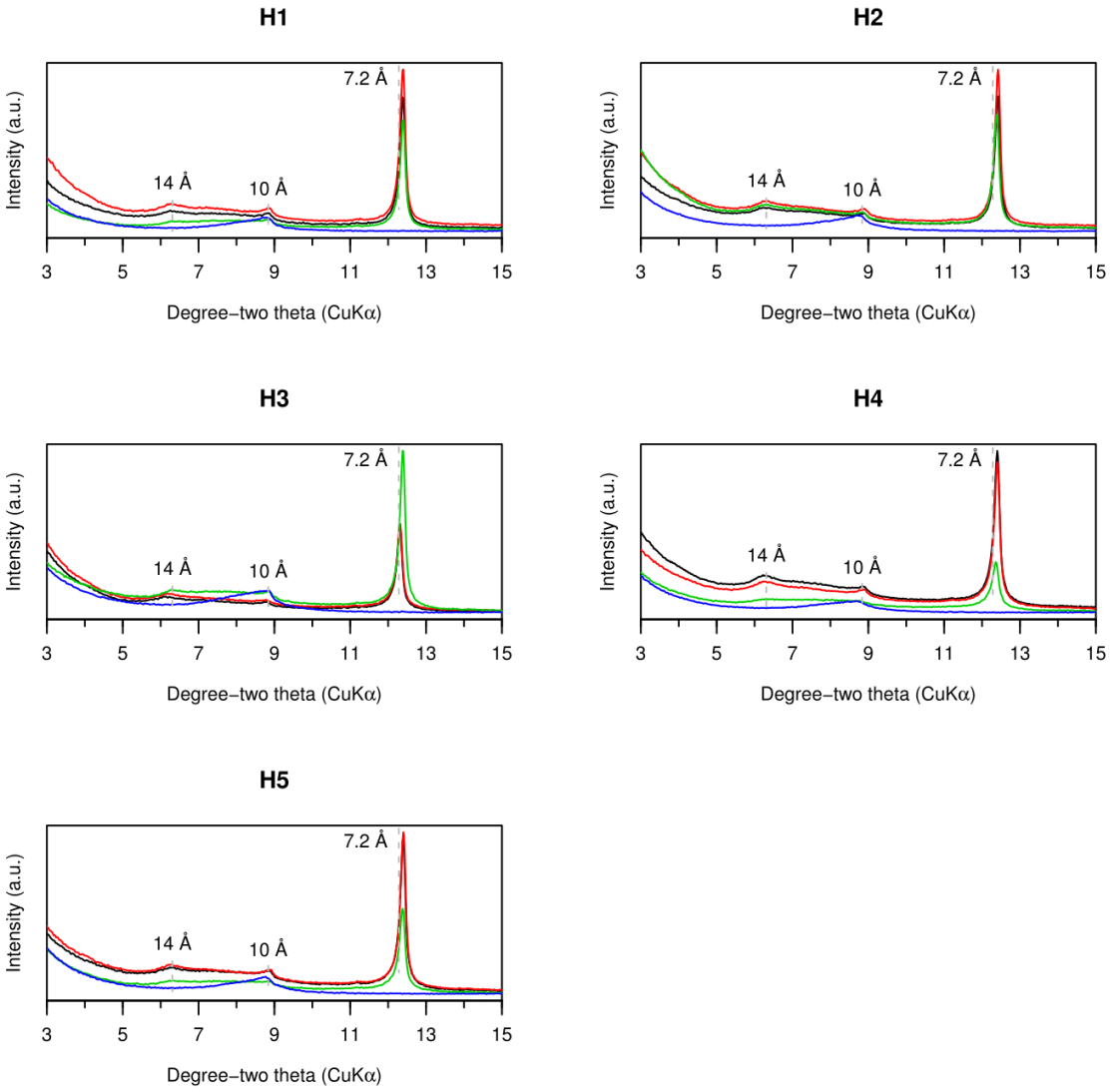


Figure 25 Oriented XRD scan following various pretreatments for Site 6.

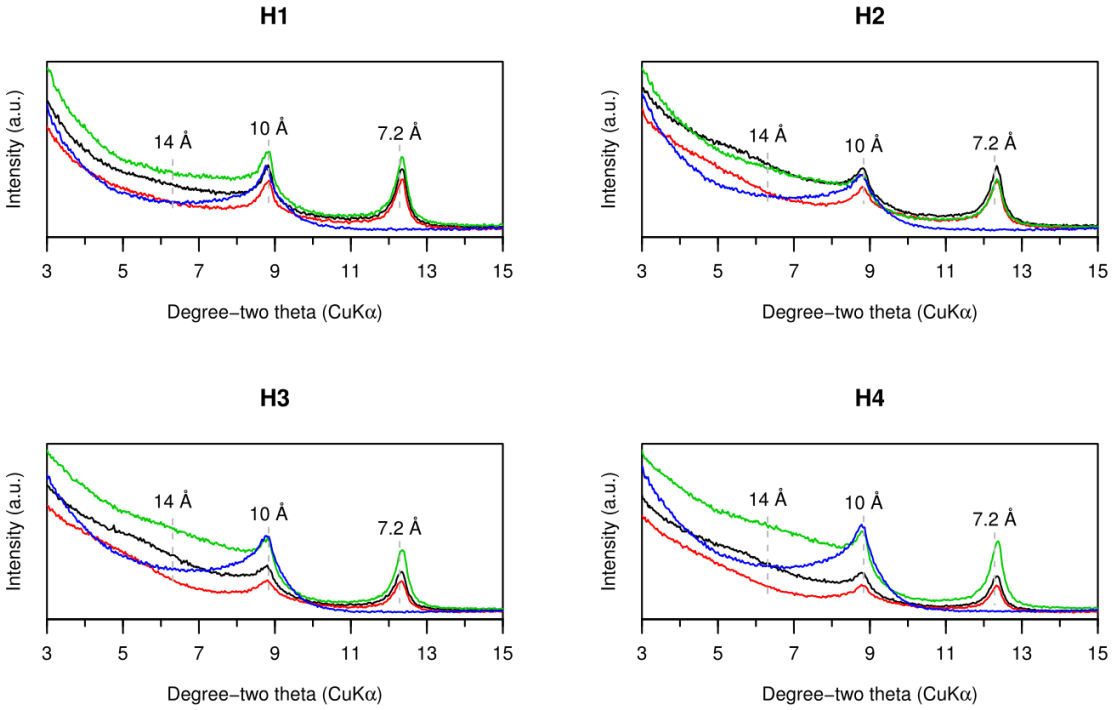


Figure 26 Oriented XRD scan following various pretreatments for Site 7.

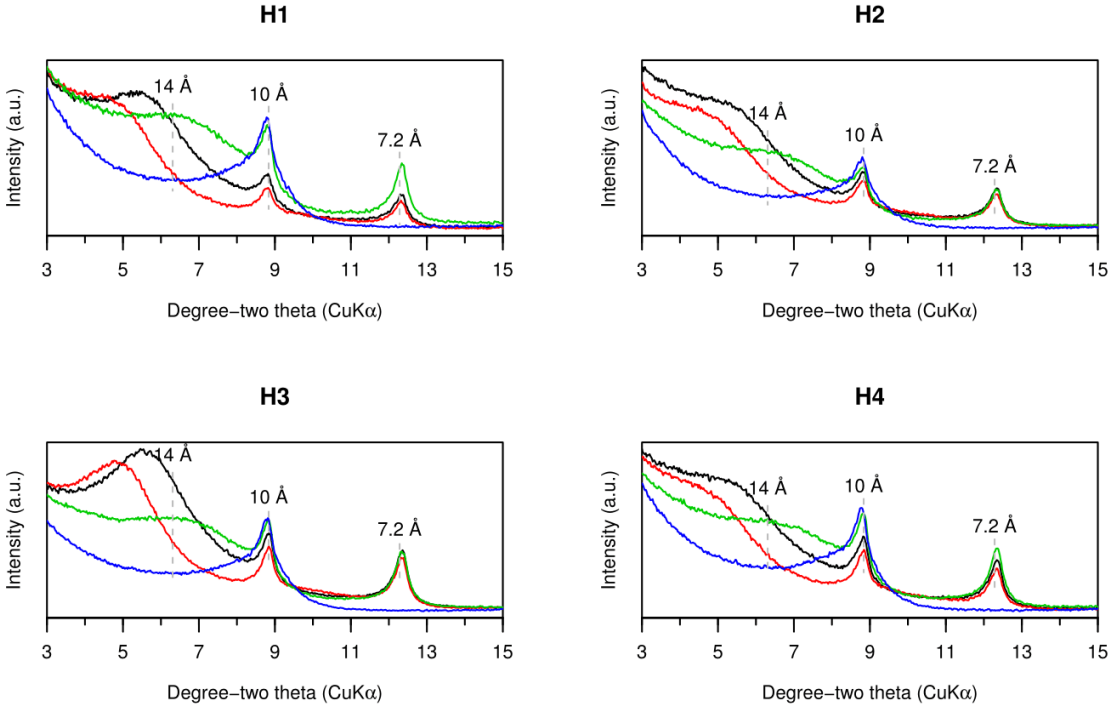


Figure 27 Oriented XRD scan following various pretreatments for 8.

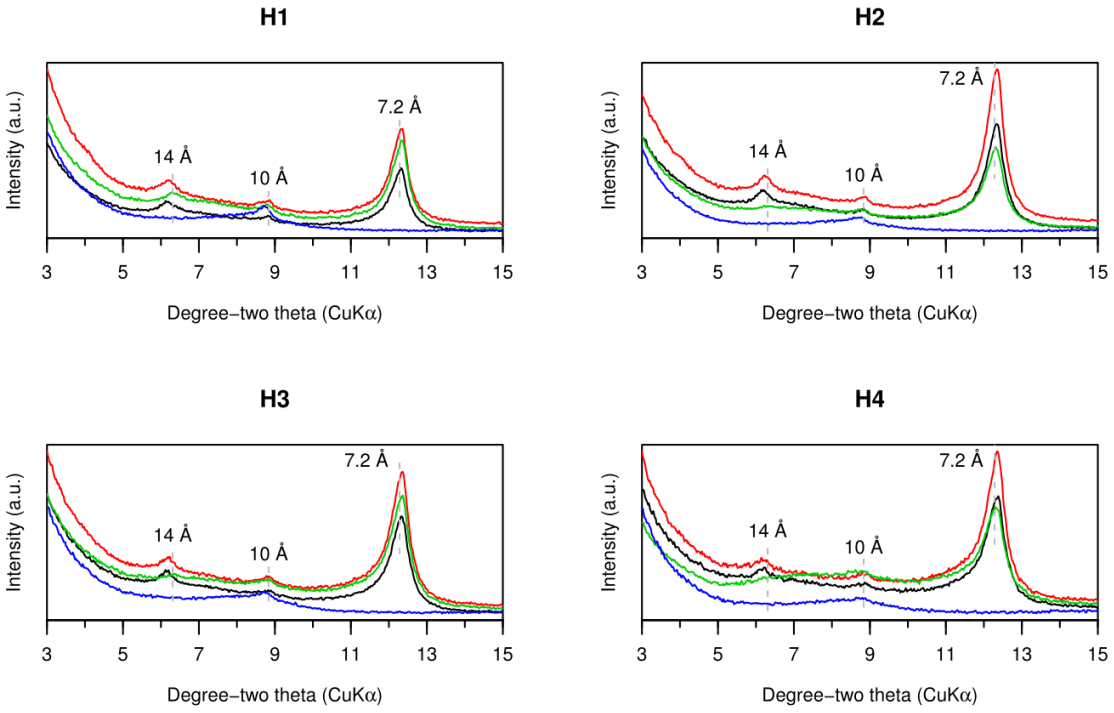


Figure 28 Oriented XRD scan following various pretreatments for Site 9.

Appendix

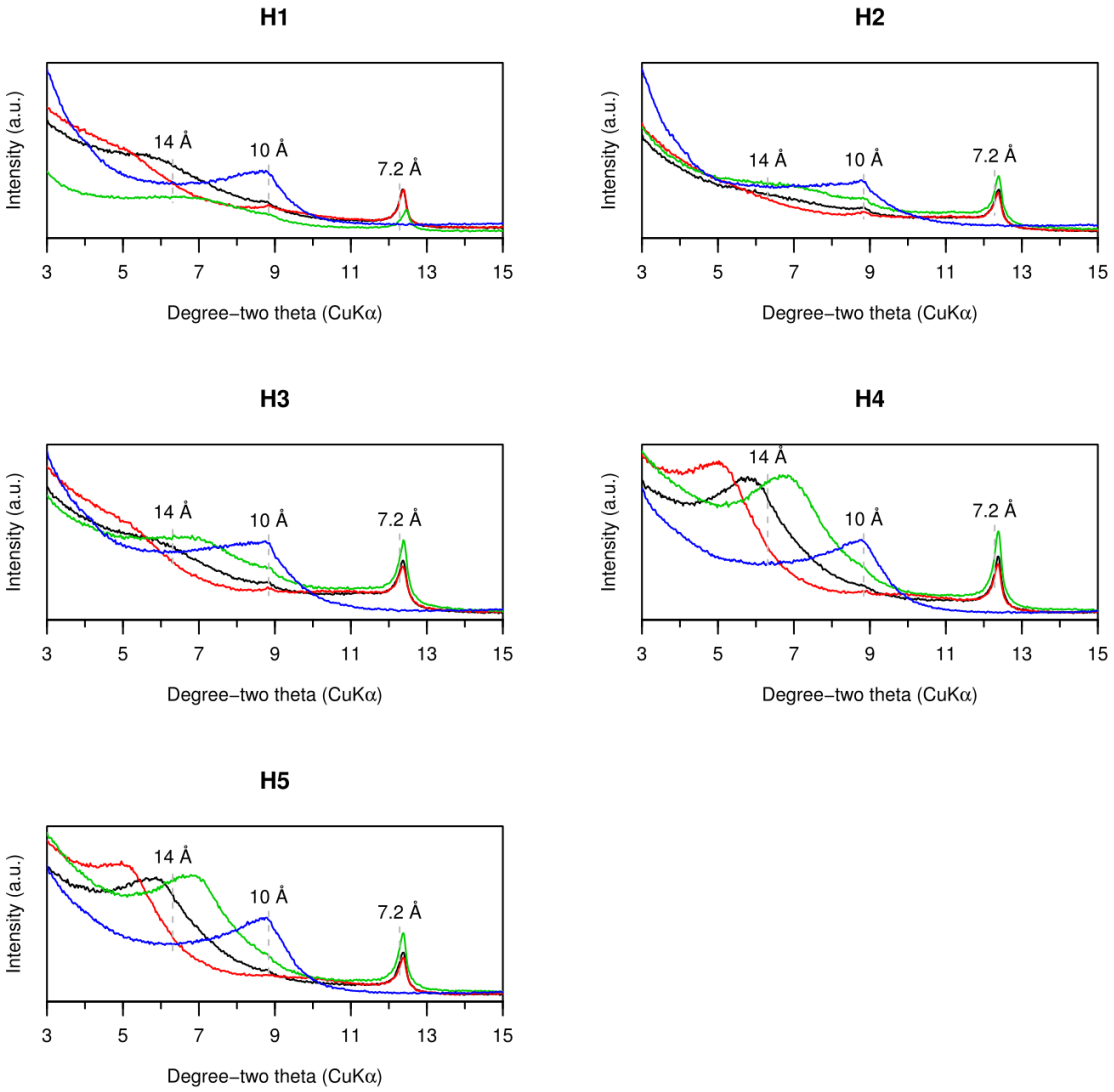


Figure 29 Oriented XRD scan following various pretreatments for Site 10.

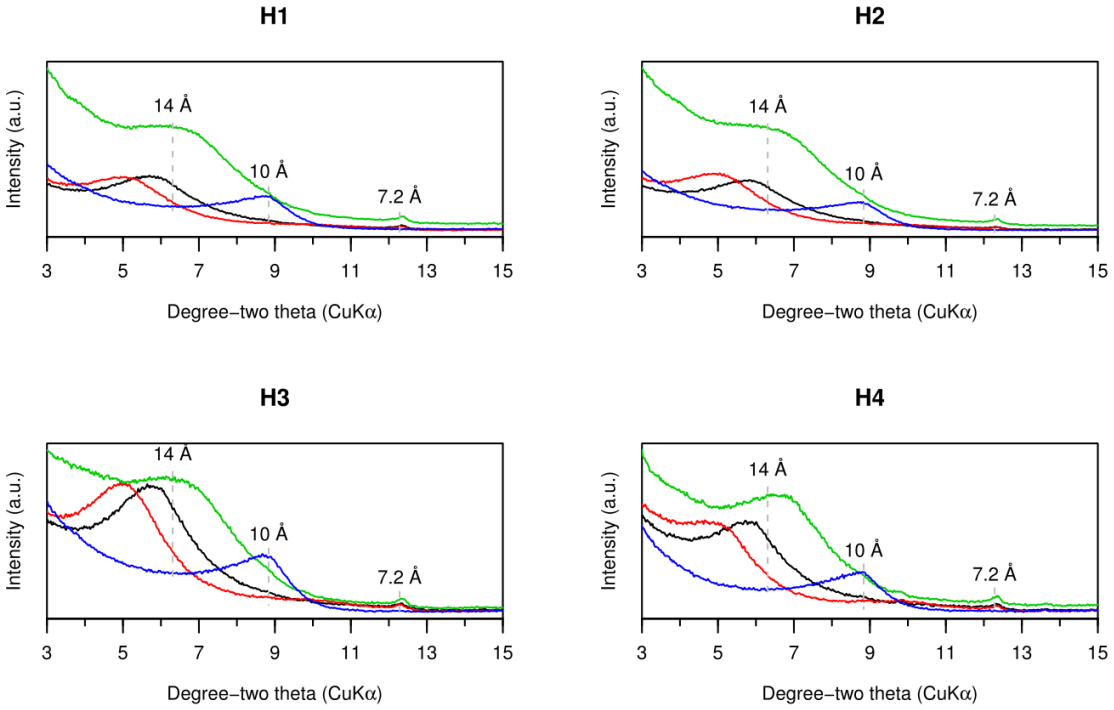


Figure 30 Oriented XRD scan following various pretreatments for Site 11.

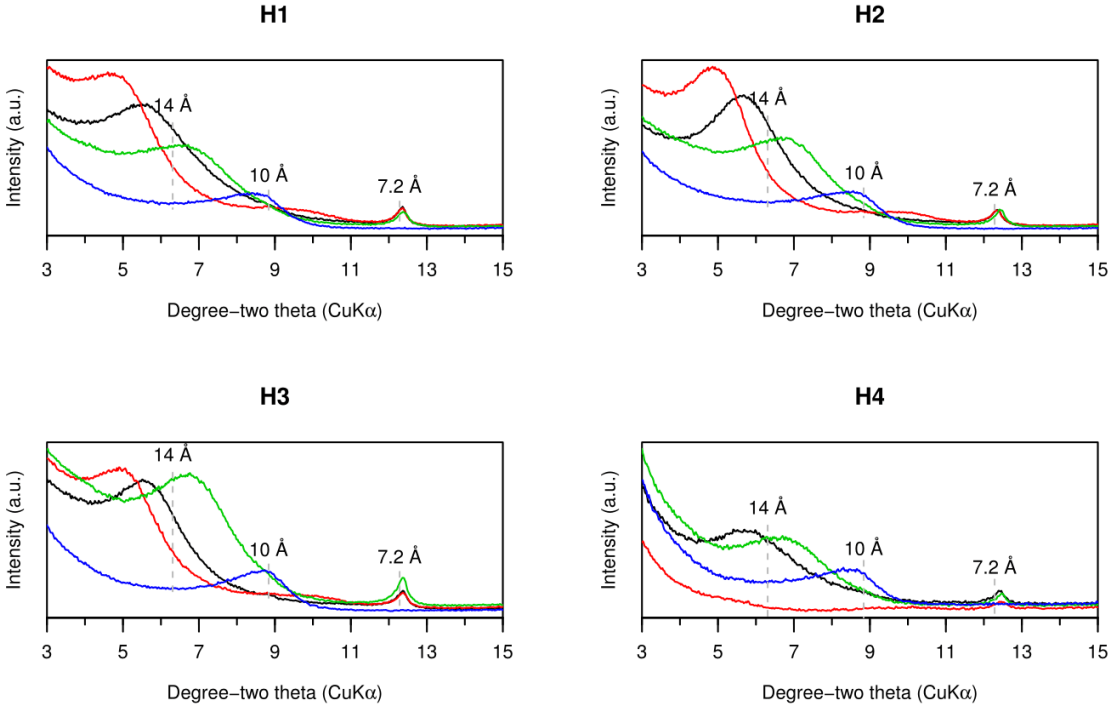


Figure 31 Oriented XRD scan following various pretreatments for 12.

Appendix

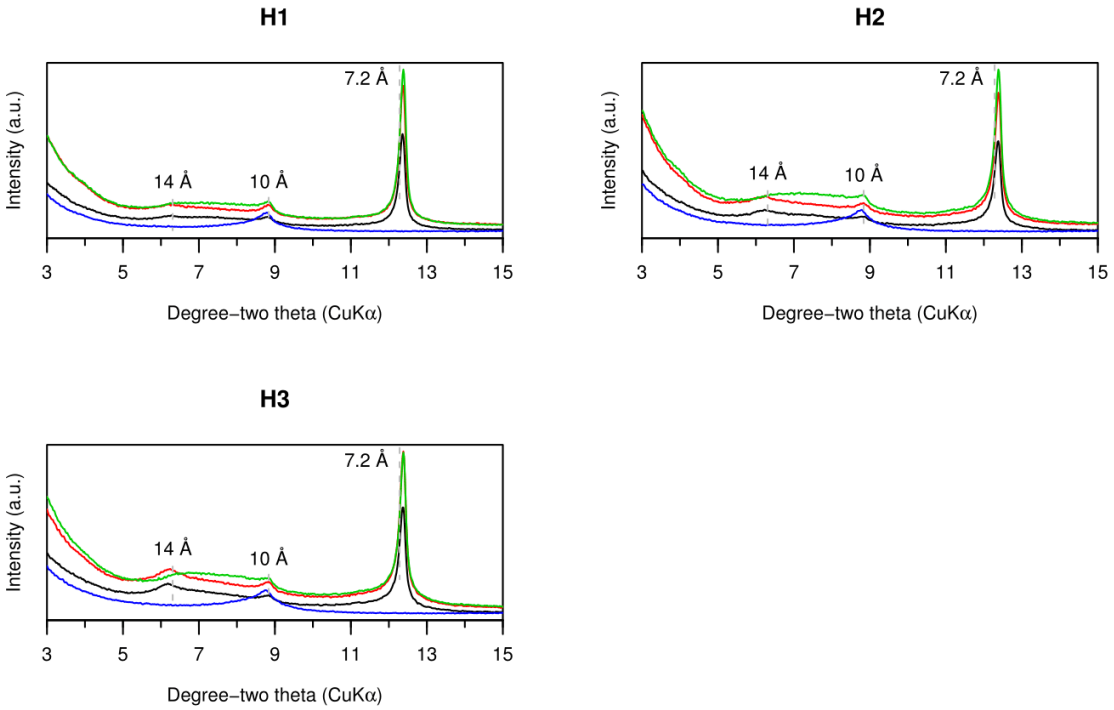


Figure 32 Oriented XRD scan following various pretreatments for Site 13.

Appendix

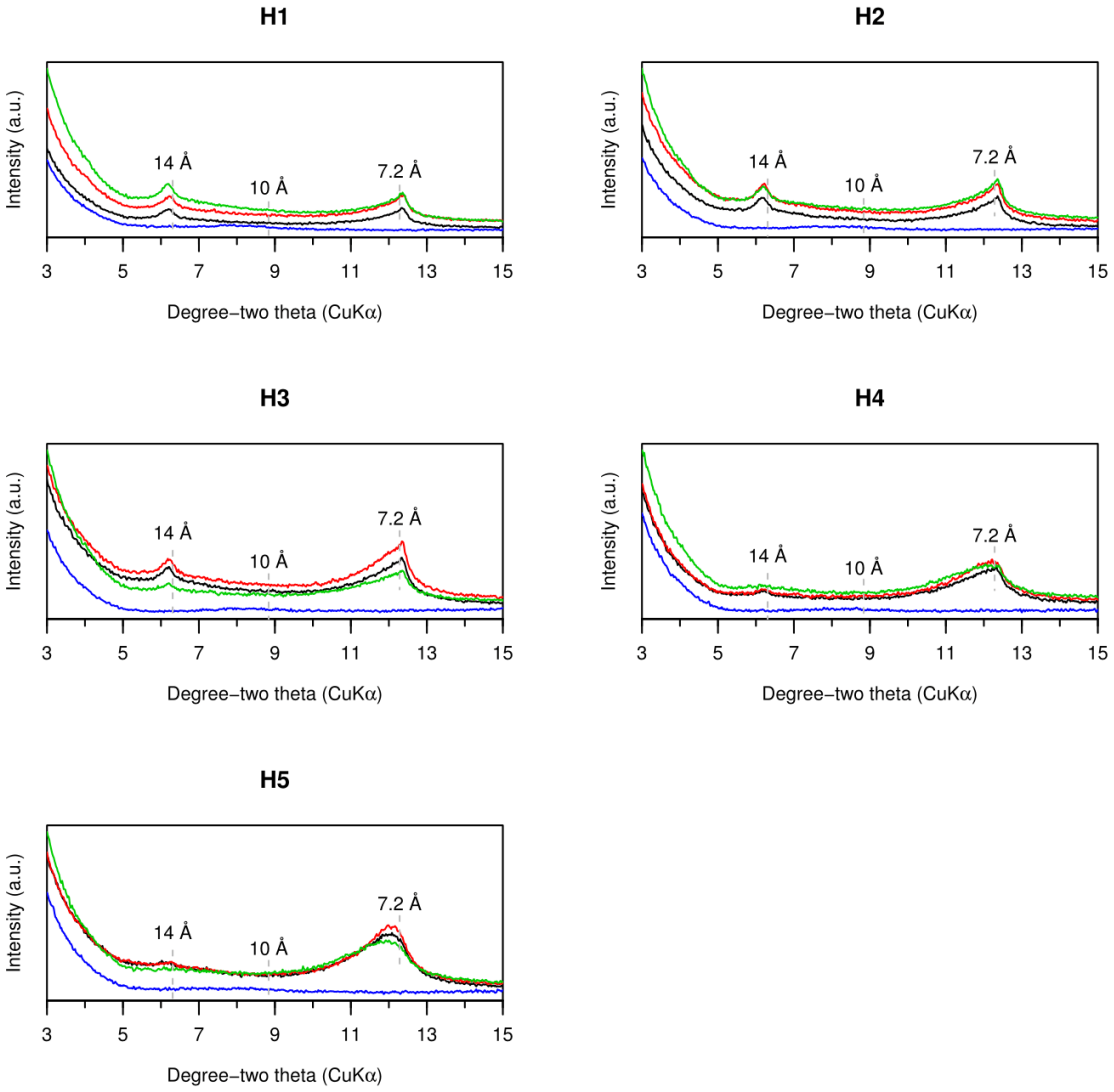


Figure 33 Oriented XRD scan following various pretreatments for Site 14.

Appendix

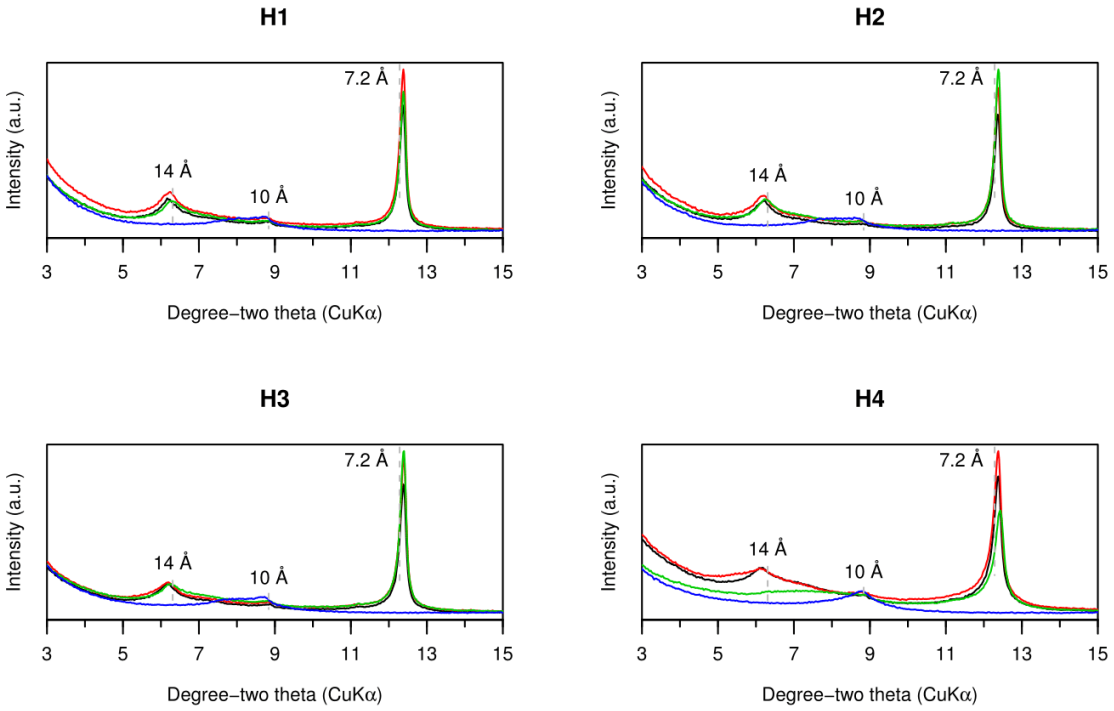


Figure 34 Oriented XRD scan following various pretreatments for Site 15.

B3 XRD semi-quantitative mineral composition

Table 1 XRD semi-quantitative mineral composition of all fifteen soil profiles. Auxiliary minerals are also indicated

Site	Hor	Ka	Sm	Il	He	Go	Ca	Gy	Q	Auxiliary minerals
Site 1	H1	tr	xxx	tr	-	-	tr	-	xx	vermiculite (tr)
	H2	tr	xxx	-	-	-	x	-	xx	anorthite (tr), vermiculite (tr)
	H3	tr	xx	-	-	-	xx	-	xx	anorthite (tr), vermiculite (tr)
	H4	tr	xx	-	-	-	xxxx	-	x	vermiculite (tr)
Site 2	H1	x	tr	tr	-	-	-	-	xxxx	albite (tr)
	H2	xxx	x	tr	tr	-	-	-	xx	
	H3	xxx	x	tr	tr	-	-	-	xx	
	H4	xxx	x	tr	-	-	-	-	xx	albite (tr)
Site 3	H0	xx	-	-	-	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)
	H1	xx	-	-	tr	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)
	H2	xxx	-	-	tr	tr	-	-	xxx	Ti-oxide (tr), HILV (tr)
	H3	xxx	-	tr	tr	x	-	-	xx	Ti-oxide (tr), HILV (tr)
	H4	xxxx	-	tr	tr	x	-	-	xx	Ti-oxide (tr), HILV (tr)
Site 4	H1	tr	tr	tr	-	-	-	-	xxxxx	
	H2	tr	-	tr	-	-	-	-	xxxxx	
	H3	tr	-	tr	-	-	-	-	xxxxx	K-feldspar (tr)
	H4	tr	-	tr	-	-	-	-	xxxxx	
	H5	xxxx	-	tr	tr	tr	-	-	xx	Ti-oxide (tr)
	H6	xxxx	-	tr	tr	tr	-	-	xx	Ti-oxide (tr)
Site 5	H1	x	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr)
	H2	x	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr)
	H3	x	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)
	H4	x	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)

- not present, tr, trace <5%, x 5-20%, xx 20-40%, xxx 40-60%, xxxx 60-80%, xxxxx 80-100%. *Hydroxy-interlayered vermiculite (HILV).

Appendix

Site	Hor	Ka	Sm	Il	He	Go	Ca	Gy	Q	Auxiliary minerals
Site 6	H1	xx	-	tr	-	tr	-	-	xxxx	albite (tr), Ti-oxide (tr), HILV (tr)
	H2	xx	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)
	H3	xx	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)
	H4	xx	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)
	H5	xx	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)
Site 7	H1	x	tr	x	-	-	-	-	xxxx	Ti-oxide (tr)
	H2	xx	x	xx	-	-	-	-	xx	
	H3	xx	x	xx	-	-	xx	-	xx	
	H4	xx	x	xx	-	-	xx	-	xx	
Site 8	H1	x	xxx	x	-	-	-	-	xx	K-feldspar (tr), albite (tr)
	H2	x	xxx	x	-	-	-	-	xx	albite (tr)
	H3	x	xxx	x	-	-	tr	-	xx	albite (tr)
	H4	x	xxx	x	-	-	tr	-	xx	albite (x)
Site 9	H1	x	-	x	tr	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)
	H2	xx	-	x	tr	tr	-	-	xxx	Ti-oxide (tr), HILV (x)
	H3	xxx	-	x	tr	tr	-	-	xxx	HILV (x)
	H4	xxx	-	x	tr	tr	-	-	xx	HILV (x)
Site 10	H1	x	x	tr	tr	-	-	-	xxx	Ti-oxide (tr)
	H2	x	x	tr	tr	-	-	-	xxx	Ti-oxide (tr)
	H3	x	xx	tr	tr	-	tr	-	xx	Ti-oxide (tr)
	H4	x	xxx	-	tr	-	xx	-	xx	Ti-oxide (tr)
	H5	x	xxx	-	tr	-	tr	-	xx	Ti-oxide (tr)
Site 11	H1	tr	xxxx	-	tr	tr	tr	-	x	Ti-oxide (tr)
	H2	tr	xxxx	-	tr	-	tr	-	x	Ti-oxide (tr)
	H3	tr	xxxx	-	tr	-	tr	-	x	Ti-oxide (tr)
	H4	tr	xxxx	-	tr	-	x	-	x	albite (x), Ti-oxide (tr)

- not present, tr, trace <5%, x 5-20%, xx 20-40%, xxx 40-60%, xxxx 60-80%, xxxxx 80-100%. *Hydroxy-interlayered vermiculite (HILV).

Appendix

Site	Hor	Ka	Sm	Il	He	Go	Ca	Gy	Q	Auxiliary minerals
Site 12	H1	x	xxxx	-	-	-	-	-	x	
	H2	x	xxxx	-	-	-	tr	-	x	
	H3	x	xxxx	-	-	-	tr	-	x	
	H4	x	xxxx	-	-	-	tr	x	x	
Site 13	H1	xx	-	x	tr	-	-	-	xxxx	Ti-oxide (tr), HILV (tr)
	H2	xxx	-	x	tr	-	-	-	xxx	Ti-oxide (tr), HILV (tr)
	H3	xxx	-	x	tr	-	-	-	xxx	Ti-oxide (tr), HILV (tr)
Site 14	H1	x	-	-	xx	xx	-	-	x	Ti-oxide (tr), gibbsite (x)
	H2	xx	-	-	xx	xx	-	-	x	Ti-oxide (tr), gibbsite (x)
	H3	xx	-	-	xx	xx	-	-	x	Ti-oxide (tr), gibbsite (tr)
	H4	xx	-	-	xx	xx	-	-	tr	Ti-oxide (tr), gibbsite (tr)
	H5	xx	-	-	xx	xx	-	-	tr	Ti-oxide (tr), gibbsite (tr)
Site 15	H1	x	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr)
	H2	x	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr)
	H3	x	-	tr	-	tr	-	-	xxxx	Ti-oxide (tr), HILV (tr)
	H4	xx	-	tr	-	x	-	-	xx	Ti-oxide (tr), HILV (tr)

- not present, tr, trace <5%, x 5-20%, xx 20-40%, xxx 40-60%, xxxx 60-80%, xxxxx 80-100%. *Hydroxy-interlayered vermiculite (HILV).

B4 Fusion model predictions

Fusion model predictions for all fifteen sites are presented. A legend is provided indicating the prediction of: Ka –kaolinite; Sm – smectite; Il – illite; He – hæmatite; Go – goethite; Ca - $CaCO_3$; Gy – gypsum; and Q - quartz. Horizontal black lines indicate field observed horizons.

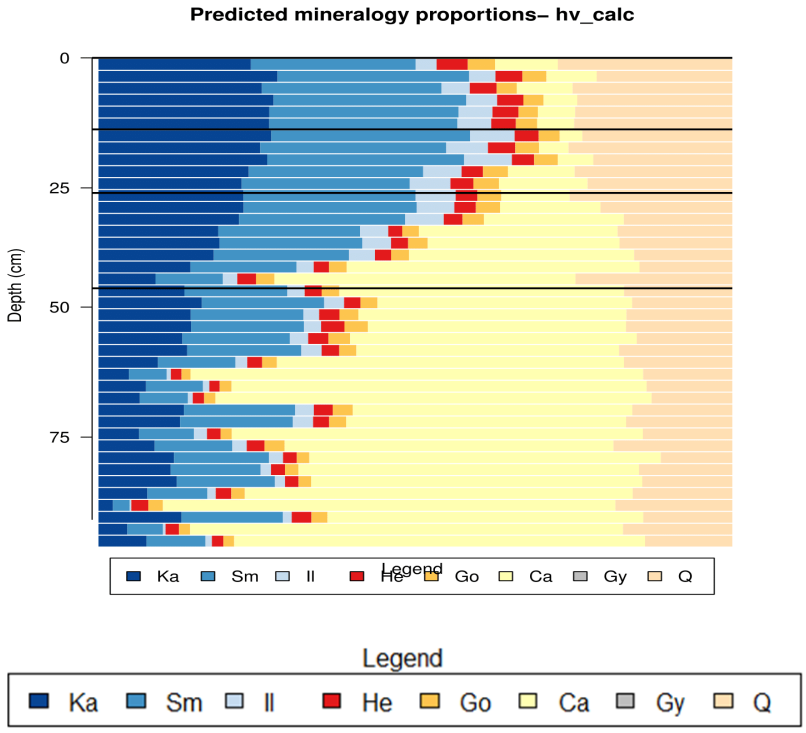


Figure 35 Fusion model predictions for Site 1.

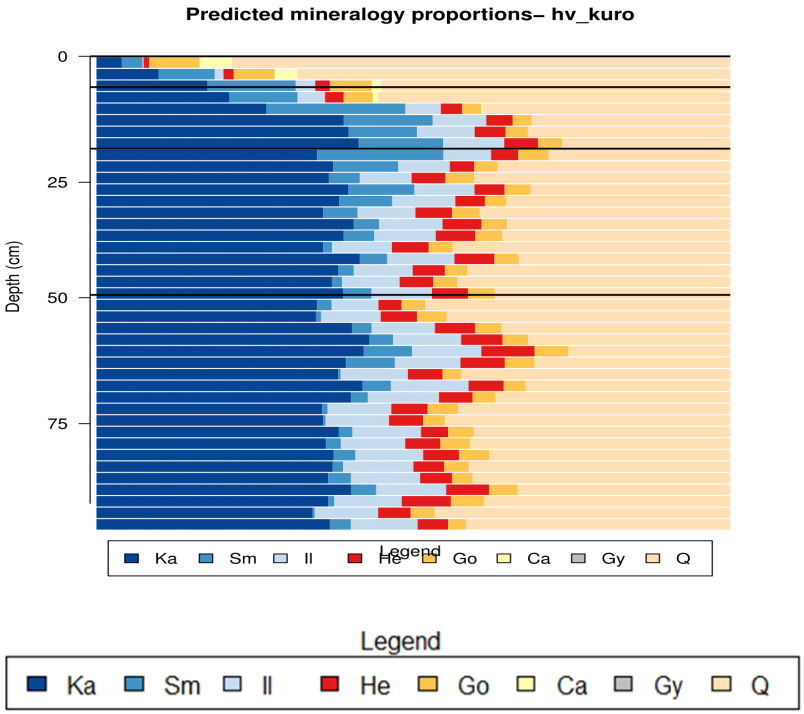


Figure 36 Fusion model predictions for Site 2.

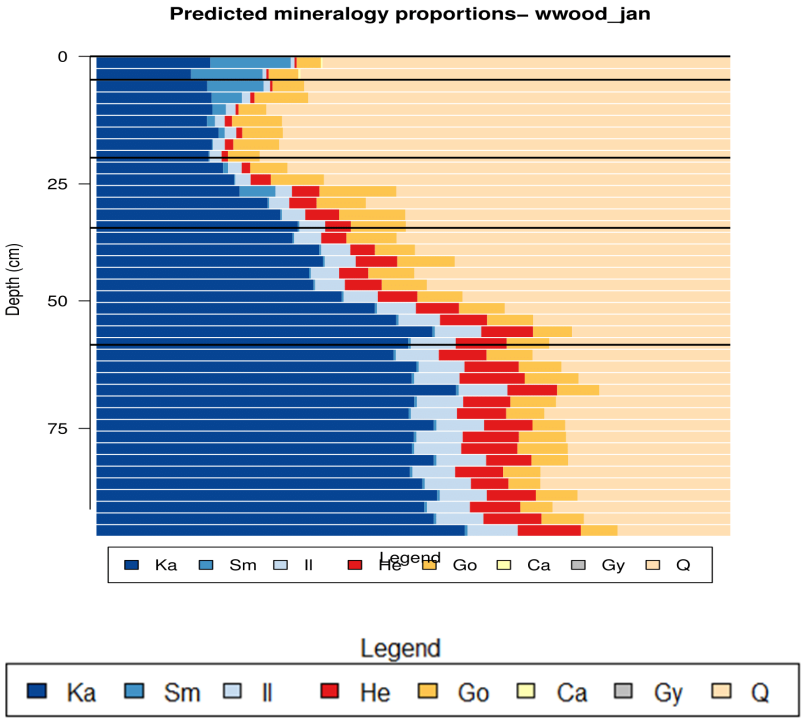


Figure 37 Fusion model predictions for Site 3.

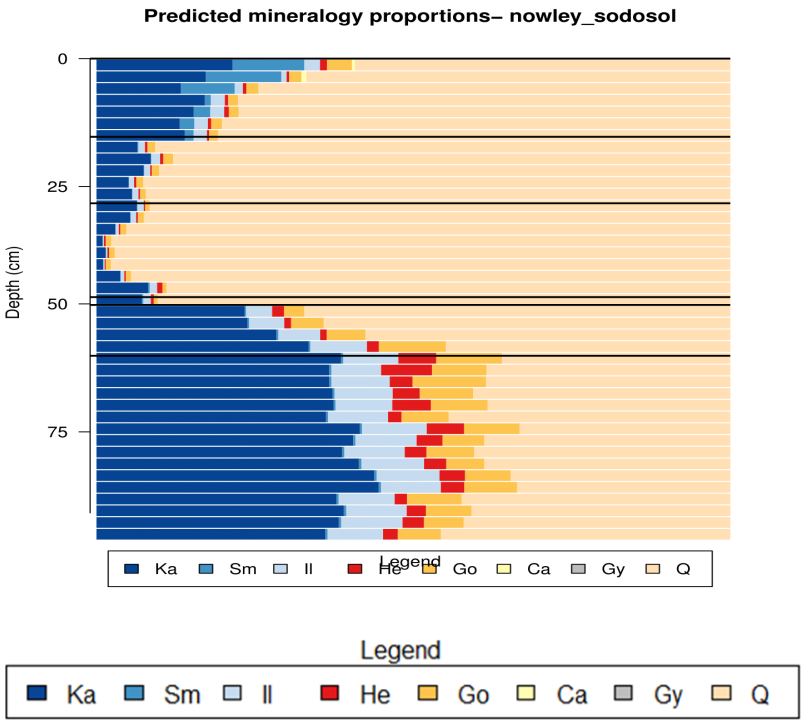


Figure 38 Fusion model predictions for Site 4.

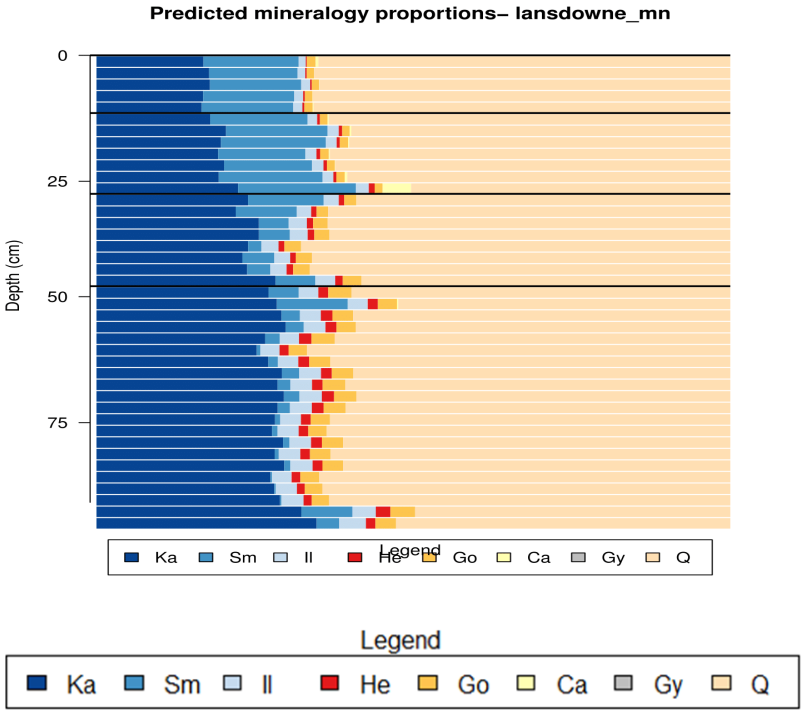


Figure 39 Fusion model predictions for Site 5.

Appendix

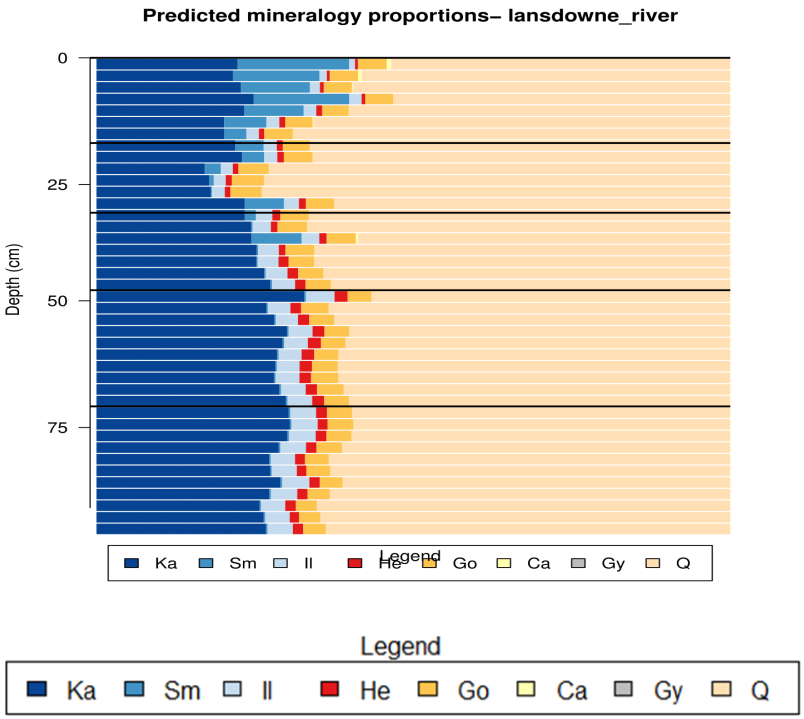


Figure 40 Fusion model predictions for Site 6.

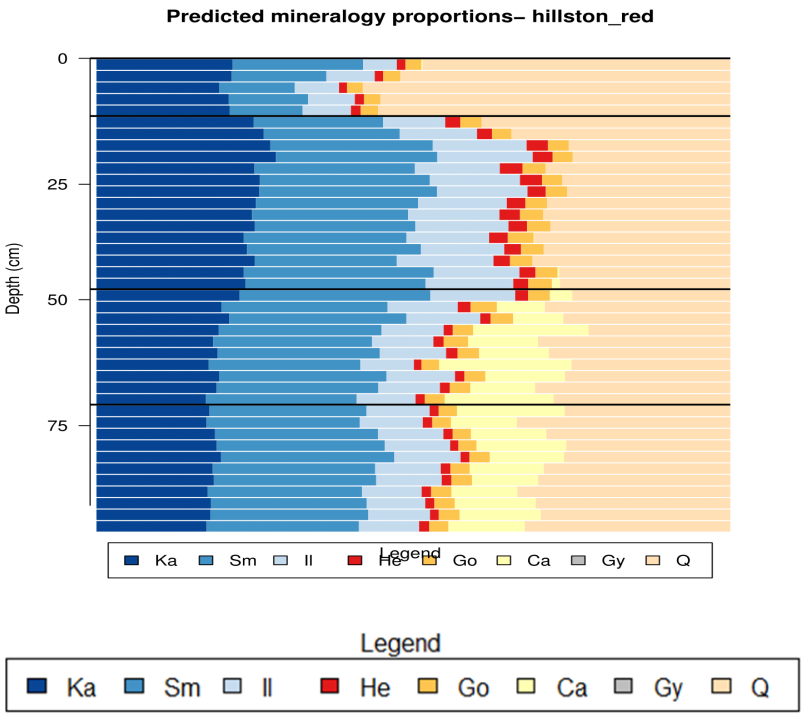


Figure 41 Fusion model predictions for Site 7.

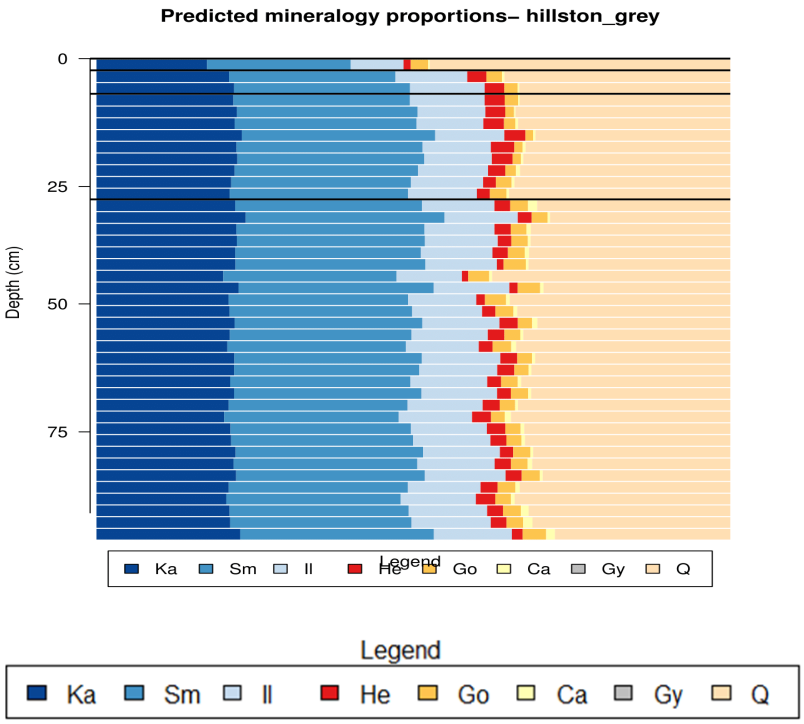


Figure 42 Fusion model predictions for 8.

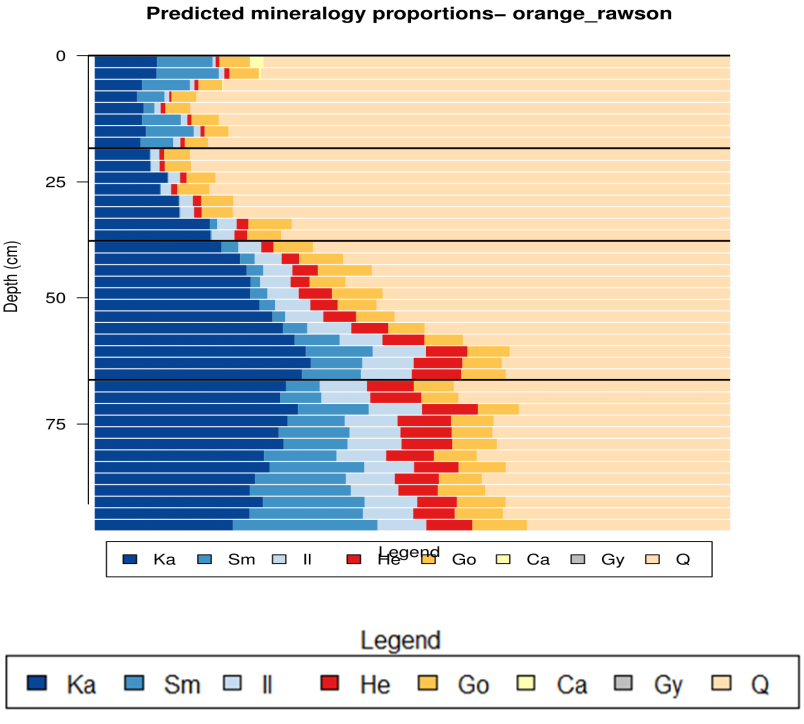


Figure 43 Fusion model predictions for Site 9.

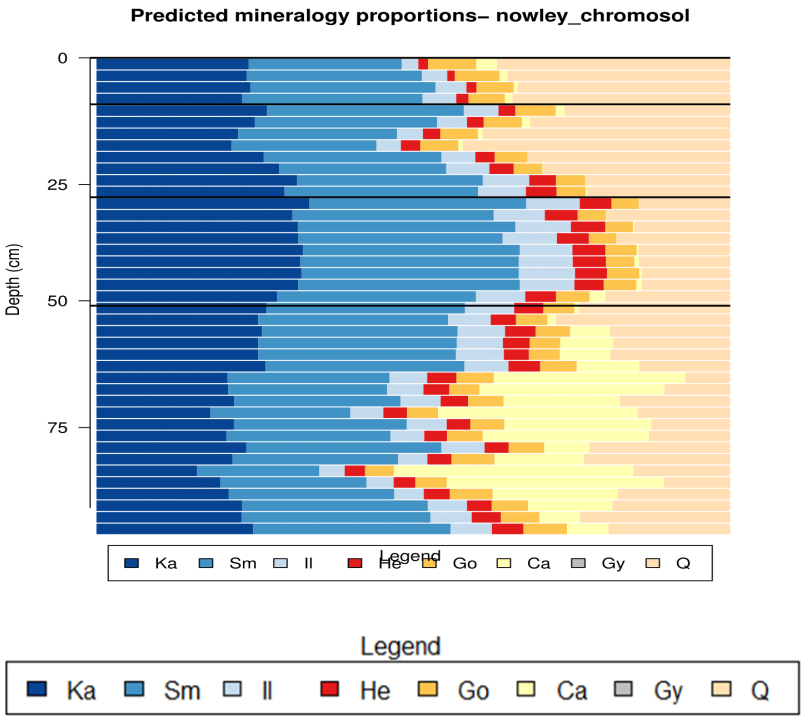


Figure 44 Fusion model predictions for Site 10.

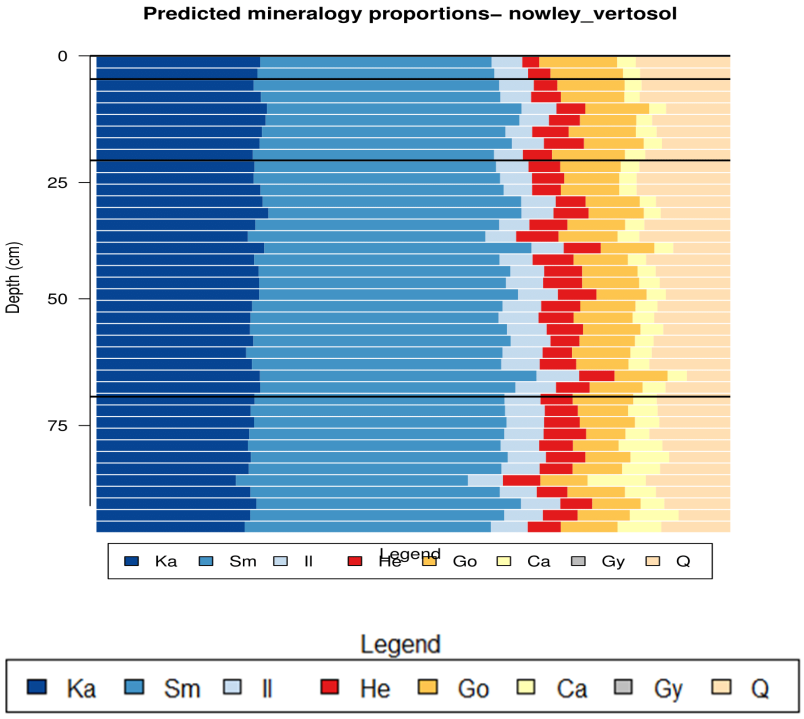


Figure 45 Fusion model predictions for Site 11.

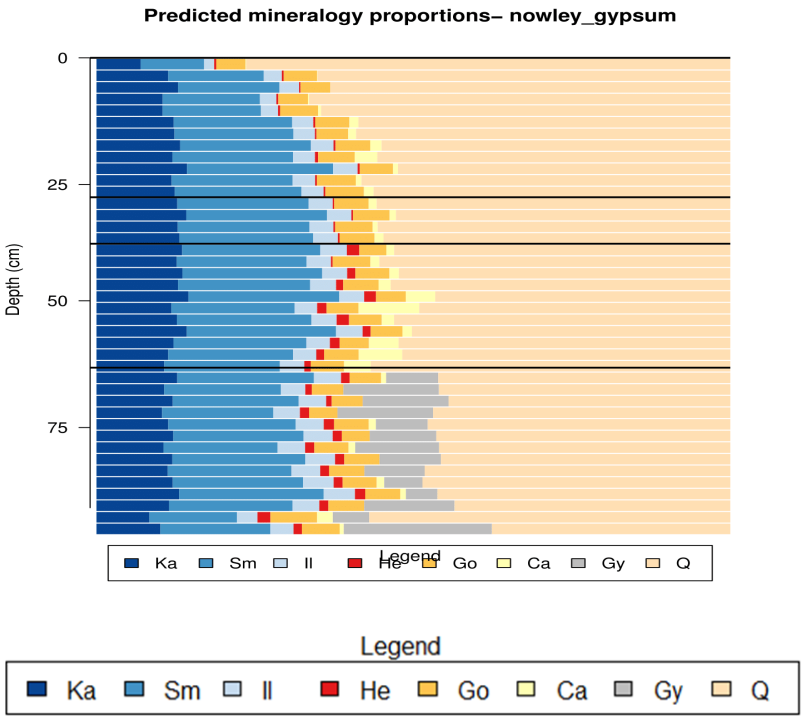


Figure 46 Fusion model predictions for 12.

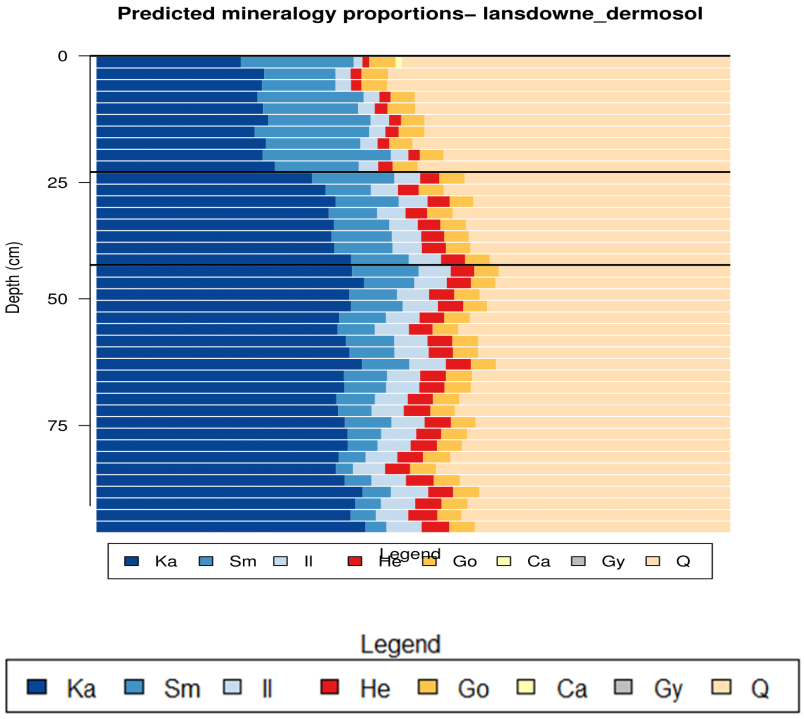


Figure 47 Fusion model predictions for Site 13.

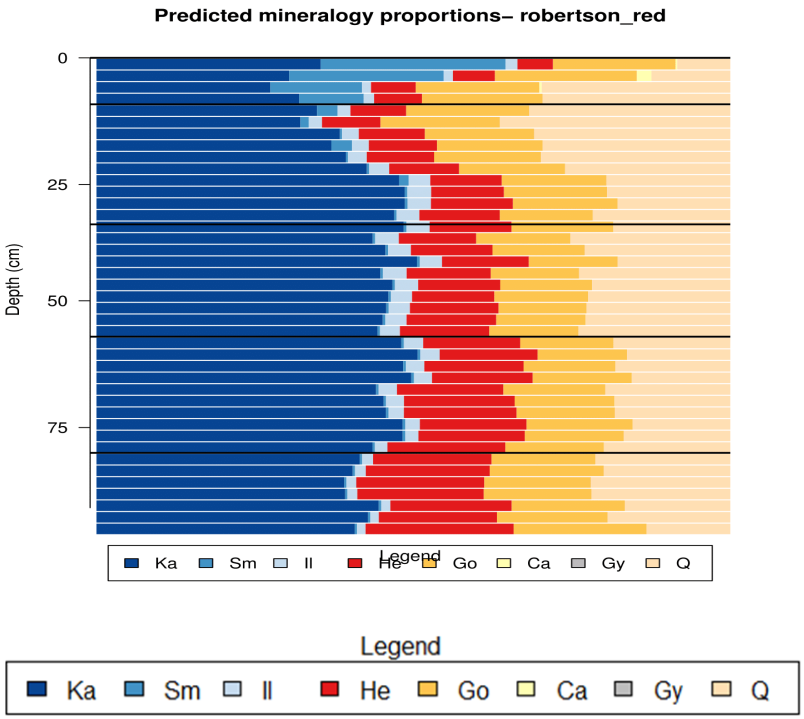


Figure 48 Fusion model predictions for Site 14.

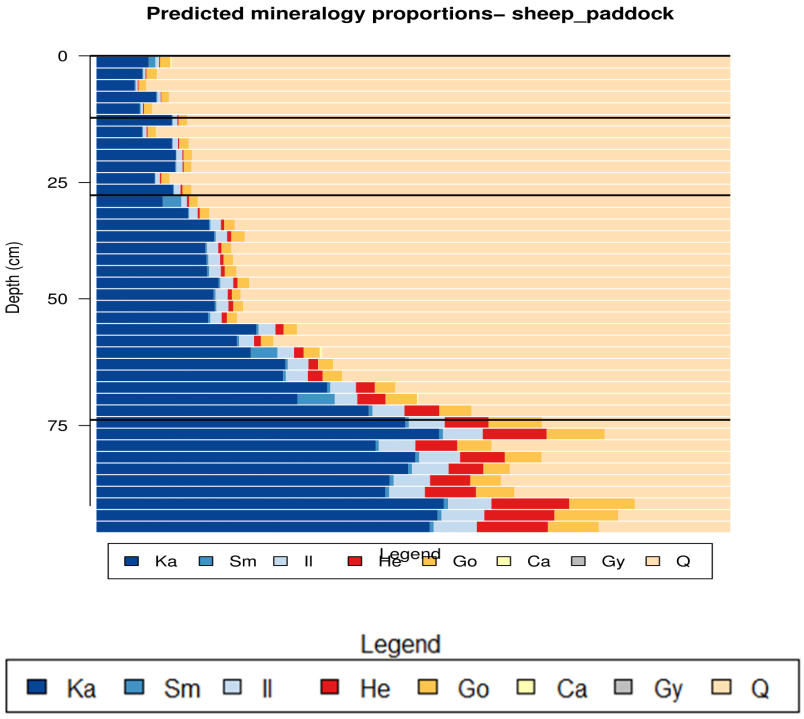


Figure 49 Fusion model predictions for Site 15.

Appendix C

C1 Unique property codes and units

awc – plant available water capacity ($cm^3\ cm^{-3}$)

bd – bulk density ($g\ cm^{-3}$)

buffering_cap – pH buffering capacity ($cmol\ H^+\ kg^{-1}\ pH\ unit^{-1}$)

cec – cation exchange capacity ($cmol_{(+)}\ kg^{-1}$)

clay_2 – clay content (<2 μm) ($g\ 100\ g^{-1}$)

cole – coefficient of linear extensibility

ec_1_5 – electrical conductivity (1:5 H_2O) ($dS\ m^{-1}$)

ece – electrical conductivity of saturated soil extract ($dS\ m^{-1}$)

k_sat – saturated hydraulic conductivity ($mm\ hr^{-1}$)

linear_shrinkage – linear shrinkage ($mm\ 100\ mm^{-1}$)

lower_depth – sample lower depth (cm)

mid_depth – sample mid-depth (cm)

om – organic matter ($g\ 100\ g^{-1}$)

ph_h2o – pH (1:5 H_2O)

ph_cacl2 – pH (1:5 $CaCl_2$)

sand_20_2000 – sand content (20 – 2,000 μm) ($g\ 100\ g^{-1}$)

sand_50_2000 – sand content (50 – 2,000 μm) ($g\ 100\ g^{-1}$)

silt_2_20 – sand content (2 - 20 μm) ($g\ 100\ g^{-1}$)

silt_2_50 – sand content (2 – 50 μm) ($g\ 100\ g^{-1}$)

smc_alpha – van Genuchten model parameter α

smc_m – van Genuchten model parameter *m*. ($m = 1-1/n$)

smc_n – van Genuchten model parameter *n*

smc_theta_r – van Genuchten model parameter θ_r

smc_theta_s – van Genuchten model parameter θ_s

Appendix

theta_cll – moisture content at crop lower limit ($cm^3\ cm^{-3}$)

theta_dul – moisture content at drainage upper limit ($cm^3\ cm^{-3}$)

theta_fc10 – moisture content at field capacity, matric potential -10 kPa ($cm^3\ cm^{-3}$)

theta_fc33– moisture content at field capacity, matric potential -33 kPa ($cm^3\ cm^{-3}$)

theta_pwp – moisture content at permanent wilting point ($cm^3\ cm^{-3}$)

theta_sp – moisture content at sticky point ($cm^3\ cm^{-3}$)

tot_oc – organic carbon ($g\ 100\ g^{-1}$)

upper_depth – sample upper depth (cm)

C2 List of PTFs and other calculations used in SPEC-SINFERS

List of PTFs in R language. Provenace is a from internal PTF construction and the published literature. Some of the equations are simple calculations and not PTFs *sensu stricto*, such as the calculation of organic matter from organic carbon (PTF_35). The PTFs are listed in the order in which they were created. The PTF number has no influence on the construction of the SPEC-SINFERS network. An example of a function containing an if-then-else statement is shown in PTF_39, separate PTFs are applied depending whether the mid-depth of the sample is above or below 20 cm.

PTF_1

theta_dul = 0.2739 + 0.005033 * clay_2 + 3.158 * 10^-5 * sand_20_2000 * cec - 1.96 * 10^-5 * sand_20_2000 ^2 - 0.00256 * clay_2 * bd

Internal PTF constructed from APSRU database

PTF_2

theta_dul = 0.2358 + 0.002572 * cec + 0.001001 * clay_2 - 1.70 * 10^-7 * sand_20_2000 ^3

Internal PTF constructed from APSRU database

PTF_3

theta_dul = 0.374 + 0.01182 * bd + 0.00365 * clay_2 + 6.09 * 10^-5 * sand_20_2000 * clay_2 - 0.00339 * sand_20_2000 - 0.00192 * bd ^2 * clay_2

Internal PTF constructed from APSRU database

PTF_4

theta_dul = 0.2082 + 0.02757 * tot_oc + 0.002666 * clay_2 - 1.73 * 10^-7 * sand_20_2000 ^3

Internal PTF constructed from APSRU database

PTF_5

theta_dul = 0.364 + 4.828 * 10^-5 * sand_20_2000 * clay_2 - 0.00296 * sand_20_2000

Internal PTF constructed from APSRU database

PTF_6

theta_fc10 = 0.5255 - 2.76 * 10^-5 * sand_20_2000 ^2 - 0.05195 * bd ^2

Internal PTF constructed from database of Australian hydraulic properties

Appendix

PTF_7

theta_fc10 = 0.4795 - 3.873 * 10^-5 * sand_20_2000 ^2 - 6.701 * 10^-7 * clay_2 ^2 * sand_20_2000

Internal PTF constructed from database of Australian hydraulic properties

PTF_8

theta_cll = 0.1476 + 9.002 * 10^-5 * clay_2 ^2 - 0.00115 * sand_20_2000 - 9.752 * 10^-7 * clay_2 ^3

Internal PTF constructed from APSRU database

PTF_9

theta_cll = 0.6151 * theta_dul - 0.02192

Internal PTF constructed from APSRU database

PTF_10

theta_pwp = 0.1766 + 0.00255 * clay_2 - 0.001487 * sand_20_2000

Internal PTF constructed from database of Australian hydraulic properties

PTF_11

theta_pwp = (0.814 * theta_fc10 - 0.07996)

Internal PTF constructed from database of Australian hydraulic properties

PTF_12

theta_fc10 = (32.35 + 43.64 * tanh(0.5 * (5.34 - 0.02178 * clay_2 + 2.12 * tot_oc - 0.4192 * ph_h2o)) + 81.23 * tanh(0.5 * (-4.581 + 0.02389 * clay_2 + 0.06029 * tot_oc + 0.30392 * ph_h2o)))/100

Internal PTF constructed from database of Australian hydraulic properties

PTF_13

theta_fc33 = 0.003320110612731 - 0.3396 * tanh(0.5 * (-0.9705 - 0.8529 * bd - 0.00827 * clay_2 + 0.01994 * sand_20_2000)) + 0.1629 * tanh(0.5 * (3.71 - 3.19 * bd + 0.01205 * clay_2 + 0.01617 * sand_20_2000)) - 0.1272 * tanh(0.5 * (-3.94 - 0.5067 * bd + 0.02158 * clay_2 + 0.04978 * sand_20_2000))

Internal PTF constructed from database of Australian hydraulic properties

Appendix

PTF_14

theta_pwp = -0.1554 - 0.7221 * tanh(0.5 * (-0.9705 - 0.8529 * bd - 0.00827 * clay_2 + 0.01994 * sand_20_2000)) + 0.1325 * tanh(0.5 * (3.71 - 3.19 * bd + 0.01205 * clay_2 + 0.01617 * sand_20_2000)) + 0.1720 * tanh(0.5 * (-3.94 - 0.5067 * bd + 0.02158 * clay_2 + 0.04978 * sand_20_2000))

Internal PTF constructed from database of Australian hydraulic properties

PTF_15

smc_theta_r = (0.3697 * tanh(-0.0167 * clay_2 - 0.0259 * sand_20_2000 + 0.5587 * bd + 1.86) - 0.2543 * tanh(-0.0074 * clay_2 - 0.0061 * sand_20_2000 + 0.9869 * bd - 1.47) - 0.2099 * tanh(-0.0653 * clay_2 - 0.0063 * sand_20_2000 - 5.30 * bd + 9.40) - 0.2032)^2

Internal PTF constructed from database of Australian hydraulic properties

PTF_16

smc_theta_s = 0.1958 * tanh(-0.0167 * clay_2 - 0.0259 * sand_20_2000 + 0.5587 * bd + 1.86) - 0.4692 * tanh(-0.0074 * clay_2 - 0.0061 * sand_20_2000 + 0.9869 * bd - 1.47) + 0.0063 * tanh(-0.0653 * clay_2 - 0.0063 * sand_20_2000 - 5.30 * bd + 9.40) + 0.0495

Internal PTF constructed from database of Australian hydraulic properties

PTF_17

smc_alpha = exp(-2.07 * tanh(-0.0167 * clay_2 - 0.0259 * sand_20_2000 + 0.5587 * bd + 1.86) - 1.123 * tanh(-0.0074 * clay_2 - 0.0061 * sand_20_2000 + 0.9869 * bd - 1.47) - 0.3819 *tanh(-0.0653 * clay_2 - 0.0063 * sand_20_2000 - 5.30 * bd + 9.40) - 2.57)

Internal PTF constructed from database of Australian hydraulic properties

PTF_18

smc_n = exp((-1.33 * tanh((-0.0167 * clay_2 - 0.0259 * sand_20_2000) + 0.5587 * bd + 1.86) + 1.58 * tanh(((-0.0074 * clay_2 - 0.0061 * sand_20_2000) + 0.9869 * bd) - 1.47) + 0.2151 * tanh((-0.0653 * clay_2 - 0.0063 * sand_20_2000 - 5.30 * bd) + 9.40)) - 0.016) + 1

Internal PTF constructed from database of Australian hydraulic properties

PTF_19

linear_shrinkage = 8.23 + 18.49 * tanh(0.5 * (-0.8995 + 0.03462 * clay_2 - 0.00191 * sand_20_2000))

Internal PTF constructed from the National Soil Collection Database (TERN)

Appendix

PTF_20

cole = 0.06392 - 0.03560 * tanh(0.5 * (5.54 - 0.1381 * clay_2 + 0.01538 * sand_20_2000))

Internal PTF constructed from the National Soil Collection Database (TERN)

PTF_22

k_sat = exp(2.41 - 8.12 * tanh(0.5 * (-3.96 + 2.86 * theta_fc10 + 1.90 * bd)) - 3.67 * tanh(0.5 * (-14.40 + 20.90 * theta_fc10 + 3.68 * bd)))

Internal PTF constructed from database of Australian hydraulic properties

PTF_35

om = 1.72 * tot_oc

Nelson, D.W. and Sommers, L.E. 1982. "Total carbon, organic carbon and organic matter". In *Methods of Soil Analysis*, Part 2. (ed. A.L. Page) pp. 539-79. American Society of Agronomy, Madison, WI.

PTF_36

bd = 100 /(om/0.223 + (100 - om) / (1.35128606477631 + 0.00451974677070142 * sand_20_2000 + (sand_20_2000 - 44.652494600432) * ((sand_20_2000 - 44.652494600432) * -0.0000613723924995459) + 0.0596420803366252 * log(mid_depth)))

Tranter, G., Minasny, B., McBratney, A. B., Murphy, B., McKenzie, N. J., Grundy, M., Brough, D. (2007). Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use and Management 23(4), 437-443.

PTF_37

mid_depth = (upper_depth + lower_depth)/2

Simple calculation

PTF_38

silt_2_20 = 100 - sand_20_2000 - clay_2

Simple calculation

PTF_39

buffering_cap = if(mid_depth[1]<20){(0.955 * tot_oc + 0.011 * clay_2) * 1.2}else{(12.79 - 0.19 * clay_2 - 0.7 * tot_oc - 0.03 * silt_2_20 + 0.74 * silt_2_20 * tot_oc) * 0.06}

Noble, A., Cannon, M., Muller, D. (1997). Evidence of accelerated soil acidification under Stylosanthes-dominated pastures. Soil Research 35(6), 1309-1322.

Appendix

PTF_40

awc = theta_fc10 - theta_pwp

Simple calculation

PTF_41

theta_sp = -2.659e-01 + 6.362e-02 * clay_2 - 1.883e-03 * clay_2 ^2 + 1.914e-05 * clay_2 ^3

Slavich, P., Petterson, G. (1993). Estimating the electrical conductivity of saturated paste extracts from 1: 5 soil, water suspensions and texture. Soil Research 31(1), 73-81.

PTF_42

ece = ec_1_5 * (2.46 + 3.03/ theta_sp)

Slavich, P., Petterson, G. (1993). Estimating the electrical conductivity of saturated paste extracts from 1: 5 soil, water suspensions and texture. Soil Research 31(1), 73-81.

PTF_43

ph_cacl2 = -0.05 + 0.9 * ph_h2o + 0.14 * log(ec_1_5)

Minasny, B., McBratney, A., Brough, D., Jacquier, D. (2011). Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. European Journal of Soil Science 62(5), 728-732.

PTF_44

silt_2_50 = 2.26 * silt_2_20 + (5.55 * silt_2_20 + 1.513 * (silt_2_20)^2)/(0.9966 - 1.236 * silt_2_20 - 1.349 * sand_20_2000)

Padarian, J., Minasny, B., McBratney, A. (2012). Using genetic programming to transform from Australian to USDA/FAO soil particle-size classification system. Soil Research 50(6), 443-446.

PTF_45

sand_50_2000 = sand_20_2000 - (silt_2_50 - silt_2_20)

Simple calculation

PTF_101

smc_m = 1 - (1/ smc_n)

Simple calculation