

School of Molecular and Life Sciences

**Spectroscopy with Chemometrics for the
Forensic Analysis of Sandy Soils**

**Talia Grace Newland
0000-0003-3231-6686**

**This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University**

June 2023

Declaration

To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for any other degree or diploma in any university.

Signature:

Date: 16th June 2023

Acknowledgement of Country

We acknowledge that Curtin University works across hundreds of traditional lands and custodial groups in Australia, and with First Nations people around the globe. We wish to pay our deepest respects to their ancestors and members of their communities, past, present, and to their emerging leaders. Our passion and commitment to work with all Australians and peoples from across the world, including our First Nations peoples are at the core of the work we do, reflective of our institutions' values and commitment to our role as leaders in the Reconciliation space in Australia.

Abstract

Soil is a valuable source of forensic trace evidence due to its transferability and extensive heterogeneity across different locations, however, it is often underutilised within criminal investigations due to challenges associated with its analysis and interpretation. While there are many different types of soils encountered around the world, no single method is suitable for the forensic analysis of all soils. There is also minimal research exploring the analysis of specifically arid, sandy soils and their role as geological evidence in a forensic context, like those found on the Swan Coastal Plain in Perth, Western Australia. Analysis of these soils is especially challenging due to their quartz-dominated nature and very low levels of clay and organic matter, and this is made further difficult by the trace quantities of soil commonly encountered in forensic casework. It is therefore important to have reliable, validated methods for the forensic analysis of these soils, as well as for the interpretation and communication of the results generated. This thesis explores a multi-faceted approach that utilises spectroscopic techniques in combination with chemometrics to demonstrate objective characterisation and differentiation of arid, sandy soil samples for forensic purposes. The findings educate on the potential value of sandy soils when used for site discrimination in a forensic context, in order to allow for the optimised use of soil evidence in forensic laboratories in locations that experience soils like these.

Microspectrophotometry (MSP) was performed on the quartz-recovered fine fractions of soils to provide an objective colour measurement, and this data was subsequently analysed using principal component analysis (PCA) and linear discriminant analysis (LDA) for the discrimination of soils based on their locations. Only a few soils were able to be differentiated due to noisy, irreproducible reflectance spectra, and most of these soils were distinctly coloured. MSP spectra were also converted into $L^* a^* b^*$ colorimetric values prior to chemometric analysis, however, this method was unable to provide any beneficial separation over visual examination of the soil. Hence, MSP analysis was recommended for use in forensic

pairwise comparisons to detect subtle differences in colour, or to provide a statistical measure of differentiation for highly coloured soils.

Vibrational spectroscopy was used to chemically characterise the quartz-recovered fine fractions of the same soils. Initially, Raman spectroscopy was performed on these soil samples. However, it was unsuccessful at providing any chemical information due to fluorescence swamping of the spectra. Background corrections showed no beneficial effect. Attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy was performed in combination with PCA and LDA, and successfully illustrated many inorganic and organic components within the soils that allowed for their differentiation. This resulted in the full discrimination of all soil samples based on their original locations, and the classification of 93.2% of samples to their correct locations. The remaining incorrectly classified samples displayed variability between duplicate samples, indicating sampling inconsistencies or high levels of heterogeneity within these sites, however, this could potentially be improved through the inclusion of additional principal components (PCs) when conducting LDA. Overall, ATR-FTIR analysis showed great potential for further differentiation of visually similar sandy soils compared to MSP.

The quartz-recovered fine fractions of the same soils were analysed using X-ray diffraction (XRD). When combined with PCA, data allowed for the discrimination of most soil samples based on their locations, however, some sites showed significant intra-site variability. Due to limitations on the number of samples, LDA predictive models were not able to accurately classify soils to their original locations, despite the degree of discrimination achieved. However, the XRD data showed improved discrimination of some soils that were more challenging to differentiate using the ATR-FTIR data, highlighting the potential of a multi-model, sequential approach. A reproducibility study was also conducted by comparing results obtained through the use of three different XRD instruments that varied in age and capabilities. This preliminary research showed variations in the quality of data did affect the results, however, all instrumentation achieved similar overall discrimination of soils and predictive accuracies, demonstrating robustness of the method.

A combined analytical sequence was then proposed and applied to a blinded case simulation to showcase its suitability for use within the context of forensic casework. Using the combined spectroscopic and chemometric sequence, an unknown “recovered” sample was able to be associated with a single reference soil and discriminated from the other soils. The multiple different techniques employed produced varying results, however, these results offered complementary information that was able to be interpreted collectively to discriminate between the reference soils and identify the source of the “recovered” soil with greater confidence. The use of chemometric methods to interpret the data provided an objective, statistical measure of the similarities and/or variation between samples that could be presented as weighting for evidence in court. Hence, this approach was deemed suitable for incorporation into forensic casework to provide discrimination between similar sandy soil samples in an objective manner.

Acknowledgements

I would like to take this opportunity to thank the people who have helped and supported me throughout this journey.

Firstly, thank you to my supervisory team. Simon, your encouragement and guidance over the last decade has been so appreciated. You have kept me on track with your constant check-ins, and your sense of humour helped soften the blows of any challenges faced. My time at university was made much more amusing with you calling me out mid-lecture recording for not attending your classes, and faking a heart attack to hide your shame of losing to me in laser tag. Kari, thank you for driving me all around Perth to collect endless amounts of soil, and devoting your weekends and nights to editing my work despite already being incredibly busy. Thank you for imparting all your knowledge about dirt onto me – I have learned so much from you and am extremely grateful. Thanks also goes out to ChemCentre for allowing me to use their laboratories and resources, with an extra special mention to Ryan Fillingham, for warming up the MSP for me before I arrived, troubleshooting any issues encountered, and constantly letting me into the secure access rooms.

To the Curtin Forensic and Analytical Research Group, thank you for making my journey a thousand times more enjoyable, by willingly partaking in blind Coca-Cola tasting competitions and Rhiannon and I's lengthy quiz nights. Thank you to the members who also volunteered to proof-read my thesis and tidy up my ramblings along the way. Rhiannon, I would not have survived the UK without you (literally) – thank you for watching nearly 200 episodes of The Office with me, running around town to find me Pepto Bismol and Hydrolyte when I was too sick to leave bed, and accompanying me to both the hospital and the Spice Girls concert (on the same day). Thank you also to Josh, for being my desk buddy (when I actually came in) and making lunch times way more fun, and to Georgina, for coaching me in everything chemometrics and answering every question I had with minimal eye-rolling.

Thank you to my family for being so supportive and assuring me I could do anything I put my mind to. I also appreciate you feigning both interest in and understanding of my thesis. To all my friends who are surely sick of hearing about my research by now, thank you for listening and pretending not to be. Most importantly, to Ryan – I could not have finished this without you. Thank you for listening to my late-night rambles, providing me constant assurance that I could do this, and making me laugh every day. I promise to be slightly less stressed from now onwards. And finally, to my little puppy Sundae, for consistently lifting my spirits with all your cheekiness and undivided love. I probably would have finished writing much quicker if you weren't trying to shove my laptop off my lap to claim it for yourself, but I love the cuddles too much.

I would also like to acknowledge the support of the Australian Government for providing me with a Research Training Program (RTP) Scholarship, Curtin University for the Curtin Research Scholarship (CRS), and Rowe Scientific for the HDR Stipend Scholarship, all of which helped to fund this research.

Contribution of Others

The following have contributed to the thesis, with contributions being listed in the form of CRediT (Contributor Roles Taxonomy) statements for each chapter – see <https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement>

Chapters 1 – 7:

Professor Simon W. Lewis (Principal Supervisor): Conceptualisation, Project Administration, Supervision, Visualization, Writing – Review & Editing.

Dr Kari Pitts (Associate Supervisor): Conceptualisation, Supervision, Resources, Design & Creation of Blinded Case Simulation, Writing – Review & Editing.

Chapter 1:

Portions of this chapter have been published in *Encyclopedia of Analytical Science (Third Edition)*, see Appendix for details.

Chapters 1 – 5, & 7:

Portions of these chapters have been published in *Forensic Chemistry*, see Appendix for details.

Chapters 1, 2 & 4:

Portions of these chapters have been submitted for publication in *Forensic Science International: Reports*, see Appendix for details.

Chapters 2, 6 & 7:

Portions of these chapters have been published in *Forensic Chemistry*, see Appendix for details.

Publications

This dissertation contains work which has been submitted for publication in the following peer reviewed journals:

K. Pitts, S. Lewis, and **T. Newland**. "Geochemistry | Soil and Mineralogical Analysis." In *Encyclopedia of Analytical Science (Third Edition)*, P. Worsfold, C. Poole, A. Townshend, and M. Miró, 2019. p. 292-301. Oxford: Academic Press. DOI: [10.1016/B978-0-12-409547-2.14201-5](https://doi.org/10.1016/B978-0-12-409547-2.14201-5).

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics for the forensic analysis of Western Australian sandy soils." *Forensic Chemistry*, 2022. 28: 100412. DOI: [10.1016/j.forc.2022.100412](https://doi.org/10.1016/j.forc.2022.100412). (A preprint of this article can be viewed on Chem Rxiv at: [10.26434/chemrxiv-2021-bv3mr-v2](https://chemrxiv.org/abs/10.26434/chemrxiv-2021-bv3mr-v2)).

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics: Application to simulated forensic soil casework." *Forensic Chemistry*, 2023. 33: 100481. DOI: [10.1016/j.forc.2023.100481](https://doi.org/10.1016/j.forc.2023.100481). (A preprint of this article can be viewed on Chem Rxiv at: [10.26434/chemrxiv-2022-00bqj](https://chemrxiv.org/abs/10.26434/chemrxiv-2022-00bqj)).

T. G. Newland, K. Pitts, and S. W. Lewis. "Negative result: Application of Raman spectroscopy to the forensic analysis of an arid, sandy, soil." *Forensic Science International: Reports*, 2022. Submitted for publication.

Conference Presentations

Selected aspects of the work contained within this thesis have been presented at the following conferences:

Oral Presentations

Multimodal spectroscopic analysis of Western Australian sandy soils for forensic purposes. Australian & New Zealand Forensic Science Society Symposium, Brisbane, 2022.

Forensic analysis of the inorganic content of Western Australian soils. 10th Annual Forensic Science Symposium, Florida International University, 2021 (online).

Poster Presentations

Forensic analysis of the inorganic content of Western Australian soils. Australian & New Zealand Forensic Science Society - Western Australian Forensic Science Forum, Perth, 2021.

Forensic analysis of the inorganic content of Western Australian soils. Royal Society of Chemistry - Twitter Conference, 2021 (online).

The application of spectroscopic techniques and chemometrics to the forensic analysis of the inorganic content of Western Australian soils. School of Molecular and Life Sciences – Science Symposium, Curtin University, 2020.

The application of spectroscopic techniques and chemometrics to the forensic analysis of soils. Australian & New Zealand Forensic Science Society Symposium, Perth, 2018.

The application of spectroscopic techniques and chemometrics to the forensic analysis of soils. Royal Society of Chemistry - Twitter Conference, 2018 (online).

Table of Contents

Declaration	i
Acknowledgement of Country	ii
Abstract	iii
Acknowledgements	vi
Contribution of Others	viii
Publications	ix
Conference Presentations	x
Table of Contents	xi
List of Figures	xvi
List of Tables	xxiii
List of Abbreviations	xxvi
Chapter 1. Introduction	1
1.1 <i>Soil as forensic evidence</i>	2
1.2 <i>Soil formation and composition</i>	3
1.2.1 The formation of soils	3
1.2.2 The composition of soil	4
1.2.3 Factors affecting soil composition	6
1.2.4 Composition of forensic soils	7
1.3 <i>Swan Coastal Plain</i>	8
1.3.1 Characteristics of the Swan Coastal Plain	13
1.4 <i>The forensic analysis of soils</i>	14
1.4.1 The forensic approach.....	14
1.4.2 Methods utilised for forensic soil analysis	16
1.4.2.1 Polarised light microscopy.....	19
1.4.2.2 X-ray diffraction	20
1.4.2.3 Inductively coupled plasma mass spectrometry.....	23

1.4.2.4	Scanning electron microscopy with energy dispersive X-ray spectroscopy	23
1.4.2.5	X-ray fluorescence spectroscopy	24
1.4.2.6	Chromatography	25
1.4.2.7	Optical microscopy	25
1.4.2.8	Isotope-ratio mass spectrometry	26
1.4.3	Emerging techniques for forensic soil analysis.....	27
1.4.3.1	Microspectrophotometry	27
1.4.3.2	Raman spectroscopy.....	28
1.4.3.3	Infrared spectroscopy.....	29
1.4.3.4	Microbiome analysis.....	31
1.5	<i>Use of the quartz-recovered fine fraction</i>	32
1.6	<i>Interpretation of results</i>	34
1.6.1	Chemometrics	36
1.6.1.1	Principal component analysis	36
1.6.1.2	Linear discriminant analysis.....	37
1.6.1.3	Discrimination of forensic soils in literature using chemometrics.....	38
1.6.1.4	Gaps in the literature.....	40
1.7	<i>Aims and overview</i>	40
Chapter 2.	Experimental Methods.....	44
2.1	<i>Introduction</i>	45
2.2	<i>Collection of soil samples</i>	45
2.3	<i>Sample selection and identification</i>	47
2.3.1	Primary investigations (Chapters 3 – 5)	47
2.3.2	Blinded case simulation (Chapter 6)	50
2.4	<i>Preparation of the quartz-recovered fine fraction</i>	50
2.5	<i>Instrumental methods.....</i>	52
2.5.1	Microspectrophotometry (MSP)	52
2.5.2	Raman spectroscopy	52
2.5.3	Attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy	53
2.5.4	X-ray diffraction (XRD).....	53
2.5.4.1	ChemCentre.....	53
2.5.4.2	Commonwealth Scientific and Industrial Research Organisation (CSIRO).....	54
2.6	<i>Data analysis.....</i>	55
2.6.1	Principal component analysis.....	55

2.6.2	Linear discriminant analysis	55
2.7	<i>Analysis sequence</i>	56
2.8	<i>Preliminary considerations</i>	57
Chapter 3.	Analysis of the Quartz-Recovered Fine Fraction of Western Australian Soils using Microspectrophotometry Paired with Chemometrics	58
3.1	<i>Introduction</i>	59
3.2	<i>Experimental</i>	62
3.3	<i>Results and discussion</i>	63
3.3.1	Chemometric analysis of soils using complete MSP spectra.....	63
3.3.1.1	Principal component analysis	63
3.3.1.2	Linear discriminant analysis.....	69
3.3.2	Chemometric analysis of soils using average MSP spectra	74
3.3.2.1	Principal component analysis	74
3.3.2.2	Linear discriminant analysis.....	79
3.3.3	Chemometric analysis of soils using MSP generated L*a*b* values.....	85
3.3.3.1	Principal component analysis	85
3.3.3.2	Linear discriminant analysis.....	90
3.4	<i>Conclusions</i>	96
Chapter 4.	Analysis of the Quartz-Recovered Fine Fraction of Western Australian Soils using Vibrational Spectroscopy Paired with Chemometrics.....	99
4.1	<i>Introduction</i>	100
4.2	<i>Experimental</i>	102
4.3	<i>Results and discussion</i>	102
4.3.1	Analysis of soils using Raman spectroscopy	102
4.3.2	Analysis of soils using ATR-FTIR spectroscopy	104
4.3.3	Principal component analysis.....	106
4.3.3.1	Discrimination using location-based chemical data	106
4.3.3.2	Discrimination using feature-based data.....	113
4.3.3.3	Repeated attempts at discrimination using location-based chemical data.....	115
4.3.4	Linear discriminant analysis	117
4.4	<i>Conclusions</i>	122
Chapter 5.	Analysis of the Quartz-Recovered Fine Fraction of Western Australian Soils using X- Ray Diffraction Paired with Chemometrics	125

5.1	<i>Introduction</i>	126
5.2	<i>Experimental</i>	128
5.3	<i>Results and discussion</i>	129
5.3.1	Analysis of soils using XRD	129
5.3.2	Principal component analysis	133
5.3.2.1	Discrimination using location-based ChemCentre data	133
5.3.2.2	Discrimination using feature-based ChemCentre data	137
5.3.2.3	Discrimination using CSIRO data	140
5.3.3	Linear discriminant analysis	145
5.3.3.1	Predicting CSIRO's validation samples onto the XRD LDA model	145
5.3.3.2	Predicting ChemCentre's validation samples onto the XRD LDA model	150
5.3.4	Reproducibility of data	155
5.4	<i>Conclusions</i>	156
Chapter 6. Application of Developed Analytical Sequence Incorporating Chemometrics to a Forensic Case Simulation		159
6.1	<i>Introduction</i>	160
6.2	<i>Experimental</i>	163
6.2.1	Sample selection	163
6.2.2	Sample analysis	163
6.3	<i>Results and discussion</i>	164
6.3.1	Visual examination of soil samples	165
6.3.2	Microspectrophotometric analysis of soil samples	167
6.3.2.1	Principal component analysis utilising entire MSP spectra	168
6.3.2.2	Principal component analysis utilising average MSP spectra	170
6.3.2.3	Principal component analysis utilising L*a*b* colourimetric values	172
6.3.2.4	Linear discriminant analysis	174
6.3.2.5	Summary of results	176
6.3.3	Infrared spectroscopic analysis of soil samples	177
6.3.3.1	Principal component analysis	178
6.3.3.2	Linear discriminant analysis	181
6.3.3.3	Summary of results	181
6.3.4	X-ray diffractive analysis of soil samples	182
6.3.4.1	Principal component analysis utilising ChemCentre XRD patterns	185
6.3.4.2	Principal component analysis utilising CSIRO XRD patterns	187
6.3.4.3	Linear discriminant analysis	188

6.3.4.4	Summary of results.....	190
6.3.5	Comparison of results with true source.....	191
6.4	<i>Conclusions</i>	192
Chapter 7.	Conclusions and Future Work	195
7.1	<i>Conclusions</i>	196
7.1.1	Investigations into spectroscopic techniques in combination with chemometrics ...	197
7.1.2	Application to blinded case simulation and forensic casework	201
7.2	<i>Future work</i>	203
7.3	<i>Summary</i>	204
References		206
Appendices		222

List of Figures

Figure 1.1 The process of soil formation.....	4
Figure 1.2 Diagram detailing the different dune systems of Western Australia's Swan Coastal Plain, in which the capital city, Perth, is situated (from Pitts, Lewis, and Newland (46)).	9
Figure 1.3 Soil from the Quindalup dune system, located within the Swan Coastal Plain in Perth, Western Australia (from Middle (49)).	10
Figure 1.4 Soil from the Spearwood dune system, located within the Swan Coastal Plain in Perth, Western Australia (from Middle (49)).	11
Figure 1.5 Soil from the Bassendean dune system, located within the Swan Coastal Plain in Perth, Western Australia (from Middle (49)).	12
Figure 1.6 Soil from the Pinjarra Plain, located within the Swan Coastal Plain in Perth, Western Australia (from Middle (49)).	13
Figure 1.7 Flow diagram of a 'typical' forensic soil examination protocol (adapted from Pitts, Lewis, and Newland (46)).	17
Figure 1.8 (a) Plane polarized light microscope image of a suspected metamorphic granulite; (b) cross-polarized light microscope image of the same thin section, showing interference colours ranging from first order whites to higher order blues, pinks, greens and oranges (from Pitts, Lewis, and Newland (46)). Hypersthene, augite, labradorite, quartz and miscellaneous amphiboles were indicated in the sample (field of view is 1.1 mm).	20
Figure 1.9 SEM image of soil particles showing the detailed textural information it may provide, along with chemical identification of the coatings of some grains determined by EDX (from Dawson and Hillier (17)).	24
Figure 1.10 Photomicrograph of Nuphar water lily spore (from Bull, Parker, and Morgan (39)).	26
Figure 1.11 Infographic based on method published by Pitts & Clarke (94): The forensic discrimination of quartz sands from the Swan Coastal Plain, Western Australia.....	33
Figure 1.12 Flow diagram showing a summary of the overall methodology undertaken throughout this thesis.....	41
Figure 2.1 Map of the Perth metropolitan region in Western Australia, illustrating the 23 locations where the soil was sampled for this study.	46
Figure 2.2 The wooden frame used to mark out the areas for collection of the soil samples, illustrated in situ on the soil surface after clearing of the bulky organic material. The five points of sampling are annotated within the frame.	47

Figure 2.3 Soil samples obtained from nine of the different locations used throughout this study, illustrating the visual colour differences.	47
Figure 3.1 A page from a Munsell colour chart, depicting the differing chroma and value shades of 5YR (from “How to Read a Munsell Colour Chart” (115)). The example referenced in the text, 5YR 4/4, is highlighted.	61
Figure 3.2 A 3D model of the CIELAB colour space, where L defines the lightness of a colour, a* defines the green to red axis, and b* the blue to yellow axis (from Yélamos et al. (113)).	61
Figure 3.3 Raw MSP reflectance spectra showing the variability in composition of a selection of soil samples collected from differing locations within the Swan Coastal Plain.	63
Figure 3.4 Scree plot depicting the cumulative variance in the MSP dataset retained by each PC.	64
Figure 3.5 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from different locations based on their corresponding MSP spectra.	65
Figure 3.6 Factor loadings plot of PCs 1-3 for PCA of the soil MSP reflectance dataset.	66
Figure 3.7 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding MSP spectra.	68
Figure 3.8 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding MSP spectra.	68
Figure 3.9 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding MSP spectra.	68
Figure 3.10 Raw calibration and validation MSP spectra obtained from Location 17’s soils (most similar calibration sample pictured).	74
Figure 3.11 Scree plot depicting the cumulative variance in the MSP average spectra dataset retained by each PC.	74
Figure 3.12 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from different locations based on their corresponding average MSP spectra.	75
Figure 3.13 Factor loadings plot of PCs 1-3 for PCA of the soil average MSP reflectance dataset.	76
Figure 3.14 2-dimensional PCA scores plot showing the variability of soil samples from different locations based on their corresponding average MSP spectra. Circle indicates Location 22 which, upon removal of PC-3, now overlaps with Location 18.	77
Figure 3.15 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding average MSP spectra.	78

Figure 3.16 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding average MSP spectra.....	78
Figure 3.17 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding average MSP spectra.....	78
Figure 3.18 Scree plot depicting the cumulative variance in the L*a*b* dataset retained by each PC.	85
Figure 3.19 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from different locations based on their corresponding L*a*b* values, derived from their average MSP spectra.....	87
Figure 3.20 Factor loadings plot of PCs 1-3 for PCA of the soil L*a*b* values, generated from average MSP reflectance spectra.....	88
Figure 3.21 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding L*a*b* values determined from their average MSP spectra.....	89
Figure 3.22 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding L*a*b* values determined from their average MSP spectra.....	89
Figure 3.23 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding L*a*b* values determined from their average MSP spectra.....	89
Figure 3.24 2-dimensional PCA scores plot showing the variability of soil samples from different locations based on their corresponding L*a*b* values, derived from their average MSP spectra.....	93
Figure 3.25 Calibration (average values) and validation L*a*b* values obtained from Location 6's soils.....	96
Figure 4.1 Raman spectra (unprocessed) of four different soil samples from the Swan Coastal Plain, illustrating the high degree of fluorescence that is masking potential characteristic peaks. ..	103
Figure 4.2 The effects of laser intensity (high vs. low power) and integration time (in seconds) on Sample 0's Raman spectra.....	104
Figure 4.3 The effects of baseline correction (baseline offset vs linear baseline) on Sample 4b and 5b's Raman spectra, with visible peaks remaining unchanged and unable to be enhanced across all spectra.	104

Figure 4.4 Baseline corrected and normalised ATR-FTIR absorbance spectra showing the similarities in composition of a selection of soil samples collected from differing locations within the Swan Coastal Plain. Annotations were based off peak assignments outlined in Table 4.1.	105
Figure 4.5 Scree plot depicting the cumulative variance in the ATR-FTIR dataset retained by each PC.	106
Figure 4.6 3-dimensional PCA scores plots generated using the first five PCs, showing the variability of soil samples from different locations based on their corresponding ATR-FTIR spectra.	107
Figure 4.7 Factor loadings plot of PCs 1-3 for PCA of the soil ATR-FTIR dataset, along with representative spectra obtained from samples situated at the extreme points of the scores plot. The main peaks of interest have been highlighted and annotated with their contributing compounds.	111
Figure 4.8 Factor loadings plot of PCs 4-5 for PCA of the soil ATR-FTIR dataset, along with representative spectra obtained from samples that achieved discrimination within the scores plot based on the variation attributed to PCs 4 and 5. The main peaks responsible for the separation between samples have been highlighted and annotated with the contributing compounds.	112
Figure 4.9 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding ATR-FTIR spectra.	113
Figure 4.10 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding ATR-FTIR spectra.	114
Figure 4.11 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding ATR-FTIR spectra.	115
Figure 4.12 Scree plot depicting the cumulative variance retained by each PC in the reduced-sample soil ATR-FTIR dataset.	115
Figure 4.13 3-dimensional PCA scores plot (shown from two perspectives) generated using the first three PCs, displaying the variability of a reduced number of soil samples from different locations based on their corresponding ATR-FTIR spectra.	116
Figure 4.14 Factor loadings plot of PCs 1-3 for PCA of the reduced-sample soil ATR-FTIR dataset.	117
Figure 5.1 Baseline corrected and normalised XRD patterns obtained from ChemCentre, showing the variability in composition of a selection of soil samples collected from differing locations within the Swan Coastal Plain. Annotations were based off assignments outlined in Table 5.1.	131

Figure 5.2 Baseline corrected and normalised XRD patterns obtained from CSIRO, showing the variability in composition of a selection of soil samples collected from differing locations within the Swan Coastal Plain. Annotations were based off assignments outlined in Table 5.1.	131
Figure 5.3 Scree plot depicting the cumulative variance in the ChemCentre XRD dataset retained by each PC.	133
Figure 5.4 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations based on their corresponding ChemCentre XRD patterns.	134
Figure 5.5 Factor loadings plot of PCs 1-4 for PCA of the ChemCentre XRD dataset, with the main reflections of interest highlighted and annotated with their contributing compounds.	137
Figure 5.6 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding ChemCentre XRD patterns.	138
Figure 5.7 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples analysed with different XRD instrumentation, based on their corresponding ChemCentre XRD patterns.....	139
Figure 5.8 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding ChemCentre XRD patterns.	140
Figure 5.9 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding ChemCentre XRD patterns.	140
Figure 5.10 Scree plot depicting the cumulative variance in the CSIRO XRD dataset retained by each PC.....	141
Figure 5.11 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations based on their corresponding CSIRO XRD patterns. .	142
Figure 5.12 XRD patterns obtained from soil samples from Locations 4, 11, and 19, illustrating both the variation in patterns between different instrumentation, and between duplicate samples.	143
Figure 5.13 Factor loadings plot of PCs 1-4 for PCA of the CSIRO XRD dataset, with the main reflections of interest highlighted and annotated with their contributing compounds.	144
Figure 5.14 XRD patterns obtained from soil samples from Locations 6, 8, and 13, illustrating both the variation in patterns between different instrumentation, and between duplicate samples.	152

Figure 6.1 Flow diagram outlining the overall analysis sequence, previously developed throughout Chapters 3 to 5, that was applied to soil samples within the blinded case simulation.	164
Figure 6.2 Baseline corrected and normalised MSP reflectance spectra showing the variability in composition of the soil samples collected from differing locations representative of a blinded case simulation.	168
Figure 6.3 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding MSP spectra.	169
Figure 6.4 Factor loadings plot of PCs 1-4 for PCA of the blinded case simulation MSP reflectance dataset.	169
Figure 6.5 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding average MSP spectra.	171
Figure 6.6 Factor loadings plot of PCs 1-4 for PCA of the blinded case simulation average MSP reflectance dataset.	171
Figure 6.7 3-dimensional PCA scores plots generated using the first three PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding L*a*b* chromaticity values.	173
Figure 6.8 Factor loadings plot of PCs 1-3 for PCA of the blinded case simulation L*a*b* chromaticity values dataset.	173
Figure 6.9 Baseline corrected and normalised ATR-FTIR absorbance spectra showing the variability in composition of the soil samples collected from differing locations representative of a blinded case simulation. Annotations were based off peak assignments outlined in Chapter 4 - Table 4.1.	178
Figure 6.10 3-dimensional PCA scores plots generated using the first six PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding ATR-FTIR spectra.	179
Figure 6.11 Factor loadings plot of PCs 1-3 for PCA of the blinded case simulation ATR-FTIR absorbance dataset, with the main peaks of interest highlighted and annotated with their contributing compounds. Annotations were based off peak assignments outlined in Chapter 4 - Table 4.1.	180
Figure 6.12 Baseline corrected and normalised XRD patterns obtained from ChemCentre, showing the variability in composition of a selection of soil samples collected from differing locations	

representative of a blinded case simulation. Annotations were based off reflection assignments outlined in Chapter 5 - Table 5.1.	184
Figure 6.13 Baseline corrected and normalised XRD patterns obtained from CSIRO, showing the variability in composition of a selection of soil samples collected from differing locations representative of a blinded case simulation. Annotations were based off reflection assignments outlined in Chapter 5 - Table 5.1.	184
Figure 6.14 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding ChemCentre XRD patterns.	186
Figure 6.15 Factor loadings plot of PCs 1-4 for PCA of the blinded case simulation ChemCentre XRD dataset, with the main reflections of interest highlighted and annotated with their contributing compounds. Annotations were based off reflection assignments outlined in Chapter 5 - Table 5.1.	186
Figure 6.16 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding CSIRO XRD patterns.	187
Figure 6.17 Factor loadings plot of PCs 1-4 for PCA of the blinded case simulation CSIRO XRD dataset, with the main reflections of interest highlighted and annotated with their contributing compounds. Annotations were based off reflection assignments outlined in Chapter 5 - Table 5.1.	188
Figure 7.1 Analytical sequence for examination of the quartz-recovered fine fraction of sandy forensic soil samples from the Swan Coastal Plain in Perth, Western Australia, with the information obtained through analysis at each stage.	201
Figure 7.2 Flowchart illustrating how each stage of analysis within the sequence allowed for an association of exclusion to be made between the suspect recovered soil and the reference soils in the blind case simulation in Chapter 6.	203

List of Tables

Table 1.1 Methods commonly utilised for the inorganic analysis of forensic soil (from Pitts, Lewis, and Newland (46)).	18
Table 1.2 Methods utilised for the organic analysis of forensic soil (from Pitts, Lewis, and Newland (46)).	19
Table 2.1 The soil samples used in trials throughout this study, with their associated locations and visual descriptions. Two samples were analysed from each location, except for Location 0 (due to limited sample size). *Bushland is shortened to “bush” in annotated figures throughout this thesis.	48
Table 2.2 The soil samples used in the blinded case simulation, with their associated known locations and visual descriptions.	49
Table 2.3 Masses of quartz grains picked from each sample that were used for preparation of the quartz-recovered fine fractions (to five decimal places).	51
Table 3.1 Number of correct vs incorrect location classifications for samples in the MSP calibration set using a 7-PC LDA model (percentages rounded to nearest whole number).	70
Table 3.2 Discriminant values of Location 17’s replicates from the MSP calibration dataset (rounded to three decimal places), with correct classifications shaded green. The last column demonstrates how far away the next closest classification was, as a percentage of the lowest discriminant value obtained.	71
Table 3.3 Number of correct vs incorrect location predictions for samples in the MSP validation set using a 7-PC LDA model (percentages rounded to nearest whole number).	73
Table 3.4 Number of correct vs incorrect location classifications for samples in the average MSP calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).	80
Table 3.5 Discriminant values of Location 10, 15, and 17’s samples from the average MSP calibration dataset (rounded to three decimal places), with correct classifications shaded green. The last column demonstrates how far away the next closest classification was, as a percentage of the lowest discriminant value obtained.	81
Table 3.6 Number of correct vs incorrect location predictions for samples in the average MSP validation set using a 2-PC LDA model (percentages rounded to nearest whole number).	83
Table 3.7 Discriminant values of Location 4, 7, 12, 16, and 22’s samples from the average MSP validation dataset (rounded to three decimal places), with correct predictions shaded green.	

The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.....	84
Table 3.8 The soil samples analysed with MSP and their associated L*a*b* values generated from their average MSP spectra (rounded to four decimal places).....	86
Table 3.9 The soil sample replicates utilised for the LDA validation set and their associated L*a*b* values generated from their MSP spectra (rounded to four decimal places).	91
Table 3.10 Number of correct vs incorrect location classifications for samples in the L*a*b* calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).	92
Table 3.11 Number of correct vs incorrect location predictions for samples in the L*a*b* validation set using a 2-PC LDA model (percentages rounded to nearest whole number).	94
Table 3.12 Discriminant values of Location 16, 17, and 22's samples from the L*a*b* validation dataset (rounded to three decimal places), with correct predictions shaded green. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.	95
Table 4.1 Common minerals found in Swan Coastal Plain soils and their associated IR peaks.	106
Table 4.2 Number of correct vs incorrect location classifications for samples in the ATR-FTIR calibration set using a 4-PC LDA model (percentages rounded to nearest whole number).	119
Table 4.3 Number of correct vs incorrect location predictions for samples in the ATR-FTIR validation set using a 4-PC LDA model (percentages rounded to nearest whole number).	120
Table 4.4 Discriminant values of replicates from the ATR-FTIR validation dataset (rounded to three decimal places), with correct predictions shaded green and incorrect predictions shaded red. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.....	121
Table 5.1 Common minerals found in Swan Coastal Plain soils, and their most intense identifying XRD reflections ¹ (rounded to one decimal place) that were used as a guide to identify minerals within the XRD patterns.	132
Table 5.2 Number of correct vs incorrect location classifications for samples in the first combined XRD calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).	146
Table 5.3 Number of correct vs incorrect location predictions for samples in the CSIRO validation set using a 2-PC LDA model (percentages rounded to nearest whole number).	148
Table 5.4 Discriminant values of replicates from the CSIRO XRD validation dataset (rounded to three decimal places), with correct predictions shaded green and incorrect predictions shaded red.	

The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.....	149
Table 5.5 Number of correct vs incorrect location classifications for samples in the second combined XRD calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).	151
Table 5.6 Number of correct vs incorrect location predictions for samples in the ChemCentre validation set using a 2-PC LDA model (percentages rounded to nearest whole number).....	153
Table 5.7 Discriminant values of replicates from the ChemCentre XRD validation dataset (rounded to three decimal places), with correct predictions shaded green and incorrect predictions shaded red. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.....	154
Table 6.1 The soils utilised for the blinded case simulation, with their associated information. Details regarding the suspect recovered soil were kept confidential until all analyses were completed, and results were presented.....	165
Table 6.2 Visual descriptions and photographs of the soil samples provided for analysis within the blinded case simulation.....	166
Table 6.3 Predictions for samples in the MSP blinded case simulation validation set using a 4-PC LDA model.....	174
Table 6.4 Predictions for samples in the average MSP blinded case simulation validation set using a 2-PC LDA model.....	175
Table 6.5 Predictions for samples in the L*a*b* chromaticity values blinded case simulation validation set using a 2-PC LDA model.....	176
Table 6.6 Predictions for samples in the ATR-FTIR blinded case simulation validation set using a 6-PC LDA model.....	181
Table 6.7 Predictions for samples in the ChemCentre XRD blinded case simulation validation set using a 2-PC LDA model.....	189
Table 6.8 Predictions for samples in the CSIRO XRD blinded case simulation validation set using a 2-PC LDA model.....	190

List of Abbreviations

AES	Atomic emission spectrometry
APD	Automatic powder diffractometer
ASRIS	Australian Soil Resource Information System
ATR	Attenuated total reflectance
CDFA	Canonical discriminant function analysis
CIE	Commission Internationale d'Éclairage (International Commission on Illumination)
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DNA	Deoxyribonucleic acid
DRIFTS	Diffuse reflectance infrared Fourier transform spectroscopy
EBSD	Electron back-scattered diffraction
EDX	Energy dispersive X-ray
FTIR	Fourier transform infrared
GC	Gas chromatography
GPS	Global Positioning System
HCA	Hierarchical cluster analysis
HPLC	High performance liquid chromatography
HTS	High throughput sequencing
ICDD	International Centre for Diffraction Data
ICP	Inductively coupled plasma
IR	Infrared
IRMS	Isotope-ratio mass spectrometry
LA	Laser ablation
LC	Liquid chromatography
LDA	Linear discriminant analysis
LIBS	Laser-induced breakdown spectroscopy
MC	Multiple comparison

MIR	Mid-infrared
MS	Mass spectrometry
MSP	Microspectrophotometry
NAS	National Academy of Sciences
NATA	National Association of Testing Authorities
NIFS	National Institute of Forensic Science
NIPALS	Non-linear iterative partial least squares
NIST	National Institute of Standards and Technology
OES	Optical emission spectrometry
OM	Optical microscopy
PC	Principal component
PCA	Principal component analysis
PCAST	President's Council of Advisors on Science and Technology
PDF	Powder Diffraction File
PLM	Polarised light microscopy
QEMSCAN	Quantitative evaluation of minerals by scanning electron microscopy
rNA	Ribonucleic acid
SEM	Scanning electron microscopy
SERS	Surface-enhanced Raman spectroscopy
THM	Thermally assisted hydrolysis and methylation
UPLC	Ultra performance liquid chromatography
USA	United States of America
UV	Ultraviolet
VSC	Video spectral comparator
XRD	X-ray diffraction
XRF	X-ray fluorescence
2D	Two-dimensional
3D	Three-dimensional

Chapter 1. Introduction

Portions of this chapter have been published, or submitted for publication, in the following articles:

K. Pitts, S. Lewis, and **T. Newland**. "Geochemistry | Soil and Mineralogical Analysis." In *Encyclopedia of Analytical Science (Third Edition)*, P. Worsfold, C. Poole, A. Townshend, and M. Miró, 2019. p. 292-301. Oxford: Academic Press.

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics for the forensic analysis of Western Australian sandy soils." *Forensic Chemistry*, 2022. 28: 100412.

T. G. Newland, K. Pitts, and S. W. Lewis. "Negative result: Application of Raman spectroscopy to the forensic analysis of an arid, sandy, soil." *Forensic Science International: Reports*, 2022. Submitted for publication.

1.1 Soil as forensic evidence

Soil is a valuable but underutilised source of forensic trace evidence (1-5). It can be used in criminal investigations to establish connections between people, places, and objects, and is generally retrieved from a suspect's clothing, footwear, vehicle, or crime scene (2-4, 6-10). Soil evidence can be of significant value to investigators due to its heterogeneity and transferability (10, 11). It can be analysed by a range of different techniques to collect large amounts of detailed information on its composition, which is known to vary extensively with location (10, 11). Even though neighbouring soil systems tend to be closely related, they each may have distinctive geology, flora and fauna, topography, and history that could allow for their differentiation and identification (3, 10, 12, 13). When this data is interpreted correctly, soil can make a great tool for reconstructing previous events, as well as for forensic intelligence (12, 14-18). However, despite its huge potential, soil is not frequently used as a form of forensic trace evidence due to challenges associated with the reliability of its analysis and interpretation (19, 20).

In a forensic context, the majority of soil evidence is urban in origin, containing a mixture of inorganic mineral grains, organic matter, a living microbiome, and man-made materials (1, 3, 6, 21). The underlying chemical and physical properties of soil are naturally in a state of change, leading to a highly complex and dynamic material (2, 8-10, 12, 13). As a result of this, the development, implementation, and acceptance of standardised soil analysis procedures has been challenging (12, 16, 21, 22). While several techniques hold the potential for forensic soil analysis, rigorous method validation is a prerequisite for these methods to be routinely adopted (8, 9, 13). There has also been a reported increase in the prevalence of soil evidence in recent international criminal investigations, leading to increased demand for more reliable techniques for objective forensic examination of soils (4, 8, 13).

While there is a current demand for the increased use of soil evidence to assist in solving crime, there is still an absence of any in-depth investigations in published literature into arid, sandy soils, such as those found on the Swan Coastal Plain, and their role as geological evidence in a forensic context. Understanding the frequency

and nature of a particular trace is central to being able to evaluate its source (23, 24); this is particularly true for soils, as their properties can fluctuate greatly across different locations. Soil evidence is also often overlooked for its potential to be used as an investigative tool throughout criminal investigations (14, 24). If the science behind the analysis of forensic soil samples is strong enough, it could be used in instances to lead police to previously undiscovered locations of interest for further examination (8, 12, 14).

This thesis aims to educate on the potential value of sandy soils when used for site discrimination in a forensic context, in order to allow for the reintroduction of soil evidence back into forensic laboratories in dry, arid locations that experience soils like these. By developing methods that can reliably and objectively analyse these sandy soils and identify their differentiating features, it is possible to produce results that are admissible as forensic evidence in court.

1.2 Soil formation and composition

This section is intended to provide a broad overview of soil formation (pedogenesis) and composition, as this forms the basis of the discussion on why soils differ by location. As with other forms of forensic trace evidence, such as fibres, it is important to understand their origin processes and prevalence in order to better interpret their likely characteristic features (25, 26). However, this thesis only focusses on the comparison of surface soils in a forensic context, as these are the traces that are most likely to be transferred during the occurrence of a crime. Hence, the following sections do not cover the geological pedogenesis process in detail, nor do they provide extensive information on soil horizons and classification systems, as this was considered to be outside the scope of this thesis.

1.2.1 The formation of soils

Soil can be generally defined as the loose materials, or regolith, found above the bedrock of a given geological region (1, 13). The formation of soils is related to the products of weathering; in-situ rock is broken down through a variety of processes, such as wind and water erosion, and plant and animal activities. The unconsolidated

regolith may combine with organic matter shed from neighbouring plants and animals to form a young soil. Time, additional sedimentation, biota, and the leaching action of water will then create layers known as soil horizons, distinguished by colour, structure, texture, and/or chemistry (Figure 1.1). In addition to in-situ soil formation, weathering and erosion of rock particles may also occur until the rock particles reach a size that is small enough to be transported and deposited by rainfall, ocean currents, rivers, and wind. The transported regolith may then also combine with organic matter, either during the process of transportation, such as within a river system, or after deposition, such as a wind-blown (aeolian) dune.

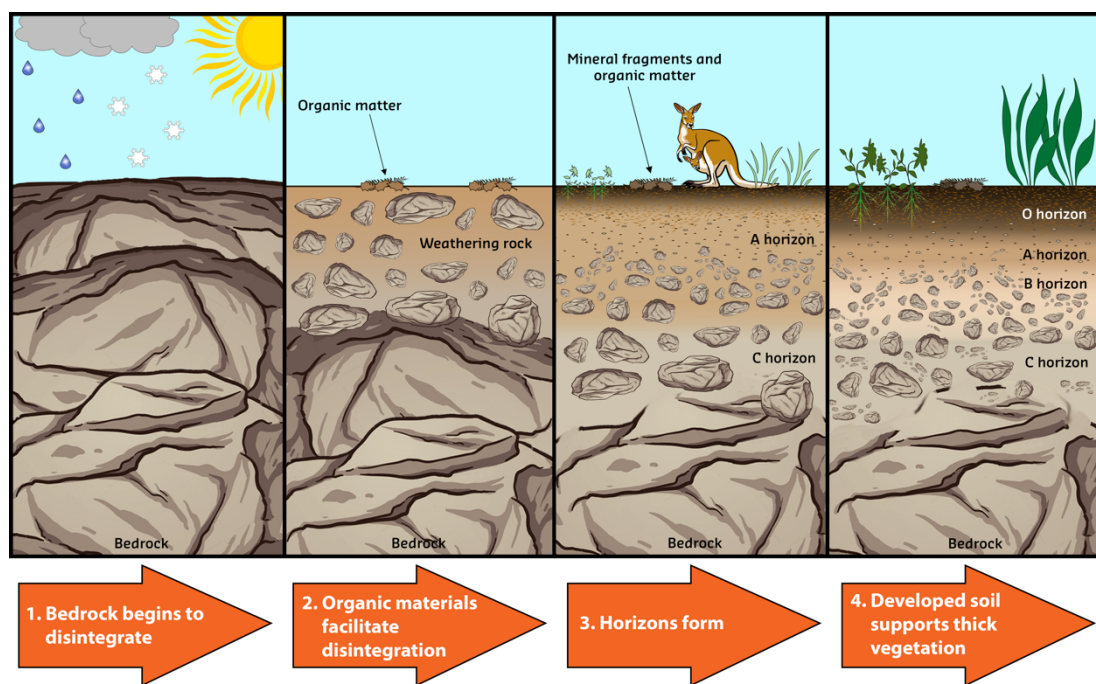


Figure 1.1 The process of soil formation.

1.2.2 The composition of soil

Soil is made up of varying proportions of inorganic rock particles and mineral grains, chemical precipitates, organic plant and animal matter, and bacteria (3, 6, 13). The inorganic components of soil are derived through weathering of the underlying parent material to produce minerals and rock fragments, as well as through anthropogenic means, such as the addition of builder’s sand (17, 27). Minerals are solid, naturally occurring elements, or compounds that are made up of several different elements, arranged in specific crystal structures. Their physical and

chemical properties are determined by both their chemical compositions and crystal structures (17). Primary minerals are those that have formed from primary igneous rocks via solidification processes, and secondary minerals are then formed from further weathering or alteration of these primary minerals (28). While there are several thousand different minerals that exist, only 20 – 30 of these are commonly found in soils above trace concentrations, such as quartz, feldspars, iron oxides, aluminium oxides, carbonates, and clay minerals (1, 17, 29). A typical soil sample, however, will contain a mixture of only 5 – 10 different major minerals present in amounts that can be detected by routine analysis techniques (1, 17, 29).

The organic components of soil originate from surrounding plant and animal material (8). Soil organic matter consists of mixtures of lipid and wax compounds, proteins, carbohydrates, humic acids, cellulose, hemicellulose, complex polysaccharides, lignin, suberin, and cutin, derived from living plants, roots, and organisms, as well as decomposing plant and animal remains (humus) (6, 17). The biological activity of microorganisms present within the soil degrades the residual plant and animal matter, which can then significantly affect the soil structure, fertility, and water holding capacity (30). Characteristic patterns of these compounds found within the surrounding vegetation are mirrored by the surrounding soil, providing information on the site history and associated environmental conditions (6, 17, 31). These compounds can persist for long periods of time and can vary over spaces as small as several metres, making them useful markers for site discrimination (6, 17, 32). Managed soils that have been purposefully altered can also contain organic compounds from the application of manures, composts, or mulches to the soil (17). For these reasons, surface soil tends to contain higher levels of organic content than the underlying layers (33).

Additionally, soils contain vast communities of microbes, fungi, and other microorganisms, which can vary widely over extremely small spaces (2, 12, 34). These populations are determined and supported by the nutrients and waste material produced by the surrounding flora and fauna (31). The presence of some plants can also affect the conditions of the soil, such as moisture content, which will directly impact soil organisms (31). These living organisms are responsible for the production

of proteins, which in their complete sets are referred to as 'proteomes' (35). These can vary qualitatively and quantitatively at any given time, under defined conditions (35). They also deposit an abundance of deoxyribonucleic acid (DNA) in the soil, adding to the DNA left behind by other animals that live in, dig through, defecate, or die and decay in the soil (18). While present in abundance, these microbial communities are very sensitive to exogeneous factors, and can be dramatically altered by time, weather, or disturbance of the soil by human practises, such as farming (12, 17).

Because of the prevalence of soil, there are many other materials that are commonly encountered within its mixture. These can include pollen spores, glass fragments, asphalt or brick fragments, fossils, road marking beads, or other artificial materials introduced anthropogenically, such as litter (21, 29). These are dependent on the type of land use surrounding the soil and the accessibility of the location to the public.

1.2.3 Factors affecting soil composition

The composition of soil is largely site-specific and highly complex, with its mineralogy, elemental distribution, and organic matter varying depending on the underlying parent source material, age and weathering of deposits, climate, topography, land use, local pollution, the surrounding vegetation, the microorganisms present in the soil, animal residues, and artefacts of human activities (6, 10, 13, 14, 19, 36). For this reason, soils that are in neighbouring locations may have very distinctive compositions if they are subject to different external factors.

As discussed above, parent source material (underlying bedrock) will determine the major types of minerals present within the soil as it is eroded into small particles (17, 27, 31). It will also affect the overall soil texture. Biological activity at the site will considerably influence the soil composition, largely related to the organic matter and microbial communities (6, 17, 31). This can range from soil disturbance due to the movement of larger animals, as well as their defecation and decay, to mechanisms

with which the microbes and plant systems extract or deposit nutrients within the soil.

The other primary factors influencing soil composition are climate, topography, and time. The climate that the soil is exposed to, e.g., temperatures, rainfall, wind, and humidity, has a direct effect on its weathering; soils that are exposed to heavier rainfall with strong drainage will typically experience higher levels of leaching, and regions with stronger wind currents will generally produce more weathered soils that are mixed in origin. The topography, or relief, of a landscape will also affect the rate of erosion, transport, and deposition of sediments, which in turn will alter the soil composition on a local scale. Lastly, the ageing of soils over extensive periods of time will also directly affect the composition of the soil; as soil ages, it is transported across locations and exposed to the elements, until it is finally buried by more material and compacted underneath the ground's surface. Older soils generally have higher levels of mineral breakdown, and therefore contain lesser amounts of clays, nutrients, and organic matter. Studies have shown that significant temporal variation can be observed in the organic content of soils (32).

1.2.4 Composition of forensic soils

Soils that are commonly utilised as trace evidence within forensic investigations tend to originate from areas nearby the main population density, and are therefore generally urban soils, rather than agricultural (24). This means that forensic soils are particularly affected by human activity, movement, and contamination (14). Soil samples are often recovered from public parks, garden beds, and road verges, and are characterised by their high levels of heterogeneity resulting from the addition of various materials to these sites, such as mulch or fertilizers, and mixing of the soil material through activities such as landscaping or high levels of traffic (24, 37). Soil samples collected from crime scenes are also often contaminated with other forms of trace evidence, such as glass, fibres, or hair. These should be separated and analysed appropriately alongside the soil, to maximise the potential value of the evidence (17).

Often, by the time a forensic soil sample is collected for analysis, it has been unintentionally sampled and then mixed with multiple other sources (10, 14, 17, 38, 39); for example, soil collected from a person of interest's shoe will represent all instances where the shoe made contact with soil, both before and after the alleged event in question, and the portion of the soil adhering to the shoe may preferentially favour particles of a certain size based on the width of the tread marks. Sample may also be lost during subsequent activities involving the use of the item, prior to forensic analysis (39). Even if the item was not subject to further use between the crime and the collection of evidence, these events can be separated by lengthy periods of time, causing temporal mismatches in the soil conditions (40). Forensic soil samples are therefore generally not a true representation of the bulk soil material (10, 14, 17, 38).

1.3 Swan Coastal Plain

One example of a region that contains several soil systems in close proximity is that of the Swan Coastal Plain in Perth, Western Australia. Approximately 200 million years ago, the Australian continental plate started drifting away from the African and Indian plates, causing a rift valley known as the Perth Basin (41). This basin now extends over 170 000 km², from the easterly Darling Fault line, past the current coastline and into the Indian ocean (41). Over time, the coastline has gradually grown westwards as more sediments have been deposited over the top of the basin, creating the Swan Coastal Plain. Three deposition processes are responsible for this: marine (sediment brought to the shore from the ocean), aeolian (sediment wind-blown inland from the shore), and alluvial (sediment from the easterly upland areas brought down by rivers, gravity, and rain) (42-44). Whilst this sedimentary system is currently stable, these processes are still occurring today. The Swan Coastal Plain now encompasses the majority of the Perth metropolitan region, and encapsulates four main soil systems; the Quindalup, Spearwood, and Bassendean dunes, as well as the Pinjarra Plain (Figure 1.2) (36, 45). While the four systems are closely related, they each have distinctive geology, flora and fauna, and topography that allow for their differentiation and identification.

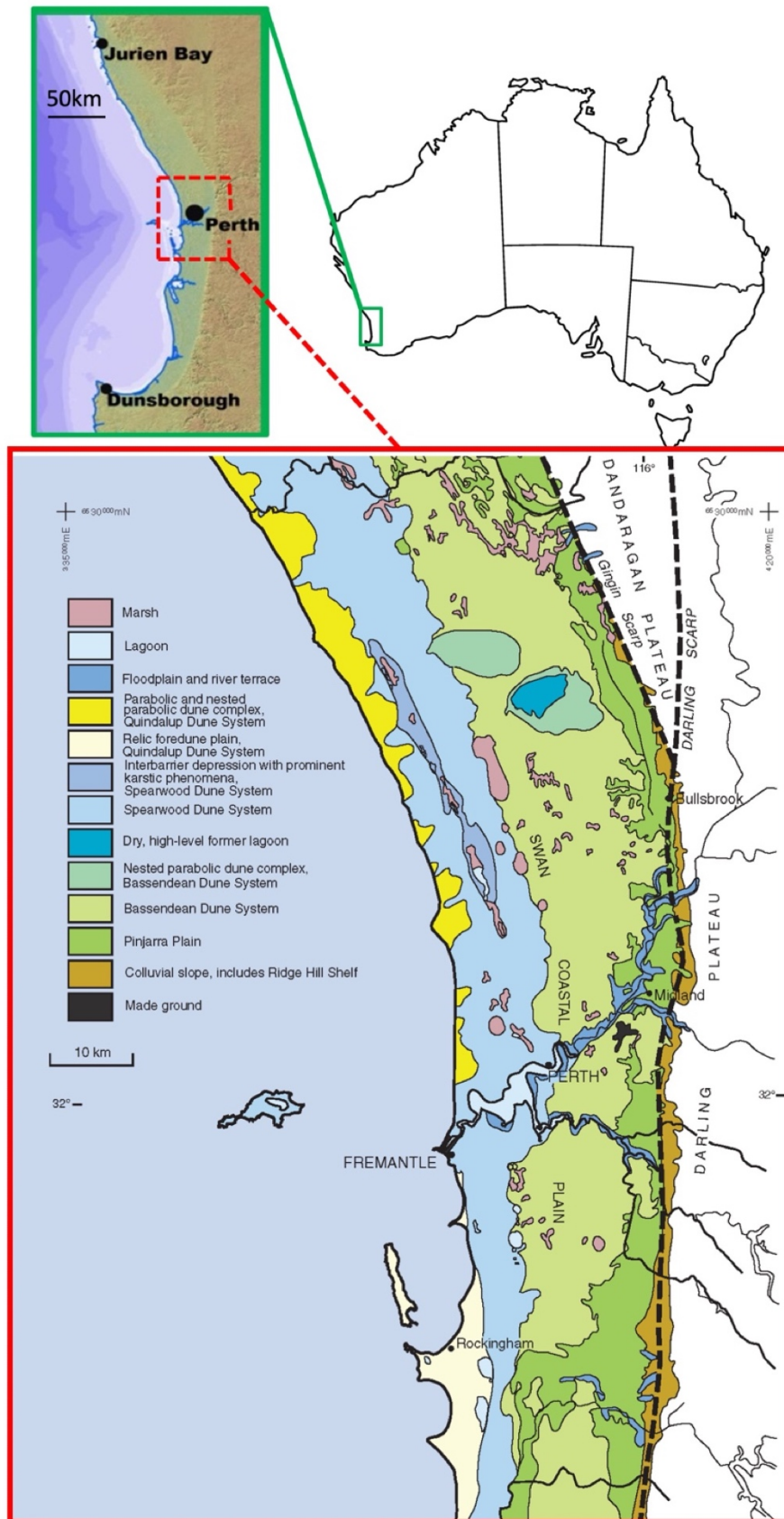


Figure 1.2 Diagram detailing the different dune systems of Western Australia's Swan Coastal Plain, in which the capital city, Perth, is situated (from Pitts, Lewis, and Newland (46)).

The Quindalup dunes (pictured in Figure 1.3) are situated nearest to the coast, and are the youngest of all the systems; their formation began at the end of the last ice age approximately 10,000 years ago, however, in most places along the coast they are less than 1000 years old (41, 43). These marine/aeolian, cream-coloured soils are mostly composed of medium to fine-grained quartz sand and calcium carbonate marine fragments, causing them to be quite alkaline (43, 44, 47). The degree of soil development is apparent through the immature soil profiles, minimal quantities of humus, and high carbonate soil content (42, 47). Calcite is underneath the surfaces in the form of limestone pipes throughout, as well as a solid base at the bottom of the dunes, known as the Tamala Limestone (48). This limestone can be dissolved by water to form hollowed-out caves along the coastline (44).



Figure 1.3 Soil from the Quindalup dune system, located within the Swan Coastal Plain in Perth, Western Australia (from Middle (49)).

Slightly inland of the Quindalup dunes are the Spearwood Dunes (pictured in Figure 1.4), formed during the Pleistocene glacial and interglacial periods approximately 80,000 – 40,000 years ago (36, 43). These dunes are raised higher than the other dune systems and have the most variation in their soils, containing a mixture of red/brown, yellow, and pale yellow/grey sands (42, 43). The medium to coarse sand

grains are generally coated with layers of iron and aluminium oxides, which determines the colour of the sands and also increases their capacity to retain phosphorus (50). As they age however, the iron and aluminium coatings are leached, causing them to lose their distinct colouring over time (41, 43). For this reason, the darkest coloured sands occur to the west of the dune system, and the lighter coloured sands to the east. The Spearwood dunes are similarly situated on Tamala limestone bedrock, however, only the western margin of the dune system contains limestone pipes throughout, produced by calcium carbonate from former marine material (36, 42, 51). The soils within this system are believed to have originated from complex processes, containing large amounts of quartz with some calcium carbonate shell fragments, and smaller quantities of feldspars and heavy minerals (41, 43, 52).



Figure 1.4 Soil from the Spearwood dune system, located within the Swan Coastal Plain in Perth, Western Australia (from Middle (49)).

Farthest from the coast are the Bassendean dunes (pictured in Figure 1.5); these were formed during the middle Pleistocene period approximately 800,000 – 125,000 years ago, making them the oldest of the dune systems (43, 44). Due to their age, the coatings on the sand grains have been predominantly washed away, making these soils the most leached, infertile, and acidic (43, 51). This has also made it difficult to

determine their origin, however, it is generally believed that they were formed by aeolian and/or marine processes (41, 44). The white, yellow, and grey Bassendean sands contain minimal clay or silt, very low levels of nutrient elements, and hence, little organic matter (52). These dunes exhibit low-lying hills throughout, and the regions between these can contain layers of iron and humus podzols underneath their surface, formed by the poor drainage in these areas (43, 51, 52).



Figure 1.5 Soil from the Bassendean dune system, located within the Swan Coastal Plain in Perth, Western Australia (from Middle (49)).

Bordering the eastern edge of the Bassendean dunes is the Pinjarra Plain (pictured in Figure 1.6). This region is flat and low lying, formed alluvially by the major rivers that run through the Swan Coastal Plain (42, 44, 48). The soils within this system are relatively complex. While they contain kaolinite and other clays eroded from the Darling Scarp and carried down by rivers and streams, finally being deposited as fans of alluvium, there is also sand in the soils where they overlay and mix with the Bassendean and Spearwood dunes (43, 44). Contributions from the Ridge Hill Shelf, which constitute colluvial materials from the Darling Scarp, are also found within this particular area (42, 43). This has resulted in a large amount of clay with many layers of soils underneath that range in age (53). Given the clay nature of these soils

combined with the low-lying topography, the Pinjarra region is naturally poorly drained and holds excess surface water for most of winter, but this also makes it more fertile than the other systems (43). The Pinjarra Plain occupies approximately one third of the Swan Coastal Plain, however, most of it has been cleared for agriculture in recent years, leaving only small remnants of the native vegetation (42, 53).



Figure 1.6 Soil from the Pinjarra Plain, located within the Swan Coastal Plain in Perth, Western Australia (from Middle (49)).

1.3.1 Characteristics of the Swan Coastal Plain

The capital city of Perth, in Western Australia, has a Mediterranean climate with hot, dry summers and mild, wet winters, with average maximum daily temperatures ranging from 19 – 31°C (41, 42, 44). The amount of rainfall each year averages around 790 millimetres, with over 70% of that falling between the months of May and August (41, 44). The city also experiences frequent south to south-westerly coastal winds, with the occasional stronger south-westerly through north-westerly winds (42). The landscape exhibits distinct local patterns in the vegetation, related to distance from the ocean, degree of soil development, position in the topographical landscape, and

the fire history of the region (47). There are 26 different vegetation complexes within the Swan Coastal Plain, and these are home to over 300 bird and 50,000 invertebrate species, as well as abundant reptile, amphibia, and microfauna species (45). However, there are some variations in the density of these inhabitants due to the quality of the soils in particular regions. Populations of small to medium-sized mammals have unfortunately suffered significant declines associated with introduced carnivorous predator species and human activity, such as land clearing and controlled burns (45). Off the coast, marine seagrass provides shelter for many species of tropical fish (45).

These conditions have given rise to soils that are extremely dry and nutrient-poor (52). Over time, chemical weathering has leached most of the primary minerals from the soils and concentrated the quartz, leaving almost pure quartz sand (>95%) with only thin iron oxide and residual mineral coatings (41, 52, 54). These coatings can include minerals such as kaolinite, microcline feldspar, gibbsite, goethite, mica (believed to be predominantly muscovite), hematite, vermiculite, anatase, calcite or aragonite, mixed with other forms of iron oxides, clays, magnesian calcites, and magnesian silicates (43, 52). The coastal Quindalup dunes rapidly accumulate organic carbon and nitrogen, allowing maximum quantities to be stored within the Spearwood dunes, before declining further inland (43, 55). Phosphorus levels within the soil exhibit a steady decline across the dunes as they increase in age, with some of the lowest quantities in the world apparent in the Bassendean dunes (41, 55). This has a negative influence on plant biomass and microbial populations, limiting the diversity of ecosystem processes that can occur within this region (41, 55). From a forensic standpoint, this also makes analysis of these soils challenging by reducing the organic fraction present.

1.4 The forensic analysis of soils

1.4.1 The forensic approach

Soil is one of the most challenging forms of forensic trace evidence to sample, analyse, and interpret, due to its inherent heterogeneity (16, 24). Generally, soil

samples that are recovered from items or people of interest (also called *questioned* samples) are compared to reference samples obtained from known sources of origin, and a measure of similarity is provided between these samples (15, 56, 57). These conclusions however, are dependent on a number of factors, such as the precision of the methods used for analysis, the degree of variation existing within the soils from the location of interest, and how common the observed similarities/differences are in soils originating from other locations (57). While it is possible to exclude a soil sample as having originated from a specific location, it cannot be determined with absolute certainty that a soil has originated from a single source (56).

Consideration must also be given to the representativeness of the samples provided; have specific portions of the sample been transferred or retained preferentially, under what conditions were they transferred, and how long ago did the transfer take place (4, 10, 14, 17, 38-40)? Reference soil samples should then be collected under comparable conditions, as close as possible to the crime scene, as soon as possible after the crime occurred, and at a similar depth to the soil exposed during the crime. For example, if the offender walked through a garden bed, then surface samples are appropriate, versus when dealing with a burial site, the layers below the surface should also be sampled (11, 17, 38, 56).

Forensic soil samples can also be compared to reference soil databases to provide forensic intelligence, by directing investigators to specific areas of interest that share similar characteristics with the questioned soil or excluding other areas that show significant differences (37, 58, 59). However, this method is often limited in its potential by the disparity in size between the minute soil samples collected as trace evidence and much larger samples collected for regional soil surveys (17). Additionally, soil databases would need to be highly specific and constantly updated in order to stay accurate over time, and as such, show much greater potential for use on a regional scale, rather than local (37).

1.4.2 Methods utilised for forensic soil analysis

Before the introduction of DNA evidence into courtrooms, soil samples were regularly obtained during criminal investigations and analysed using mainstream soil science methods, such as microscopic examination for particle size, shape, colour, texture, and general object identification, pH determination, and density gradient columns (4, 33, 34, 60, 61). However, the tests conducted were primarily physical and therefore could only assist with the general identification of soil type (12, 34). They also required a great deal of human judgement and interpretation, so were not entirely accepted by the courts as reliable methods for forensic soil examination (3). Newer techniques such as cathodoluminescence, thermoluminescence, and laser granulometry were experimented with briefly, but were regarded as novel methods and not routinely used for casework (61). The introduction of improved DNA evidence into courts caused a significant depreciation in the perceived forensic value of soil evidence, leading to it being largely overlooked (24, 61, 62). In recent years, it has been recognized that criminal investigations and subsequent court proceedings cannot stand on DNA evidence alone. This, and major improvement in scientific instrumentation and techniques, has encouraged the development of new approaches to forensic soil examination (38, 62, 63).

The methods utilised by forensic soil analysts today have primarily been sourced from mineralogy and soil science research, and applied to soil evidence in a forensic context (20). Techniques which are easy-to-use, cost-effective, and non-destructive to the sample, have sensitive detection limits, and require small sample sizes are particularly valued due to the nature of the evidence encountered in forensic casework (8, 17, 24, 29, 54, 64). When utilising several techniques in combination, consideration should also be given to the logical order in which the sequence of methods should be carried out (17, 24). A simplified flow diagram of a typical soil examination protocol is illustrated in Figure 1.7 (46). Whilst not standardized, most forensic soil examiners would follow a similar process, with variations depending on instrument availability, characteristics of the soils native to the areas of focus, and the expertise of the analyst. Most analysts will begin with an overall examination of the bulk characteristics of the soil, such as colour, mineral composition, and a

stereomicroscopical examination before undertaking some form of separation for a more detailed examination of the appropriate fractions (17, 24, 56). Man-made artefacts may be separated and examined by other trace evidence examiners and intact plant material by botanists. Palynologists may also process appropriate fractions of soil for the examination of pollen grain populations.

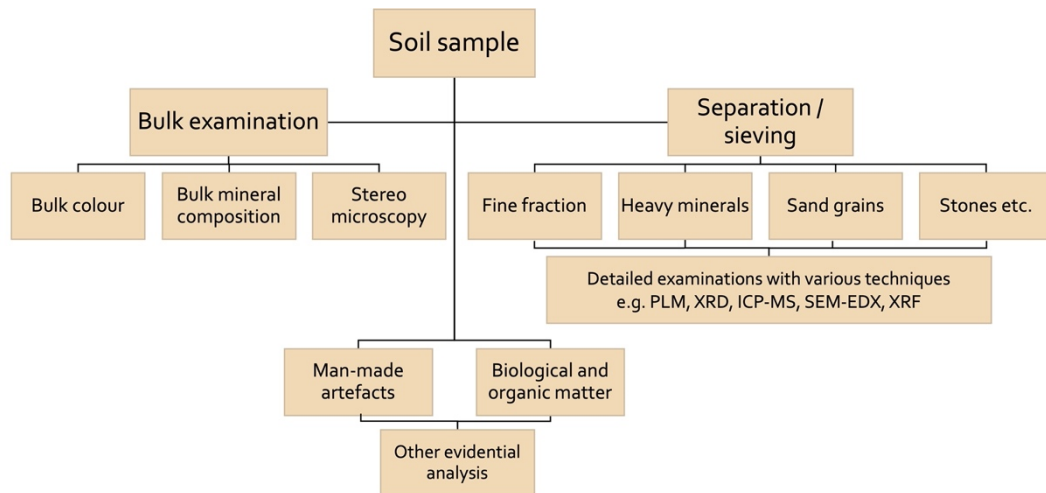


Figure 1.7 Flow diagram of a 'typical' forensic soil examination protocol (adapted from Pitts, Lewis, and Newland (46)).

In the majority of forensic applications, the inorganic content within soil has generally been the focal point of analysis rather than the organic material, which has resulted in a wide range of instrumental techniques used for its analysis (1, 2, 4, 6, 32). The inorganic component can provide information on the soil's mineral content and underlying geology, as well as specific site characterisation through the presence of additives, such as builder's sand (17). One advantage of analysing the inorganic fraction of soils is that it is generally stable for long periods of time, independent of sample storage conditions (17, 65, 66). Table 1.1 outlines some of the advantages and disadvantages associated with some of the most common methods, with further summaries of each technique detailed below, as well as in subsequent chapters.

Table 1.1 Methods commonly utilised for the inorganic analysis of forensic soil (from Pitts, Lewis, and Newland (46)).

Method	Advantages	Disadvantages
PLM	<ul style="list-style-type: none"> • Non-destructive technique • Inexpensive • Relatively fast 	<ul style="list-style-type: none"> • Only qualitative information provided • Requires expertise
XRD	<ul style="list-style-type: none"> • Both qualitative and quantitative information provided • Non-destructive technique • Able to distinguish isomorphs with differing crystal lattices 	<ul style="list-style-type: none"> • Only semi-quantitative information • Sample size required may be larger than most forensic samples collected • Requires time-consuming sample preparation
ICP-MS	<ul style="list-style-type: none"> • Both qualitative and quantitative information provided • Low detection limits 	<ul style="list-style-type: none"> • Requires time-consuming sample preparation • Destructive • Cannot differentiate isomorphs
SEM-EDX	<ul style="list-style-type: none"> • Both qualitative and semi-quantitative information provided • Imaging of samples • Non-destructive technique 	<ul style="list-style-type: none"> • Samples must be stable under vacuum • Detection limits limited to ~1% • Cannot distinguish isomorphs
XRF	<ul style="list-style-type: none"> • Semi-quantitative elemental analysis 	<ul style="list-style-type: none"> • Detection limits vary, with elements lighter than sodium difficult • Cannot distinguish isomorphs

The organic compounds present in soil have been poorly studied in the context of Australian forensic applications (2, 6, 8, 17). This has presumably been due to the transient nature of organic soil characteristics, which are temporal and unstable when exposed to varying weather conditions or physical disturbances (20). Compounds that experience faster levels of decay are generally less useful for forensic comparison than stable inorganic compounds, as they can prove difficult to characterise and analyse quantitatively and are unable to provide a robust profile of the soil (17). While there are several established methods for organic soil analysis in areas of geochemistry and soil science research, they have only been explored very recently for forensic soil examination and are some of the first published in open

literature (6, 8, 9, 32, 57, 59). Some of these methods are outlined in Table 1.2 and further sections below.

Table 1.2 *Methods utilised for the organic analysis of forensic soil (from Pitts, Lewis, and Newland (46)).*

Method	Advantages	Disadvantages
GC-MS	<ul style="list-style-type: none"> • Method well established from agricultural sciences 	<ul style="list-style-type: none"> • Larger than typical forensic sample size required • Some compounds unstable at temperatures of column • Sample preparation may be arduous
HPLC/UPLC	<ul style="list-style-type: none"> • Heat-sensitive compounds able to be detected 	<ul style="list-style-type: none"> • Methods not well established
OM	<ul style="list-style-type: none"> • Non-destructive technique • Inexpensive 	<ul style="list-style-type: none"> • Only qualitative information provided • Open to human interpretation
IRMS	<ul style="list-style-type: none"> • Able to compare isotopic level variations 	<ul style="list-style-type: none"> • Destructive • Interpretation complicated and requires good databases

1.4.2.1 Polarised light microscopy

Polarized light microscopy (PLM) was one of the earliest techniques utilised to analyse minerals, including those in soils (32, 67). Mineralogists have been using PLM for over a century to identify minerals within a rock, using characteristics such as their birefringence, relative refractive index, and optical sign, and the same methods can be applied to forensic soil samples (68-70). Different materials will exhibit varying optical properties depending on their crystal lattice and the vibrational property of the light (horizontally and/or vertically polarized, or unpolarized). Isotropic materials, such as glass or garnet, allow light to pass directly through them with the same retardation, regardless of the direction the light travels (68). Anisotropic materials, such as quartz, have different retardation values relative to their crystal lattice, and hence will alter the light as it passes through at different orientations (68, 69). This can be utilised to determine the compositional nature of the sample being examined.

For example, in plane-polarized light microscopy, where only one polarizing filter is used, pleochroic grains will change colour as a sample is reoriented (usually by rotating the stage) (68, 69). In crossed-polarized light microscopy, two polarizing filters are set at 90° relative to each other, allowing the throughput of only those light rays affected by interference effects of anisotropic materials (68). When anisotropic materials are viewed with crossed-polars, variations in colour (known as interference colours) indicate differences in birefringence, which may be due to thickness and/or composition differences (68). An examiner can use their expertise, aids such as the Michel-Levy chart, and known mineral characteristics to infer an identification of the materials present within a sample. An illustration of this is shown in Figure 1.8. PLM is also used commonly within other areas of forensic science, identifying and comparing artefacts such as fibres and glass in the same way (68, 69).

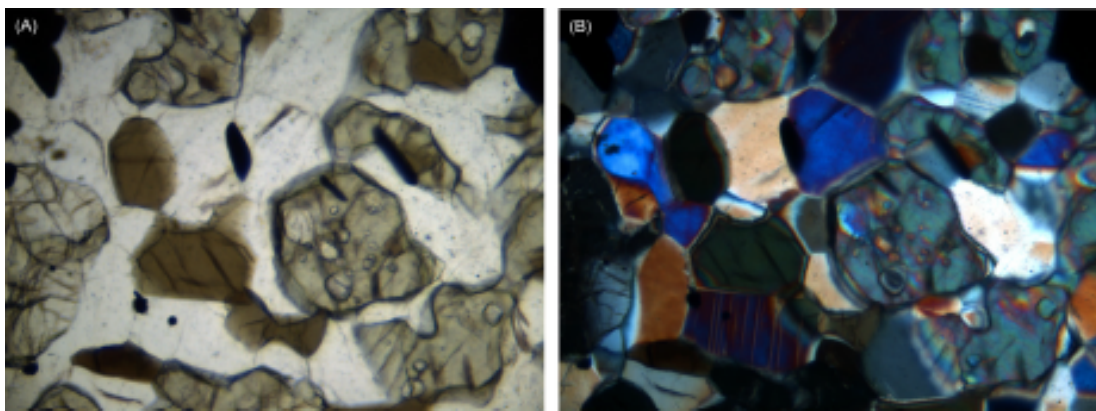


Figure 1.8 (a) Plane polarized light microscope image of a suspected metamorphic granulite; (b) cross-polarized light microscope image of the same thin section, showing interference colours ranging from first order whites to higher order blues, pinks, greens and oranges (from Pitts, Lewis, and Newland (46)). Hypersthene, augite, labradorite, quartz and miscellaneous amphiboles were indicated in the sample (field of view is 1.1 mm).

1.4.2.2 X-ray diffraction

X-ray diffraction (XRD) is the most widely implemented technique, and one of the most effective, for the identification of minerals and crystalline substances within forensic soil samples (7, 54, 56, 71). XRD analysis can be performed on fine-ground bulk samples, or on separated fractions within the soil, such as the fine-grained, clay fraction (7, 71, 72). In XRD, X-rays are directed at a soil sample. These X-rays are

diffracted by the lattice of crystalline materials to produce a characteristic pattern that displays the intensity of X-rays collected at different angles of diffraction (7, 56). The angle of diffraction is determined by the lattice spacing of the layers of atoms, also known as the 'd-spacing' (7). Each mineral and/or crystalline component produces a characteristic X-ray diffraction pattern, and as differing soils contain varying minerals, these patterns are overlaid on top of each other to produce a summation pattern that is representative of the soil as a whole (7). This diffraction pattern can therefore be used to qualitatively identify the mineral components present within the soil. In 2018, Prandel *et al.* demonstrated how Rietveld refinement of XRD data was able to characterize minerals present in deferrified clays from Brazil (72). As this was performed in a forensic context, a reduced amount of sample was utilised. Discrimination of four different sites and two horizons was achieved through principal component analysis (PCA), indicating mineral variance in kaolinite, quartz, anatase, gibbsite, and muscovite.

Additionally, XRD can be used to discriminate between mineral polymorphs, elements and their oxides, and mixtures of different crystalline substances (56, 71). The use of an internal or external standard can also allow for semi-quantitative analysis of the soil minerals (7, 71). In 2004, Ruffell and Wiltshire published a repeatability test comparing conventional XRD with quantitative XRD for determining the mineral composition of forensic soil samples within the context of a forensic investigation (7). An unknown soil collected from vehicle tyre treads was analysed and compared to 21 known samples collected from the greater Belfast area in Northern Ireland using PCA. Both XRD and quantitative XRD showed that the mineral proportions of the tyre tread sample matched only the correct location, however, quantitative XRD did fail to discriminate two locations that XRD showed were different. Hence, the authors suggested conjunctive use of both techniques to allow for maximum discrimination.

XRD has also been used extensively within forensic casework to discriminate between questioned and known samples. In 2008, Petraco *et al.* published a case report detailing an investigation involving a deceased female who was found floating in the East River in Manhattan, USA (1). XRD was utilised to compare soil samples

collected from the riverbank and local beaches to soil retrieved from a water-stained shoe in the husband's bedroom closet. Both the questioned and known specimens were shown to exhibit the same trace elemental composition, and confirmed to contain NaCl, and these results were presented as testimony at trial. Another case review was published by Melo *et al.* that utilised XRD for mineralogical analysis of soil samples involved in a murder case in Brazil (19). In 2018, a young female disappeared from Colombo after spending the night with the male suspect, and her body was later found 50 km away from the town. Soil samples were retrieved from the suspect's vehicle and compared with surface soil samples collected from the body disposal site, as he claimed he had not left the Colombo town. The XRD data showed that the samples recovered from the suspect's vehicle could be excluded as having come from the areas examined in Colombo town and indicated an association between the body disposal site and the vehicle due to the presence of quartz and K-feldspar within both samples.

Finally, there are also documented occurrences where information generated with XRD was used to help guide forensic investigators to particular regions of interest. XRD analysis of soil evidence was used to help police locate the bodies of two missing women in South Australia (56). Crime scene investigators searched the vehicle of the main suspect and located a blood-stained shovel, bracelet, and boots that were caked with pink, powdery soil. Soil scientists from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) were called in to analyse the soil adhering to these items, and XRD showed that the mineralogical composition of these soils was consistent with a mining area or quarry. The comprehensive data was compared against available Australian Soil Resource Information System (ASRIS) databases to determine regions that fit these criteria, and these areas were sampled and analysed to confirm an association. The indicated area was then monitored by investigators until the bodies were eventually uncovered through animal activity, resulting in a full confession from the suspect and conviction.

Further research involving the use of XRD for forensic soil analysis is explored in Chapter 5.

1.4.2.3 Inductively coupled plasma mass spectrometry

In more recent times, inductively coupled plasma mass spectrometry (ICP-MS) has become a well-established method for inorganic analysis in soil science and is now being utilised for the same purpose in forensics (27, 73). The technique can be employed for semi-quantitative estimation of the composition of the crystalline fraction of fine silt and clay (16). An acid-digested soil sample is nebulized into an inductively coupled plasma, where it is ionized, and then separated and quantified using a mass spectrometer (10). This method is capable of detecting elements from metals and mineral components present within the soil at concentrations down to parts per trillion (ppt) (16). However, the quantitative capabilities are often limited by incomplete silicate digestions, which generally require hazardous hydrofluoric acid for completion (11, 13, 73). The use of a laser ablation unit to introduce a small amount of solid sample directly (LA-ICP-MS) allows for solid samples to be analysed without the need for digestion, but detection limits are poorer, semi-quantitation requires matrix matched standards, and heterogeneity may cause issues (16, 73).

1.4.2.4 Scanning electron microscopy with energy dispersive X-ray spectroscopy

Another technique that is commonly utilised today in both mineralogy and forensic science is scanning electron microscopy (SEM) with energy dispersive X-ray spectroscopy (EDX) (17, 21, 62, 74). This method can provide detailed visual information on individual grains from soil samples, as well as bulk elemental composition information, allowing for inferred mineral identification and the characterization of native metals (24, 54, 71, 74). An illustration of this is shown in Figure 1.9. A high-energy beam of electrons is directed onto the soil sample for imaging using a SEM, which stimulates the emission of characteristic X-rays. The number of X-rays emitted from the sample and their associated energy can then be measured using an EDX. As the energies of the X-rays are characteristic of the electron orbitals within the emitting material, this allows the elemental composition of the soil sample to be qualitatively measured (17). If standards are available and samples are polished flat, quantitative compositions may also be determined (74). SEM-EDX only requires small quantities of soil, as these samples contain hundreds of

individual particles from which quantitative data can be obtained for comparison (71). However, this can sometimes be a disadvantage in forensics; when observing already biased samples microscopically, soils can show high levels of heterogeneity, which is hard to interpret in a broader sense to attain an overall picture of the bulk soil. More advanced electron techniques such as electron back-scattered diffraction (EBSD), quantitative evaluation of minerals by scanning electron microscopy (QEMSCAN), and microprobes have seen greater use in recent years, but are still specialized and not routine in their use (17).

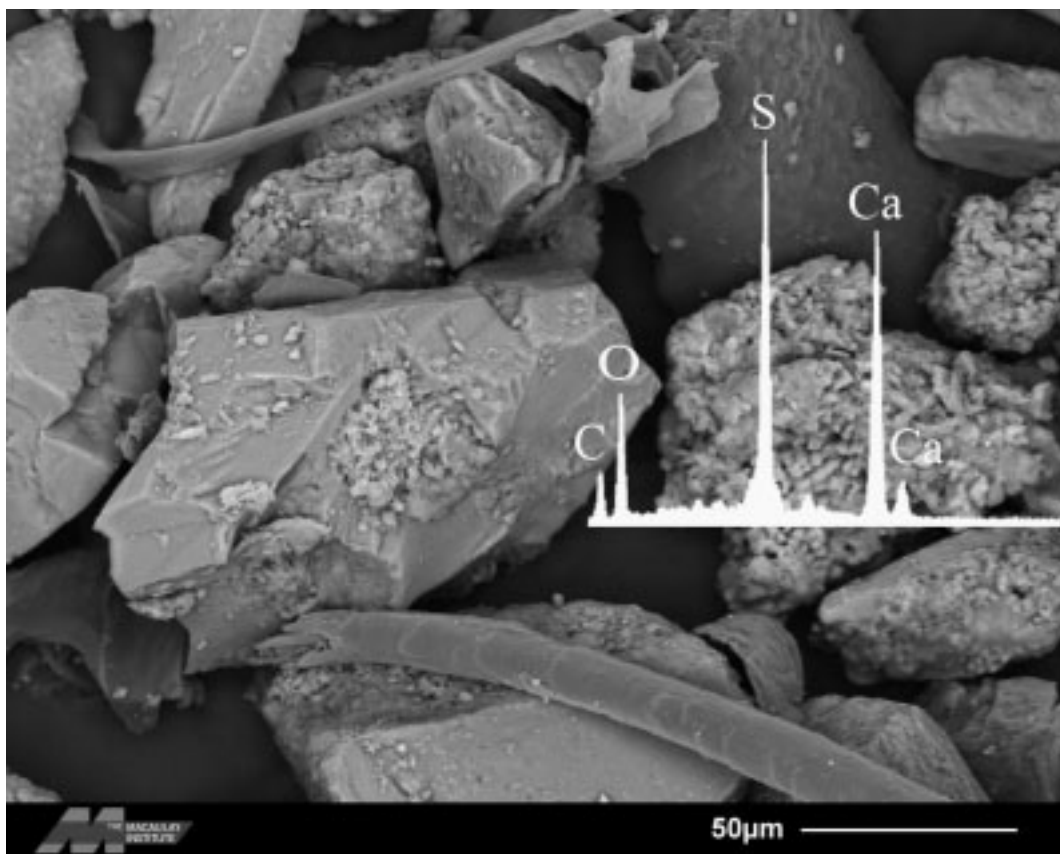


Figure 1.9 SEM image of soil particles showing the detailed textural information it may provide, along with chemical identification of the coatings of some grains determined by EDX (from Dawson and Hillier (17)).

1.4.2.5 X-ray fluorescence spectroscopy

An alternative to EDX is the use of X-ray fluorescence (XRF), which also utilises the emission of characteristic X-rays to determine bulk chemistry of a sample. Forensic scientists frequently use XRF for the elemental analysis of paints, liquids, and soil

samples (24, 62, 64). X-rays are directed at the soil sample and absorbed by inner shell electrons of the elements within the sample (13). This produces an excited ion, which spontaneously emits a characteristic X-ray fluorescence spectrum (13). XRF has better detection limits for transition metals than EDX, but still has issues with lighter elements below sodium and still relies on inferring mineral species from elemental data. Portable XRF systems have also recently been made available for field use, which are capable of detecting up to 25 elements simultaneously (13).

1.4.2.6 Chromatography

Gas chromatography (GC) is routinely used to identify wax esters and oils derived from plant matter from an agricultural science perspective (8, 13). Within soil samples, these hydrocarbons are extracted and then vaporized before interacting with the stationary phase that is coated on the walls of the GC column (13, 17, 71). This ideally causes each hydrocarbon to elute at a different time, and these times can be compared to known compounds for their identification, which in turn can indicate the types of plants that were present at the origin of the soil sample (8, 13, 17, 59, 71). Another method that can be used for this is high performance liquid chromatography (HPLC) or the similar ultra-performance liquid chromatography (UPLC), which separates and identifies these hydrocarbons in a similar way as GC, but in their liquid form (9, 32, 59). Both methods benefit from the use of a sensitive mass spectrometry detector to allow for surer identifications based on fragmentation patterns (GC-MS/LC-MS) (13, 22, 71).

1.4.2.7 Optical microscopy

For the visual identification of pollens, palynologists generally employ the use of optical microscopy (OM), which forensic scientists can also utilise to their advantage (17). Pollen and spores can be so small in size that they are not able to be detected by the naked eye, and are very resistant to removal from adhered objects, making them ideal forms of trace evidence (8). By observing pre-prepared pollen grains under a microscope (Figure 1.10), distinguishing features such as grain size and shape, the sculpturing of the outer coating, and number and size of the pores or

furrows can provide information that allows forensic scientists to identify the species of plants that were present at the origin of the soil sample (8, 15, 17). In cases where pollen grains are too small or difficult to see under a standard optical microscope, SEM can also be used.

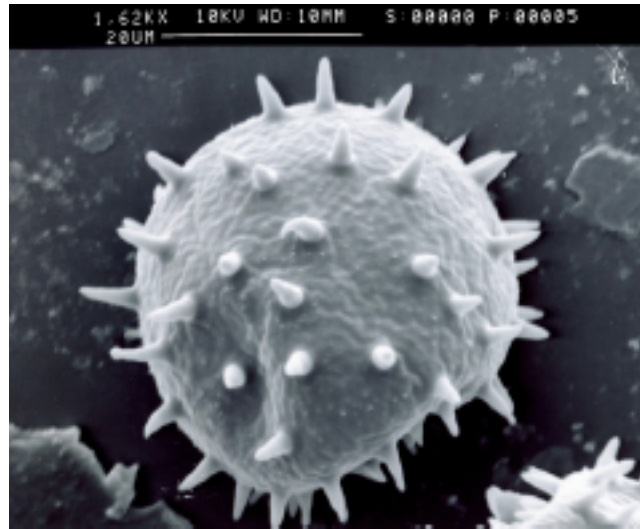


Figure 1.10 Photomicrograph of Nuphar water lily spore (from Bull, Parker, and Morgan (39)).

1.4.2.8 Isotope-ratio mass spectrometry

A technique that is commonly utilised today in geochemistry for samples with organic matter is isotope-ratio mass spectrometry (IRMS) (17). Compositional analysis of carbon isotopes within a soil sample can be undertaken to identify a soil's contributing flora and fauna. Samples are ionised, and then accelerated so that the resulting ions are separated according to their mass-to-charge ratios. As soil organic carbon is generally derived from the remains and decomposition residues of plants growing in situ, IRMS can utilise mass spectrometric methods to measure the relative abundance of stable carbon isotopes within a forensic soil sample to provide information on the type of ecological environment from which the sample originated (8, 57, 71). Other isotopic ratios, for elements such as nitrogen, hydrogen, oxygen, strontium, and lead may also be utilised for comparison of soil samples (57).

1.4.3 Emerging techniques for forensic soil analysis

In recent literature, additional methods for forensic soil analysis have been investigated, including microspectrophotometry for colour analysis (4, 15, 75), infrared and Raman spectroscopy for compositional analysis of soil samples (4, 33, 40, 64, 76, 77), and laser-induced breakdown spectroscopy (LIBS) and ICP atomic/optical emission spectrometry (ICP-AES/OES) for semi-quantitative estimation of elemental composition (11, 16, 62, 65, 73). These techniques are considered useful in certain cases, especially as part of a standard questioned versus known comparison, however, are still considered to be in the developmental stage as limited substantive validation and reliability testing has been shown. Some of these additional techniques have been further expanded on below.

1.4.3.1 Microspectrophotometry

Forensic scientists routinely use microspectrophotometry (MSP) for colour analysis of many different forms of trace evidence, including paint, fibres, and inks (4, 71). While the use of MSP has been recently applied to forensic soil samples too, it has not yet become widely adopted (4, 15, 17, 75). Soil samples are mounted on glass or quartz slides and presented to the microspectrophotometer (consisting of a microscope combined with a spectrophotometer) for analysis. Areas within the sample can be magnified and focussed using the microscope, before illumination with a controlled light source. The spectrophotometer then separates the light into specific wavelengths, and measures the intensity of the transmission or reflection of this light by the soil sample (71). The preferred method of illumination is generally transmitted light, due to the effects of surface angles and sample thickness on reflected light, however, this is not always possible with soil samples (71). The result is a spectrum displayed across the ultra-violet, visible, and near-infrared regions of the electromagnetic spectrum – an objective measure of colour. This spectrum can be further converted into numerical chromaticity values according to a range of different colour systems, such as the $L^*a^*b^*$ colour space, which is defined in further detail in Chapter 3.

Woods *et al.* demonstrated the use of MSP for the discrimination of Australian forensic soil samples, and best results were achieved through conversion of the MSP spectra to $L^*a^*b^*$ chromaticity coordinates prior to chemometric analysis (4). By utilising both MSP spectra and $L^*a^*b^*$ values, the 29 soil specimens analysed could be classified into 21 different groups, and pairwise comparisons resulted in 98% discrimination of the soils. However, results were affected by intra-sample variation, as soils can exhibit many colours when analysed microspectrally that may not be representative of the overall bulk colour of the soil. Hence, this technique was recommended only as a screening step prior to further analyses.

Further information involving the use of MSP and colour analysis for forensic soil examination is explored in Chapter 3.

1.4.3.2 Raman spectroscopy

Raman spectroscopy has been previously utilised in forensic science for the comparison of evidence, such as drugs (78, 79), paint (80, 81), and fibres (82, 83). Samples are irradiated with a monochromatic laser source, and electrons within the sample undergo excitations to virtual energy states (84, 85). As these electrons relax back down, they release energy characteristic of their vibrational modes (84, 85). Rayleigh scattering, where electrons are excited and relaxed to the ground state with no energy differential, is filtered out, while the Raman scattering, where electrons are relaxed to a different state than where they started, is dispersed onto a detector and recorded as a spectrum (84, 85). The Raman 'shift' may be either positive (Stokes) or negative (anti-Stokes), but generally it is the Stokes Raman scattering that is measured, as this is simpler to interpret and has a higher intensity (84, 85). This spectrum may then be used to infer information on the molecular structure within the sample. Novel methods such as surface-enhanced Raman spectroscopy (SERS) (surface-enhanced RS) and confocal Raman microscopy can also allow for enhanced signals and higher resolution imaging of trace materials (30, 84).

A major factor to consider when utilising Raman spectroscopy to analyse materials in a forensic context, such as soil, is the production of undesirable fluorescence

emission, which can mask signals within the spectra (84). This is one of the primary reasons that its application to forensic soils has been limited. In 2012, Edwards *et al.* published a study that demonstrated how Raman spectroscopy could be used to successfully discriminate between soil types based on both their mineralogy and the organic, water-soluble fractions of soils (77). This involved the development of oxidative preparation methods designed to remove part of the organic content of the soil in order to lessen the swamping effects of fluorescence. Due to the decreased fluorescence and improved signal-to-noise ratio, it was possible to observe some of the mineralogy present in the soil samples, which could potentially allow for their differentiation in a forensic context.

Kammrath *et al.* also published a case study in 2017 utilising a novel technique that gathered detailed morphological information on soil particle size and distribution using automated particle imaging, before probing the molecular chemistry of specific particles of interest using Raman spectroscopy (29). By obtaining Raman spectra of individual soil grains, the humic content within the soil samples was avoided, and therefore the issue of undesirable fluorescence was not encountered. PCA was then performed on the particle count data to successfully discriminate between soil samples from different sites along a single road in Connecticut, USA, highlighting the sensitivity of the technique.

Further information involving the use of Raman spectroscopy for forensic soil analysis is explored in Chapter 4.

1.4.3.3 Infrared spectroscopy

Infrared (IR) spectroscopy has been widely explored for the analysis of many different forms of forensic trace evidence, including automotive paints (86), explosives (87), makeup (88), and polymer-based identity cards (89). Despite its well-known application for the analysis of minerals (90), application to forensic soil has been limited until recently, likely due to the advancements made in portable IR technology and chemometric methods (91). In IR spectroscopy, IR radiation is directed onto a sample, which exhibits unique combinations of bonds depending on

its composition (85, 92). This radiation can be absorbed by molecules within the sample, dependent on the transition energy of the vibrating, stretching, and bending of the bonds within those molecules (85, 91, 92). The resulting absorbance or transmittance is then detected by the spectrometer, and in the case of Fourier transform infrared (FTIR), undergoes transformation using a mathematical algorithm to produce a spectrum as a function of wavenumber (92). Hence, this spectrum can indicate the molecular structure of a sample, or be compared to other samples or standards to determine similarities or differences (91, 92). The most commonly applied IR methods are attenuated total reflectance Fourier transform infrared (ATR-FTIR), which utilises a sampling accessory to reduce the need for complex sample preparation, and diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS), which can be used on powder samples with limited or no sample preparation (30, 71, 93). However, IR spectra can be challenging to accurately interpret for complex mixtures like soil, due to the overlapping of bands characteristic of different molecules within the sample.

The use of IR spectroscopy for forensic soil examination was initially applied for analysis of the organic content within soil. In 2000, Cox *et al.* demonstrated the use of a novel FTIR spectroscopic technique for discrimination of similarly coloured soils (33); this involved analysing the soil via FTIR, oxidatively pyrolyzing the sample so that all organics were degraded, reanalysing the sample, and then subtracting this FTIR spectrum from the original one that contained the organic material. This resultant spectrum therefore represented the organic portion of the sample, and could be used to increase the discrimination between soils that are inorganically similar.

More recently, it has been used to characterise the entire soil sample, incorporating analysis of both the organic and inorganic components. In 2014, Woods *et al.* demonstrated that ATR-FTIR spectroscopy combined with PCA could successfully be used to discriminate between forensic soil samples from Canberra, Australia (4). The 29 soil specimens analysed could be classified into 28 different groups, and pairwise comparisons resulted in 99.7% discrimination of soils. However, this technique was recommended as an initial screening step prior to more detailed analyses. Chauhan

et al. followed up this study in 2018 with further investigation into the effects of sample preparation prior to ATR-FTIR analysis, as well as the application of additional chemometric methods (discriminant analysis) to successfully predict a soil's origin (76). Models produced were able to both discriminate and classify 100% of Indian soil samples based on their original location. It was also shown that sieving and heating of soil can affect the spectral features considerably, so should be avoided in forensic casework if possible.

A study was then conducted by Xu *et al.* in 2020 that reinforced the conclusions made by Woods *et al.*; while ATR-FTIR combined with PCA was able to successfully discriminate between most soils, it was difficult to distinguish between some similar soils using ATR-FTIR data alone, so its use alongside other analysis methods is recommended (65). One such example of this was published in 2020 by Prandel *et al.*, who demonstrated the use of PCA on combined ATR-FTIR and energy-dispersive X-ray fluorescence data (64). This methodology was able to characterise a small amount of the clay fraction of soil. However, only specific bands within the ATR-FTIR spectra were hand-picked for chemometric analysis.

Further information involving the use of IR spectroscopy for forensic soil analysis is explored in Chapter 4.

1.4.3.4 Microbiome analysis

Another option to analyse the organic component of soil is to examine the microbiome. Only a small amount of analysis has been attempted and published in this field to date, and most has focused on the bacterial DNA/rNA present within soils, rather than the proteins (2, 3, 12, 18, 31, 40, 66). This method is based on the specific fingerprint of genetic material differing between strains of bacteria, which can be analysed to indicate which species of bacteria are present in the soil. However, soil DNA profiles can be considerably altered by many factors, including annual seasons, human activity, sample storage and handling, and the presence of human decomposition (17, 40, 66). Since this research is very recent, there are no widely accepted, standard methods in place for examination of the microbiome in forensic

soils as of yet (2, 35). This is an area that has the potential to provide a powerful means for discriminating soil samples with similar mineralogy or linking them to a common origin, and should be explored further (13, 34).

1.5 Use of the quartz-recovered fine fraction

Most of the methods that make use of bulk soils for analysis are impractical for soil samples commonly encountered in forensic casework, due to their limited size (3, 74, 94). These methods are also not suitable for analysis of soils that have minimal variation in their bulk chemistry, like the quartz sands found on the Swan Coastal Plain (54, 94, 95). In many approaches to forensic soil analysis, several techniques are undertaken on separated fractions of soil, allowing further differentiation compared to the bulk chemistry (10, 54, 73, 74).

The clay and fine silt fraction within soil samples has generally been the focal point of forensic analysis, rather than the sand fraction (20, 72, 96, 97). This is due to the smaller particle size, which aids in transferability and adherence to evidence and provides a higher number of particles for analysis from a much smaller sample (19, 20, 94). The fine clay fraction also exhibits greater variability in secondary minerals, and as such, can be a better indicator of provenance than the other fractions which are primarily comprised of weathered material derived from the parent rock (10, 19, 20, 94). In Swan Coastal Plain soils however, the clay and fine silt fraction is present in very limited amounts (1 – 5% of the soil mass) and is typically found as coatings on the sand grains, making analysis of this fraction more complicated (51, 94). Because of the difficulty in separating these clay coatings from the dominant sand fraction, most of the suggested methods that analyse the clay fraction are not able to be utilised for examination of Swan Coastal Plain soils, as they are designed for loose, separable clay fractions within the bulk soil.

Sometimes the sand fraction is able to be forensically analysed if the clay fraction available is inadequate (54). Recently, Melo *et al.* investigated the use of particle size distribution, XRD, and SEM-EDX coupled with chemometric methods to characterise and discriminate between the sand fractions of subtropical soils (54). While this

research made use of the sand fraction, which is the predominant fraction in the dry, sandy soils found in Western Australia, their method was deemed unsuitable for effective discrimination. This study highlights the difficulties faced with these types of highly weathered soils that have a homogeneous, quartz sand fraction that cannot be used to differentiate between them. Traditional methods such as XRD are not appropriate for use with the sand fraction, and hence these sandy soils require more detailed examinations to be able to distinguish between different sources (54).

For these reasons, Pitts and Clarke developed a novel method which allows for the differentiation of quartz dominated, minute, fractionated soil samples, like those that originate from the Swan Coastal Plain (94), and this has been used for forensic casework since the 1990s (97). This method, illustrated in Figure 1.11, involves extraction of the clay and fine silt coatings found on quartz grains (primary and secondary minerals <20 µm) termed the 'quartz-recovered fine fraction', which can then be further analysed and compared (94).

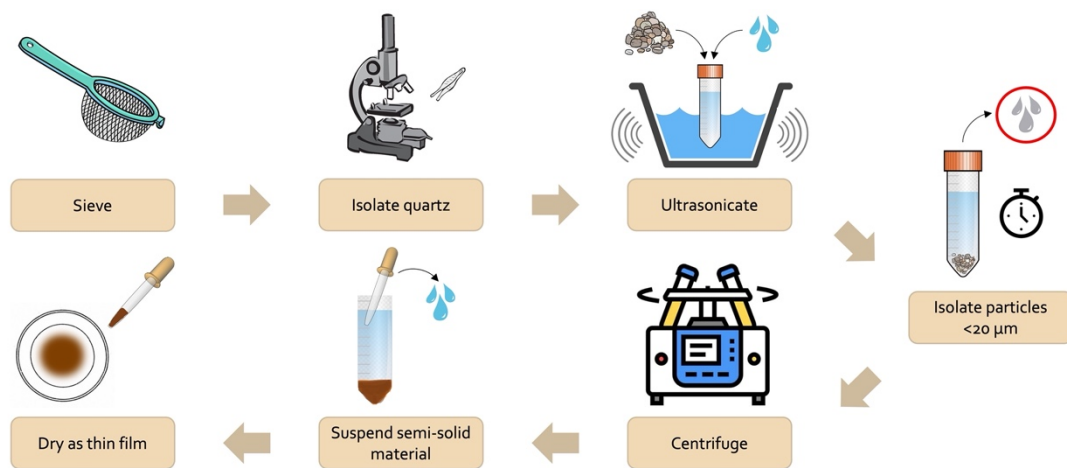


Figure 1.11 Infographic based on method published by Pitts & Clarke (94): The forensic discrimination of quartz sands from the Swan Coastal Plain, Western Australia.

Because this removes the quartz grains from the sample, reproducible XRD data that highlights mineral variation is able to be obtained from forensic-sized samples of sand (94). Although it has only been tested in combination with XRD and Raman spectroscopy so far, this technique has successfully been used within forensic casework and presented as evidence in court, as well as for proficiency testing within

the laboratory setting (94, 98). While this method was developed with soils from the Swan Coastal Plain in mind, it is also applicable to sandy soils from wider locations and potentially even different soil types from other geological environments that exhibit low levels of clay and organic matter and high quantities of quartz grains (94). Exploration of the quartz-recovered fine fraction from within sandy soils is therefore the focus of the work contained within this thesis, with an emphasis on the use of both established and emerging techniques combined with chemometric methods for the analysis of this data.

1.6 Interpretation of results

While there are many valuable techniques for analysing soils and identifying their components, challenges arise when applying these methods to forensic soil cases, which not only require robust analytical techniques but also a reliable measure of how similar a questioned sample is to a known source (17). As with all forensic evidence, forensic soil analysis requires further interpretation of data to make comparisons between questioned versus known samples and reach conclusions about a sample's geographical origin.

The interpretation of evidence has always been a significant issue within the wider forensic community (17), and translating whether an association between samples is actually probative is of primary concern. Accurate and reliable methods for evidence examination, interpretation, and communication of findings are essential for a justice system to be wholly effective, however, to date there is minimal scientific research that works to evaluate the significance of a forensic match once it has been achieved (26, 63, 99-101). This is a problem expressed commonly by other researchers such as Dr. Paul L. Kirk who, when referring to forensic fibre analysis, stated:

“...it is possible to identify all types of textile fibre with exactness, through a variety of methods.... However, the probabilistic value of a fibre transfer between two sets of clothing, as in a crime, is still a matter of controversy, even though such transfers constitute one of the more valuable types of evidence.” - Kirk, P.L., Science, 1963, 140: 367-370.

This theme has arisen more often recently, as discussed in several reports over the last decade, such as the National Academies of Science's (NAS) report on forensic science research (99), the House of Lords' report on forensic science and the criminal justice system (101), and many peer-reviewed articles (102-105).

Additionally, there is a constant requirement to ensure the scientific foundations behind the methods utilised for forensic analysis of evidence, to allow for objective analysis and interpretation based on statistically valid comparisons, rather than subjective comparisons (99-101, 106). Many forensic disciplines rely on visual comparison of chemical data, however, it is now recognised that examiners can be prone to cognitive biases that may affect their conclusions, such as prior expectations of results or knowledge of investigative details (105). Recently, the President's Council of Advisors on Science and Technology (PCAST) published a report recommending the urgent development of additional, more efficient techniques for objective forensic examination of the six most commonly encountered forms of trace evidence – fingerprints, DNA, hair, firearms, bite marks, and footwear impressions (106). This recommendation could be applied to all forensic feature-comparison methods, including soil and geological evidence.

One way of addressing this problem is by using a Bayesian approach, which is currently used in Australia, Europe, and New Zealand for the statistical interpretation of forensic evidence such as glass. The Bayesian approach describes the probability of the presented evidence being observed given two competing hypotheses (e.g., the defendant is guilty versus the defendant is not guilty), and prior knowledge of related factors, such as the rarity of certain features (100, 107, 108). In simple terms, Bayes' theorem shows us the effect that the scientific evidence has on the odds that the defendant either did or did not commit the crime (26, 92, 108). This approach has limitations however, when applied to complex forms of trace evidence such as soil, that have unknown quantities, frequencies, and features (17).

1.6.1 Chemometrics

Another approach that assists in a more objective and transparent interpretation of soil evidence is utilising multi-variate statistical methods, or chemometrics, to evaluate the relationships between samples (6, 10, 107). Data can be analysed using unsupervised techniques to establish patterns or correlations, or supervised techniques that require pre-established groupings (6, 10, 76). These methods can provide a statistical measure of similarity between samples in a way that is accepted by the scientific community (10), and have previously been used to successfully analyse chemical data and achieve source discrimination of other forms of forensic evidence, such as fibres, automotive paints, identity documents, and explosives (80, 87, 89, 107). For these reasons, data arising from chemometric analysis has been accepted as forensic evidence presented in court (109). The statistical measures generated through chemometric analysis could even be further integrated within the standard Bayesian approach to produce an accurate, objective interpretation of the forensic evidence at hand. An overview of the techniques applied throughout this thesis is provided below.

1.6.1.1 Principal component analysis

Principal component analysis (PCA) is an unsupervised technique, in that it does not require predetermined groupings to process information (105). When PCA is used in conjunction with detailed chemical analysis methods, it can quickly simplify data and visualise its distribution to provide a quantitative measure of similarity between samples (4, 10, 13, 29, 76, 107). It works by linearly transforming multidimensional datasets into smaller numbers of orthogonal variables that represent most of the variation within the data, called principal components (PCs) (6, 10, 30, 65, 76). The number of PCs generated cannot be higher than the number of original variables, and the first PC will always represent the largest proportion of total variance, with each additional PC declining in its encapsulated variance (6, 65).

This data can be projected into 2D or 3D space using the component scores, with the principal components as the axes, in order to visualise the clustering or separation

exhibited by samples based on their similarities (10, 30). The scores can also be exported into other chemometric techniques, such as linear discriminant analysis, rather than using the original multidimensional variables (30). In addition, examination of loadings plots can provide information on the chemical factors that allow for differentiation between samples (65, 105). PCA can prove especially useful when undertaking common 'questioned versus known' comparisons, to highlight subtle similarities or differences between samples that could not be detected through human interpretation of the data.

1.6.1.2 Linear discriminant analysis

Linear discriminant analysis (LDA) is a supervised, eigen-analysis technique that requires pre-established groupings in order to best discriminate between groups of samples (10, 105). It linearly combines the variable set to determine optimal canonical axes that minimise the variance within sample groups and maximise the variance between sample groups across a training set (6, 10, 76). The model then classifies these training set samples into the known groups, outputting a confusion matrix that summarises the correct and incorrect classifications and hence provides a classification accuracy percentage (105). The resulting data can then be used to predict groupings for unknown samples based on whether they share the same characteristics as the training set samples (10, 11, 38, 76, 107). Unfortunately, LDA does have one notable disadvantage in that it involves matrix inversion, meaning that the number of variables must be smaller than the number of samples involved, so it is usually carried out after data reduction methods such as PCA; if limited samples are available then not all of the PCs that account for variation within the data can be incorporated within the model, causing some variation to be lost (30, 105).

There are several different methods that can be used to validate the predictive model for LDA, with the most common being the 'leave one out' and 'independent test set' methods. The 'leave one out' process involves removing one sample from the dataset, using all of the other samples to build the model, and then projecting the remaining sample back onto the model to classify its grouping (66). This is then repeated with each individual sample in the dataset, and the classification rates are

combined to provide an overall validation accuracy (66). The 'independent test set' method involves the use of a completely independent dataset for validating the model, i.e., the samples that are predicted onto the model are never used to build the model, so it is only performed once. This is especially suited for questioned versus known comparisons, where the data from samples of known origin can be used to create the model, and unknown samples can be predicted onto the model to assess their grouping. For this reason, as well as the fact that 'leave one out' validation can greatly overestimate model performance (87, 105), 'independent test set' LDA has been the method utilised within this thesis.

1.6.1.3 Discrimination of forensic soils in literature using chemometrics

Many studies have utilised chemometric methods for interpretation of soil evidence (2, 4, 5, 7, 8, 20, 29, 54, 57, 64, 65, 72, 74). While PCA has been the most commonly applied method, many other types of analyses have been explored, including hierarchical cluster analysis (HCA), Bray-Curtis similarities, and partial least squares.

In 2011, Jantzi and Almirall demonstrated the use of quantitative LIBS in combination with chemometric methods to successfully discriminate between bulk soil samples from Florida, USA, for forensic purposes (11). PCA was used to show clear grouping of samples based on location, and 'leave one out' LDA obtained a discrimination accuracy of 99.4%. This work was then expanded on by the original authors in 2014, where they investigated different sample preparation methods that would allow for success with smaller sample sizes when analysed using LIBS (16). These results were again subject to PCA and LDA and compared to those obtained through the use of LA-ICP-MS; all methods allowed for over 94% discrimination accuracy, showcasing the potential that both LIBS and LA-ICP-MS hold for discrimination of forensic soils when combined with chemometric analysis.

Lee *et al.* performed chemometric methods such as multiple comparison (MC), PCA, LDA, and HCA on data obtained from thermally assisted hydrolysis and methylation (THM) of soil organic matter using pyrolysis GC-MS (PyGC-MS) (6). Analysis was conducted on 40 different soils, and these were shown to cluster into six groups

through chemometric analysis. The PCA loadings also identified 11 marker compounds within the soil organic matter that were responsible for the majority of variance within the dataset, and these could be related to each of the clusters formed; the clusters were differentiated by the presence of lignin, fatty acids, and urea within the organic fraction of soils.

Reidy *et al.* conducted a mock crime scene scenario in which students were asked to conduct a comparative analysis between eight bulk soils from different sites within Mississippi, USA, using ICP-MS combined with PCA and LDA (10). While PCA allowed for clustering of the soils based on their overall region, it showed overlap between samples that were located approximately 3 kilometres apart, highlighting the comparable elemental profiles of soils derived from the same parent material despite occurring across wide-spread areas. LDA allowed for improved discrimination, showing clusters based on sites located as close as 50 metres apart, and correct classification of 100% of the unknown soil samples. The results suggested that elemental analysis of soil by ICP-MS can distinguish soils from different locations, however, success is dependent on an appropriate choice of chemometric interpretation.

More recently, Chauhan *et al.* investigated the use of ATR-FTIR spectroscopy and chemometrics to characterize, differentiate, and classify both surface and sub-surface soil samples collected from various regions across North-Western India (76). Sample preparation methods such as heating and sieving were shown to considerably affect the results, with some components in the soil experiencing degradation or removal through these processes. The use of PCA improved pairwise discrimination accuracy from 99.35% for surface samples and 97.38% for sub-surface samples, to 100% for both. Canonical discriminant function analysis (CDFA) was also used to classify soils to their original geographical regions, with 100% of samples achieving correct classification through 'leave one out' cross-validation. While ATR-FTIR spectroscopy was shown to be successful for characterising and discriminating soil samples, results benefitted from further interpretation using chemometric methods.

1.6.1.4 Gaps in the literature

While the use of chemometrics on soil evidence is prevalent within published literature, chemometric analysis has never been demonstrated on data obtained from methods analysing the quartz-recovered fine fraction within soils, so its suitability for use with different soil types, such as the dry quartz sands found within Western Australia, has not been shown.

Additionally, most published studies have previously focused on a single approach for the analysis of samples, overlooking the detail that in standard forensic practice it is far more common to utilise multiple examination methods (6, 62). Soil analysis, like other areas of forensic analysis, is typically carried out in a sequence, utilising the least destructive methods first and gathering further information as the sequence progresses (15, 73). Soil contains many different features that can allow for its differentiation and not all of these can be detected by one singular method alone, and as such, benefits from a multidisciplinary forensic approach (8, 38, 62). It would therefore be more realistic to look at a suite of instrumental methods, particularly ones that are complementary in the information they provide, and experiment with combining them in a sequence (15, 17, 73). This would enhance the value of the soil evidence and allow for increased discrimination or association (4, 17, 62). The use of chemometrics should be demonstrated alongside each stage of the sequence, including visual examination, to highlight how it can provide investigators with supporting information.

1.7 Aims and overview

In this study, investigations explore the use of well-established spectroscopic techniques to further develop a multi-faceted approach for forensic analysis of the quartz-recovered fine fraction of soils. This data is used in combination with PCA and LDA to demonstrate how chemometrics can be utilised for the objective characterisation and differentiation of sandy soil samples for forensic purposes. This approach, demonstrated in Figure 1.12, can be used by forensic laboratories to help provide better evidence of association in a courtroom context.

This can be broken down into four overall aims of the project, which are:

- Demonstrate the use of the quartz-recovered fine fraction for the forensic analysis of sandy soils with techniques utilised for the examination of other forms of forensic trace evidence, such as MSP, Raman spectroscopy, ATR-FTIR spectroscopy, and XRD.
- Demonstrate how chemometrics can be used in combination with data obtained from analysis of the quartz-recovered fine fraction to objectively interpret soil evidence.
- Develop a method that can be applied to sandy soils that combines multiple spectroscopic techniques in sequence, with chemometrics performed alongside each technique, to discriminate between soils for forensic purposes.
- Test the capabilities of the developed method through the application of a blinded case simulation, in order to assess its suitability for use within the context of forensic casework.

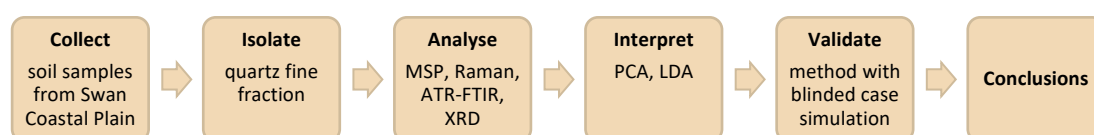


Figure 1.12 Flow diagram showing a summary of the overall methodology undertaken throughout this thesis.

Chapter 2 begins with a discussion of the experimental considerations for collecting, preparing, and conducting analysis of soil samples from the Swan Coastal Plain in Perth, Western Australia. An overview of the method for isolation of the quartz-recovered fine fraction is provided. Instrumental details and statistical methods used throughout subsequent chapters are also outlined.

Chapters 3, 4, and 5 all follow a similar methodology, exploring the use of several analysis techniques in combination with chemometric methods to discriminate between sandy soils from different locations; Chapter 3 focuses on MSP, Chapter 4

focuses on vibrational spectroscopy (Raman and ATR-FTIR), and Chapter 5 focuses on XRD and the use of both older and newer instrumentation to assess the reproducibility of the method. MSP and ATR-FTIR spectroscopy have previously not been performed on the quartz-recovered fine fractions of soils. As such, this project aims to provide information on whether these techniques can be used for the forensic analysis of specifically sandy soils. Whilst Raman spectroscopy has been investigated, it was unsuccessful at obtaining chemical information from the quartz-recovered fine fractions of soils (98), so alternate instrumental and experimental conditions are explored. The use of XRD on the quartz-recovered fine fractions has been demonstrated to successfully allow for mineral identification and differentiation (94, 98). This project therefore aims to expand on this work by reinforcing the reproducibility of the approach on new soil samples, interpreting the results using chemometrics, and evaluating the effects of different instrumentation on the results obtained. PCA is applied to data produced through all these methods of analysis to provide a quantitative measure of similarity between soil samples based on attributes such as their location, visual appearance, dune system, or land usage. LDA is then used to maximise the differentiation between soils from different locations and predict their most likely source of origin. The respective results of these chapters can be combined into a sequence to build on the information obtained and improve the overall discrimination of soils.

Chapter 6 then uses the approach developed in Chapters 3 – 5 and applies it to a blinded case simulation. A simulated *suspect recovered soil* is provided and compared to four soil samples from known origins in an attempt to associate samples of common origin and examine the discrimination of samples from other locations. Visual inspection, MSP, ATR-FTIR, and XRD are all utilised in sequence to provide complementary information on the soil samples, and PCA and LDA are performed alongside each technique to interpret the data and increase the level of discrimination achieved. This case simulation was designed by an experienced forensic soil examiner to imitate a challenging casework scenario, with samples chosen that are very similar to each other, representative of the types of samples that may be collected as a part of difficult real forensic investigation. Hence, the case

simulation is the culmination of the previous chapters and highlights the application of this work to forensic soil analysis.

Chapter 2. Experimental Methods

Portions of this chapter have been published, or submitted for publication, in the following articles:

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics for the forensic analysis of Western Australian sandy soils." *Forensic Chemistry*, 2022. 28: 100412.

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics: Application to simulated forensic soil casework." *Forensic Chemistry*, 2023. 33: 100481.

T. G. Newland, K. Pitts, and S. W. Lewis. "Negative result: Application of Raman spectroscopy to the forensic analysis of an arid, sandy, soil." *Forensic Science International: Reports*, 2022. Submitted for publication.

2.1 Introduction

This chapter summarises the experimental methods utilised throughout this study. Details of sample collection and storage, preparation of samples for analysis, and instrumental and chemometric methods are outlined, with any variations to the experimental conditions specified within the relevant chapter as required.

2.2 Collection of soil samples

Soil samples were collected from 23 locations across the Perth (Western Australia) metropolitan region (Figure 2.1), with the locations specifically chosen to allow for a selection of samples from differing dune and plant systems, as well as land usage. These locations were based on sites that had previously been sampled by ChemCentre in 1994 (36), and again in 2004 during a National Institute of Forensic Science (NIFS) funded project (97), as detailed in Bastian (110) and Pitts (98).

Sampling was carried out with respect to the considerations suggested by Pye (61), by first gently clearing the area of large organic material such as leaf litter or mulch, placing a wooden frame (30 x 30 cm to replicate the size of a footprint) onto the sample surface, and then using a small shovel to collect approximately 30 mL of the first 0–5 cm in depth of soil from one corner of the grid into a clean plastic container (polypropylene urine container, Sarstedt). This process was repeated in the other three corners and the centre of the area (Figure 2.2), producing a total of five soil samples per location. The GPS coordinates of all locations were recorded, with information on the accompanying geology, botanical influences, and land usage surrounding the sites. All samples were initially stored in a -20°C freezer and then freeze-dried (John Morris Alpha 1-4 LDplus) within two weeks of collection to minimise degradation.

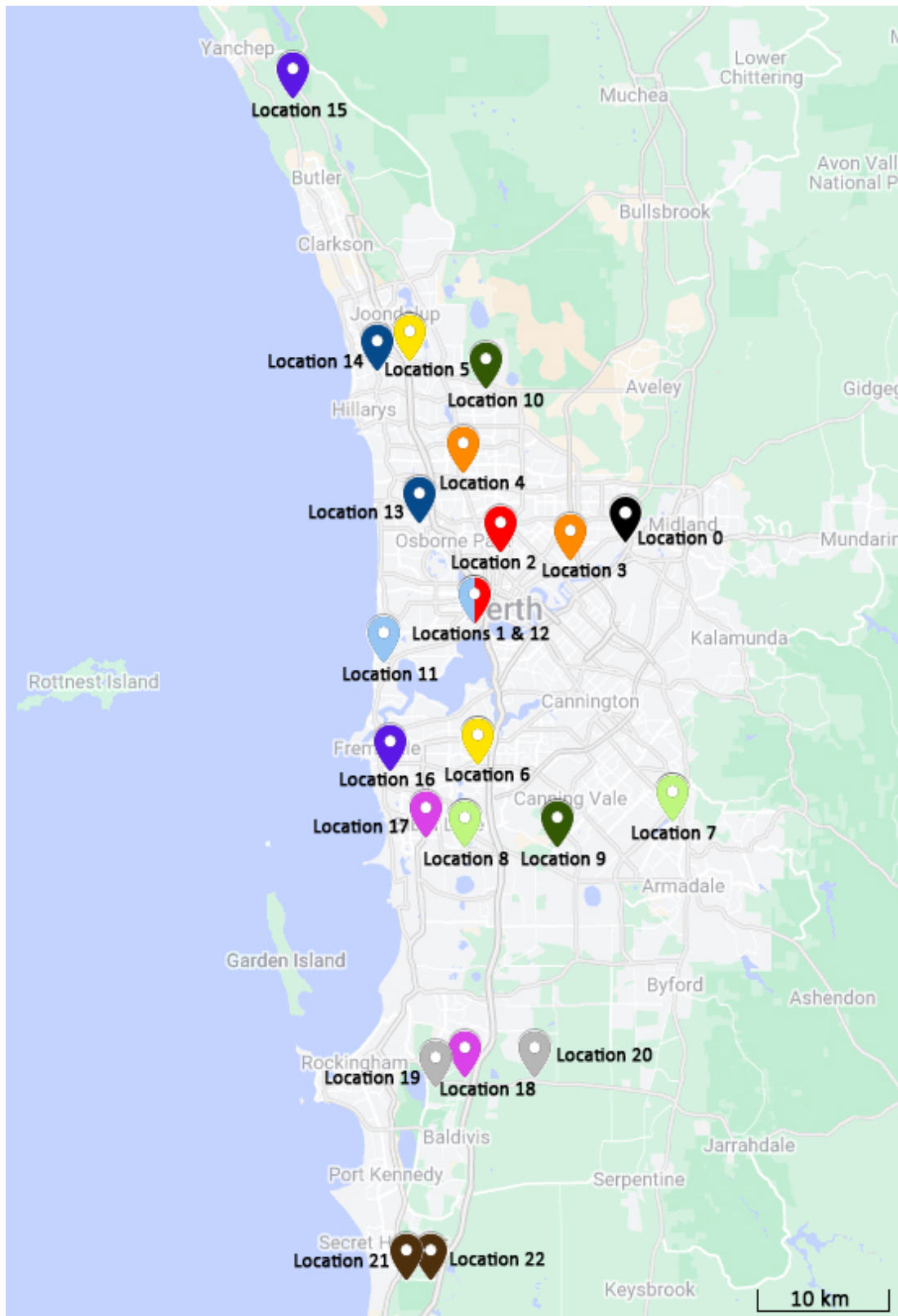


Figure 2.1 Map of the Perth metropolitan region in Western Australia, illustrating the 23 locations where the soil was sampled for this study.

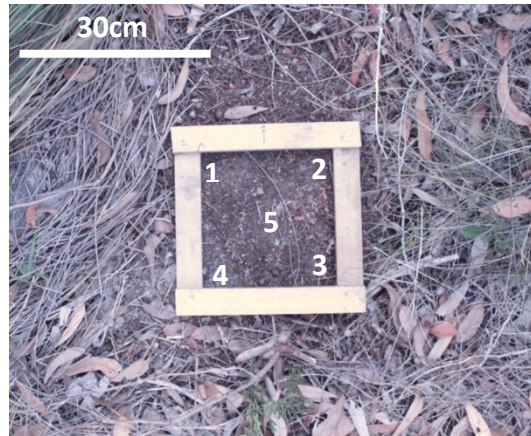


Figure 2.2 The wooden frame used to mark out the areas for collection of the soil samples, illustrated in situ on the soil surface after clearing of the bulky organic material. The five points of sampling are annotated within the frame.

2.3 Sample selection and identification

2.3.1 Primary investigations (Chapters 3 – 5)

Two soil samples, out of the five samples collected, were analysed from each of the 23 locations, except for Location 0 (full details outlined in Table 2.1). Prior to sieving and quartz recovery, samples were visually examined to note their overall colour and texture. Figure 2.3 shows a visual representation of the soil colour for nine of the sample locations included within this study. It should be noted that the assigned visual appearances for each sample are intended as a broad colour descriptor to allow for general groupings of the soils.

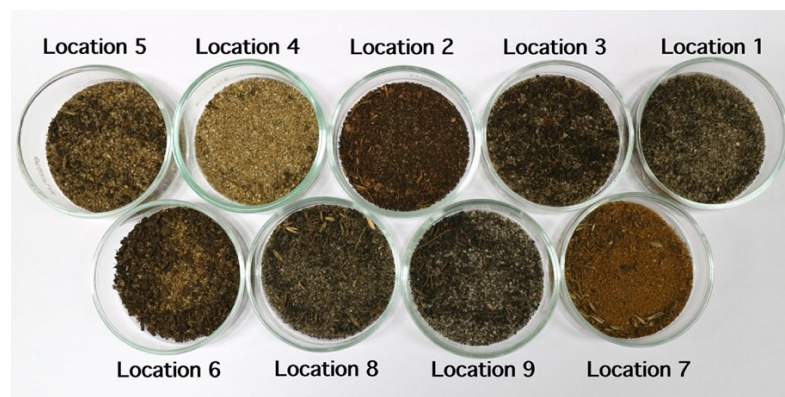


Figure 2.3 Soil samples obtained from nine of the different locations used throughout this study, illustrating the visual colour differences.

Table 2.1 The soil samples used in trials throughout this study, with their associated locations and visual descriptions. Two samples were analysed from each location, except for Location 0 (due to limited sample size). *Bushland is shortened to “bush” in annotated figures throughout this thesis.

Soil Sample	Location	Suburb	Dune System	Visual Appearance	Location Type	GPS Coordinates
0	0	Bassendean	Pinjarra Plain	Grey clay	Park (riverside)	(-31.9062577, 115.9609075)
1a, 1b	1	Kings Park	Spearwood	Grey sand	Bushland* (near park)	(-31.9646960, 115.8357950)
2a, 2b	2	Coolbinia	Spearwood	Brown sand and mulch	Garden (manicured park)	(-31.9138106, 115.8539786)
3a, 3b	3	Bayswater	Bassendean	Grey sand	Park	(-31.9176633, 115.9184918)
4a, 4b	4	Balga	Spearwood	Yellow sand	Verge	(-31.8549414, 115.8258240)
5a, 5b	5	Edgewater	Spearwood	Yellow/brown sand	Bushland*	(-31.7741340, 115.7732363)
6a, 6b	6	Murdoch University	Spearwood	Yellow/brown sand	Garden (managed)	(-32.0652384, 115.8391782)
7a, 7b	7	Champion Lakes	Pinjarra Plain	Red sand	Bushland*	(-32.1059759, 116.0022922)
8a, 8b	8	Yangebup	Spearwood	Grey sand	Bushland*	(-32.1238569, 115.8286284)
9a, 9b	9	Banjup	Bassendean	Grey sand	Bushland*	(-32.1243506, 115.9056052)
10a, 10b	10	Wangara	Spearwood	Grey sand	Cleared land (industrial)	(-31.7940448, 115.8425828)
11a, 11b	11	Cottesloe	Spearwood	Yellow sand	Verge	(-31.9900814, 115.7595074)
12a, 12b	12	Kings Park	Spearwood	Grey sand	Bushland* (near park)	(-31.9644730, 115.8360660)
13a, 13b	13	Innaloo	Spearwood	Grey sand	Park	(-31.8919048, 115.7914945)
14a, 14b	14	Kallaroo	Spearwood	Yellow/brown sand	Bushland* (near park)	(-31.7815572, 115.7537142)
15a, 15b	15	Eglinton	Quindalup	Yellow/brown sand	Dunes	(-31.5854972, 115.6820036)
16a, 16b	16	Beaconsfield	Spearwood	Yellow/brown sand	Garden (manicured park)	(-32.0687958, 115.7629386)
17a, 17b	17	Bibra Lake	Spearwood	Yellow/brown sand	Bushland*	(-32.1183926, 115.7948083)
18a, 18b	18	Baldivis	Spearwood	Light grey sand	Bushland*	(-32.2894756, 115.8276345)
19a, 19b	19	Baldivis	Spearwood	Yellow/brown sand	Bushland*	(-32.2973788, 115.8055882)

20a, 20b	20	Oldbury	Mixed (Spearwood/ Bassendean)	Brown sand and mulch	Bushland*	(-32.2894151, 115.8885882)
21a, 21b	21	Karnup	Spearwood	Brown sand and mulch	Bushland*	(-32.4299889, 115.7773041)
22a, 22b	22	Karnup	Spearwood	Grey sand	Bushland*	(-32.4341762, 115.7991339)

Table 2.2 The soil samples used in the blinded case simulation, with their associated known locations and visual descriptions.

Soil Sample	Location	Suburb	Dune System	Appearance	Location Type	GPS Coordinates
BS1a, BS1b	Alibi Site 1	North Perth	Spearwood	Grey sand	Manicured park	(-31.9180244, 115.8460303)
BS2a, BS2b	Suspect Recovered	Unknown	Unknown	Grey sand	Unknown	Unknown
BS3a, BS3b	Crime Scene	Leederville	Spearwood	Grey sand	Verge	(-31.9318972, 115.8437659)
BS4a, BS4b	Alibi Site 2	Wembley Downs	Spearwood	Grey sand	Verge / median strip	(-31.9205267, 115.7781723)
BS5a, BS5b	Potential Site of Interest	Hazelmere	Pinjarra Plain	Grey sand	Bushland, near road construction	(-31.9032140, 116.0074569)

2.3.2 Blinded case simulation (Chapter 6)

Five soil samples were provided by a third-party (identifying information can be found in Contribution of Others on page viii), along with a brief outline of the known information of these samples (Table 2.2). These soils were visually examined to determine their bulk colour, and then two representative replicates per sample were isolated for further analysis.

2.4 Preparation of the quartz-recovered fine fraction

Preparation of the quartz-recovered fine fraction was undertaken using the method developed by Pitts and Clarke (94) for the isolation and recovery of the <20 μm fraction coating found on quartz grains, as per Figure 1.11. Soil samples were passed through a 2 mm sieve to remove any larger organic material and man-made remnants. Roughly 100 – 250 mg of quartz grains were hand-picked from each sample under a microscope, weighed accurately (Table 2.3), and transferred to liquid scintillation vials. Deionised water was added to each vial to a height of approximately 1.5 cm above grain level, and samples were placed in an ultrasonic bath for 10 minutes. One by one, samples were agitated and then settled for 30 seconds before pouring the supernatant into new vials, to isolate suspended particles approximately <20 μm away from the larger grains. The liquids were centrifuged to concentrate the particulate fraction, and the supernatant was removed without disturbing the semi-solid layer below. The semi-solid material was re-suspended in the remaining liquid before being pipetted onto low background plates (single crystal silicon wafers, Philips Analytical) and allowed to air dry. The resulting films (with a typical areal concentration of approximately 1 mg cm^{-2}) were analysed in situ (X-ray diffraction) and then scraped off and homogenised before being stored in folded aluminium foil for further analysis (microspectrophotometry, Raman spectroscopy, infrared spectroscopy).

Table 2.3 Masses of quartz grains picked from each sample that were used for preparation of the quartz-recovered fine fractions (to five decimal places).

Location	Sample A mass (g)	Sample B mass (g)
0	0.02145	-
1	0.21977	0.18774
2	0.14883	0.09920
3	0.17774	0.20134
4	0.26395	0.24397
5	0.19749	0.20957
6	0.16030	0.14176
7	0.12463	0.10412
8	0.17076	0.20962
9	0.21250	0.27892
10	0.13873	0.09064
11	0.18929	0.16616
12	0.14844	0.14188
13	0.16547	0.13767
14	0.23140	0.14964
15	0.19847	0.16993
16	0.15628	0.19480
17	0.22524	0.17260
18	0.18528	0.21326
19	0.14521	0.14034
20	0.14773	0.09416
21	0.17075	0.12730
22	0.15939	0.21092
BS1 (Alibi Site 1)	0.09157	0.09274
BS2 (Suspect Recovered)	0.22998	0.21437
BS3 (Crime Scene)	0.23835	0.21038
BS4 (Alibi Site 2)	0.21887	0.25656
BS5 (Potential Site of Interest)	0.20846	0.28293

2.5 Instrumental methods

2.5.1 Microspectrophotometry (MSP)

The prepared quartz-recovered fine fractions were homogenised and mounted on new, pre-cleaned glass microscope slides (Superwhite, 1.1–1.2 x 76 x 26 mm, ProScitech) for analysis. Spectra were acquired from 310 – 800 nm using a CRAIC QDI 2000 microspectrophotometer, calibrated using NIST traceable standards, operated in reflectance mode with 150x magnification. An auto-set optimisation, dark scan, and reference scan were obtained prior to each sample analysis. Ten replicate scans were taken over different areas of each sample to account for intra-sample variation. Average spectra were produced from the replicate spectra for each sample, using CRAIC LambdaFire Microspectroscopy software (version 1.2.82.1). The average spectra were converted to L*a*b* colour co-ordinates using CRAIC LambdaFire Microspectroscopy software (version 1.2.82.1) and the following parameters: CIE 1931 2° Standard Observer over wavelength range 380 – 770 nm using D65 as the illuminant, corresponding to average daylight at a colour temperature of 6504 K. Data were exported as .csv files, or .txt files that were then converted to .xls files and collated within Microsoft Excel (Version 16.66).

2.5.2 Raman spectroscopy

The prepared quartz-recovered fine fractions were homogenised, mounted on glass slides (Superwhite glass microscope slides, 1.1–1.2 x 76 x 26 mm, ProScitech), and analysed using a WITec Confocal Raman-AFM (alpha300 RA) with a frequency-doubled Nd:YAG laser at 532 nm. Scans of each sample were undertaken from -200 – 3800 cm^{-1} , with 10 accumulations and 5 seconds exposure time. The laser power was optimised for each sample, and the exposure time/accumulations per sample were experimented with in order to give maximum signal output without overloading the detector or damaging the sample. Data acquisition and processing was performed using WITec Project FOUR software (version 4.1). Spectra were exported as .txt files, and then converted to .xls files and collated within Microsoft Excel (Version 16.66). In attempts to reduce fluorescence, baseline offset correction

and linear baseline correction were performed using The Unscrambler® X 10.5 (CAMO Software AS, Oslo, Norway).

2.5.3 Attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy

The prepared quartz-recovered fine fractions were homogenised and analysed with constant applied pressure on a Thermo Scientific Nicolet iS50 FTIR spectrometer with a single-bounce diamond ATR crystal, using sufficient sample to cover the crystal. Spectra were recorded in absorbance mode over a range of 4000 – 400 cm^{-1} , with 64 accumulated scans at a spectral resolution of 4 cm^{-1} . Three replicate scans were recorded for each sample to account for intra-sample variation. The crystal was cleaned between sample measurements using an ethanol-soaked tissue to remove contaminants and particulate matter. A background scan of the clean diamond crystal was acquired before each sample scan. Spectra were initially ATR-corrected for further chemometric analysis but ultimately resulted in unreliable data, as the refractive index of a natural heterogeneous material like soil cannot accurately be determined, so uncorrected spectra were utilised instead. Spectra were exported as .spa files.

2.5.4 X-ray diffraction (XRD)

2.5.4.1 ChemCentre

Initially, XRD analysis of samples was carried out under ambient laboratory conditions using a Philips Analytical PW1820 automatic powder diffractometer (APD) with Bragg-Brentano para-focusing geometry and Co $K\alpha$ radiation. Scanning, using a step size of 0.05° from 4 – 80° 2θ and a counting time of 12 seconds per step, was found to provide data of good quality from sample deposits. XRD patterns were zero-offset corrected using the 5.91 Å reflection of the low background plate. These were exported as .rd files using HighScore software (Malvern PANalytical, Version 4.9.0.27512), and then converted to .xls files using PowDLL Converter (Nikos Kourkoumelis, Version 2.97.0.0) and collated within Microsoft Excel (Version 16.66).

However, the instrument was upgraded halfway through analysis of the soil sample set, and the remaining samples were analysed on a Malvern PANalytical EMPYREAN III Diffractometer system equipped with a dCore-fitted Pixel3D detector, 240 mm radius theta-theta goniometer, a Reflection-transmission spinner 3.0 sample stage, and an iCore-fitted Co generator with K- α_1 1.78901 Å, K- α_2 1.79290 Å and K- β iron filter. The samples were collected at an operating current of 30 mA and tension of 40 kV at a range 4 – 80° 2 θ , 0.039° step size and approximately 12 seconds per step, with an incident beam mask of 14 mm and automatic divergence slits. These settings were chosen to best replicate the settings used during analysis with the previous Phillips XRD instrumentation. The data was K- α_2 stripped before interpretation. After data collection, the ‘new’ samples were data corrected to match the step size of the initial ‘old’ samples (0.05°) using HighScore software (Malvern PANalytical, Version 4.9.0.27512). XRD patterns were exported as .xrdml files, and then converted to .xls files using PowDLL Converter (Nikos Kourkoumelis, Version 2.97.0.0) and collated within Microsoft Excel (Version 16.66).

This entire dataset, including both old and new samples, was referred to as the “ChemCentre” XRD data throughout this thesis.

2.5.4.2 Commonwealth Scientific and Industrial Research Organisation (CSIRO)

Due to the mid-analysis upgrade in ChemCentre XRD instrumentation, samples were also analysed by CSIRO to provide a more consistent analysis source. This better allowed for variation to be attributed to different instrumentation use versus sample preparation of each batch. XRD analysis was carried out under ambient laboratory conditions using a Malvern PANalytical EMPYREAN II Diffractometer system equipped with a Pixel3D detector, 240 mm radius theta-theta goniometer, a reflection-transmission spinner sample stage, and Co generator with K- α_1 1.78901 Å, K- α_2 1.7929 Å and K- β iron filter. The samples were collected at an operating current of 40 mA and tension of 40 kV at a range 5 – 80° 2 θ , 0.053° step size and approximately 2 hour run, with an incident beam mask of 6.6 mm, automatic divergence slits, and no post-diffraction monochromator. XRD patterns were exported as .xrdml files, and then converted to .xls files using PowDLL Converter

(Nikos Kourkoumelis, Version 2.97.0.0) and collated within Microsoft Excel (Version 16.66). This dataset was referred to as the “CSIRO” XRD data throughout this thesis.

2.6 Data analysis

Chemometric analysis was conducted using The Unscrambler® X 10.5 (CAMO Software AS, Oslo, Norway). Data were imported directly from .xls, .csv, and .spa files. The regions associated with interference from the ATR diamond crystal (2350 – 1950 cm^{-1}) in ATR-FTIR trials and the low background plates (narrow regions centred on approximately 18.9°, 38.5°, and 59.3° 2θ) in XRD trials were excluded from chemometric analysis to prevent them from influencing the model (87-89, 98). Data were baseline offset corrected and unit vector normalised (except for L*a*b* values, which were left raw) to account for variations in compound concentrations and sample mass / thickness (80, 87, 89, 107). Other normalizations were explored (e.g., range normalization and area normalisation) but the performance of the model was not improved.

2.6.1 Principal component analysis

Data were mean-centred and subjected to principal component analysis (PCA) using the non-linear iterative partial least squares (NIPALS) algorithm. 2-dimensional and 3-dimensional scores plots were generated using the scores from the relevant principal components (PCs) and used to visualise the sample distribution and identify any outliers prior to further analysis. Additionally, examination of the loadings plots was conducted to identify the chemical factors that allowed for differentiation between samples.

2.6.2 Linear discriminant analysis

The discriminant models generated in Chapters 3 – 5 of this study were validated using test set validation. Each dataset was divided into two sets of data; a calibration set, and a validation set. The calibration set was used to generate the discriminant model, using a linear method on baseline corrected and normalised PCA scores, with up to seven components utilised depending on the dataset. The validation set was

then used to simulate unknown samples, which were predicted onto the model to assess its performance. For XRD, ChemCentre data was truncated from 5–80° 2 θ to match the CSIRO data, and the CSIRO data was corrected to match the step size of the ChemCentre XRD patterns (0.05°) using HighScore software (Malvern PANalytical, Version 4.9.0.27512). These were then combined to create one dataset for LDA.

The discriminant models generated in Chapter 6 of this study were validated using an independent dataset consisting of samples derived from an unknown soil. The four known soils were used as the calibration set to generate the discriminant model, using a linear method on baseline corrected and normalised PCA scores, with up to six components utilised depending on the dataset. The validation set, consisting of the *suspect recovered* soil, was then projected onto the model to predict its likely source.

2.7 Analysis sequence

In this study, analysis methods were determined by focusing on non-destructive instrumental techniques that can be used to provide information on primarily the inorganic content of soils. This involved conducting a visual examination of the soil samples, analysis of the colour instrumentally using MSP, and chemical analysis using ATR-FTIR spectroscopy and XRD. This is the order in which a forensic examiner would generally already carry out the forensic analysis of soils (13, 24). While XRD is presented as the final step in our methodology, the sample preparation was carried out with XRD as the first step due to practicality. XRD involves analysis of the whole sample as a thin film whereas MSP and ATR-FTIR require a smaller fraction of the sample as a powder. In order to minimise degradation and maximise recovery of already small samples, the quartz-recovered fine fraction suspensions were pipetted straight onto individual clean low background plates for the XRD. These were left to dry overnight, resulting in thin solid films. These films were then removed from the low background plates and stored in aluminium foil for further analyses.

2.8 Preliminary considerations

Due to the large number of samples used for analysis, data were collected over multiple different days, with every instrument appropriately calibrated before each session. It has been demonstrated that analytical precision is <0.1% for repeat spectrophotometry colour measurements made over a period of several minutes, compared to 4–5% for measurements made on the same soil samples over several days (111). In a casework scenario, it would therefore be optimal to analyse the known and questioned samples consecutively on the same day to minimise the effects of daily fluctuations in instrument performance. However, when numerous samples have been provided for analysis, this is not always possible. Therefore, it is important to note that the multiple calibrations of the instruments used may have resulted in variations to the spectra that could have impacted the results of the statistical analyses (107).

Whilst similar masses of quartz grains were collected from each sample during the preparation of the fine fraction, the sample masses varied between approximately 100 – 250 mg. Individual quartz grains in natural soil samples vary in size greatly, and so does the amount of fine fraction coating those grains (61). As a result, the final amount of fine fraction recovered from each sample was not consistent. Spectral data were baseline corrected and normalised to minimise any differences due to the thickness of sample being analysed (88, 107). To ensure representative sampling, two soil samples from each location were utilised for this study. To account for intra-sample variation, multiple replicate scans were acquired from each sample; ten scans were taken for MSP, three scans were taken for ATR-FTIR, and for XRD the entire surface of the sample film was scanned. Average MSP spectra were also generated from the combined replicate scans for each sample to minimise the influence of any outliers.

Chapter 3. Analysis of the Quartz-Recovered Fine Fraction of Western Australian Soils using Microspectrophotometry Paired with Chemometrics

Portions of this chapter have been published in the following article:

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics for the forensic analysis of Western Australian sandy soils." *Forensic Chemistry*, 2022. 28: 100412.

3.1 Introduction

In many areas of soil science analysis, soils are first characterised using general descriptors of structure, consistency, texture, and colour. Soil colour is an important tool for characterising soils, and can be extremely useful for differentiating between soil samples (75). Soils can range in colour from dark brown or grey, through to shades of red, yellow, and white. These differences in colour are the result of variations in the soil composition; quartz grains are generally coated in mixtures containing iron and aluminium oxides, organic matter, and clay, which can all affect the visual appearance of the soil (71). Iron oxides are generally responsible for the yellow, brown, and red coloured soils, with the individual shades dependant on temperature and humidity of the environment, as well as the age of the soils (53, 71). Soils in water environments generally appear lighter yellow or pale grey, as the iron oxide becomes hydrated and leaches into the water source (36, 53, 71). Darker brown or black shades of soil are generally associated with the presence of manganese or iron-manganese compounds, or can also indicate higher amounts of organic materials such as humus and decomposition products (71). Colour variations in soil samples can also be affected by sample treatment methods, such as drying, moistening, sieving, milling, or ashing (4, 15). Because of the significant variation in soil colour, colour measurements can be unreliable when used in isolation to characterise a soil sample. Several studies have noted completely different colour assessments for soils that originated very close together (1, 71), as well as very similar colour assessments for soils located kilometres apart (71). For this reason, colour determination of a sample should always be used alongside other methods of characterisation.

Munsell colour values are routinely used by soil scientists for classifying soil, where the user visually matches a sample with colour chips (4, 71, 75). The system encompasses a three-dimensional colour space that ascertains colours based on their hue (general colour), chroma (intensity of colour), and value (lightness of colour) to assign descriptive labels consisting of letters and numbers (71, 75). These terms are expressed by a number and typically one or two capital letters to represent

the hue, and then two more numbers to represent the chroma and value (e.g., 5YR 4/4, which is highlighted in Figure 3.1) (75). The process of assigning a sample a Munsell colour value is not only subjective, but can be affected by variables such as lighting conditions, moisture content of the soil, the amount of sample assessed, sample prep, the accuracy of the colour printing, fading of colour charts over time, and visual discrepancies between interpreters (4, 17, 71, 75). Any of these small variations can cause considerable deviations in the resulting assigned colour value. A study by Post *et al.* reported that soil scientists assigned the same Munsell colour value for a sample only 52% of the time (112), illustrating the degree of subjectivity surrounding the method. It is also challenging when working with forensic samples to compare them against reference colours, as forensic samples are often fractionated and present only in trace quantities (4, 71). Petraco *et al.* noted that soils originating within only 15 inches of each other were observed to have very different Munsell colour values, despite having the same mineral composition (1). While the Munsell method of colour classification is useful, it should not be the only process utilised for colour determination of soil samples, and more precise methods of colour measurement should be explored (1, 75).

The L*a*b* colour system (illustrated in Figure 3.2) is another method of measuring colour within three-dimensional space that relates to the human visual response (71). The Commission Internationale d'Éclairage (CIE), or International Commission on Illumination, introduced the CIE L*a*b* colour system in an attempt to standardise colour assessments relating to science, technology, and art (71, 113). This method produces three numerical values which are able to account for minimal variation between samples; L* is a measure of the lightness of a sample, a* encompasses the green-red components, and b* encompasses the blue-yellow components (71, 114). Whilst this method of classification is similar to Munsell, it has the potential to be more objective; a mathematical algorithm is used to convert a measured colour spectrum into chromaticity coordinates, bypassing the need for human judgement (71). It does however, require the use of a spectrophotometer to determine the colour of a sample, so forensic examination labs must be equipped with the appropriate instrumentation.

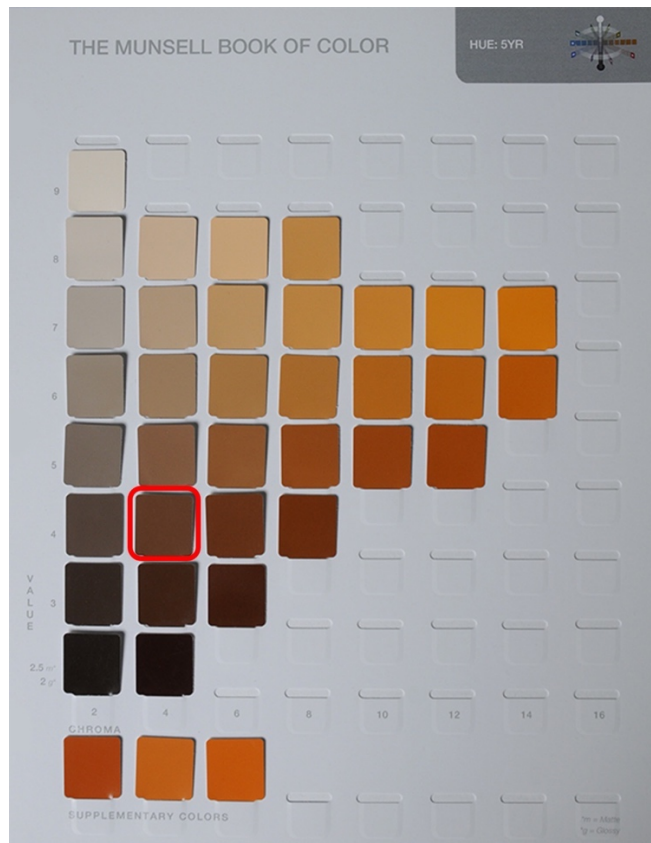


Figure 3.1 A page from a Munsell colour chart, depicting the differing chroma and value shades of 5YR (from “How to Read a Munsell Colour Chart” (115)). The example referenced in the text, 5YR 4/4, is highlighted.

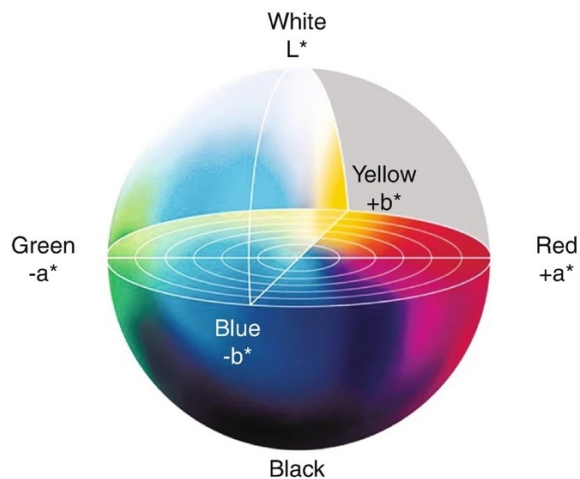


Figure 3.2 A 3D model of the CIELAB colour space, where L defines the lightness of a colour, a^* defines the green to red axis, and b^* the blue to yellow axis (from Yélamos et al. (113)).

Microspectrophotometry (MSP) is a spectroscopic technique that has been established as an examination method for forensic trace evidence, as it provides an

objective and more precise measurement of the colour of a small sample compared to visual colour comparisons (4, 71). It is regularly used for the forensic analysis of fibres, paint chips, and inks (107, 116), and more recently has been investigated for colour determination and compositional analysis of soil samples (4, 64, 76). A notable advantage of using MSP is the significantly smaller sample size requirements (4). MSP can be performed on microgram quantities of soil, making it suitable for use with forensic trace samples. However, this can become a disadvantage when working with highly variable samples, where individual measurements are not representative of the bulk soil colour. While this method is considered useful in certain cases, especially as part of a standard questioned versus known comparison, MSP analysis for forensic soil examination is still considered to be in the developmental stage, due to limited substantive validation and reliability testing. It has also never been demonstrated on the quartz-recovered fine fraction isolated from soils.

In this chapter, the use of MSP was explored for the forensic analysis of the quartz-recovered fine fraction of soils from the Swan Coastal Plain in Perth, Western Australia. Principal component analysis (PCA) was performed on the resulting individual replicate MSP spectra and their average sample spectra, to assess whether soils could be distinguished based on their location, or other attributes of the soil (such as visual appearance, location type, or dune system). MSP spectra were also converted into $L^*a^*b^*$ chromaticity values for further PCA. Linear discriminant analysis (LDA) was then used to classify and predict replicate samples using the 'independent test set' method, and the accuracy of each model was assessed.

3.2 Experimental

Soil samples were collected as outlined in Sections 2.2 and 2.3.1. The quartz-recovered fine fractions were prepared from each of the soil samples, as described in Section 2.4. MSP spectra were collected from the extracted quartz fine fractions as described in Section 2.5.1. PCA and LDA were then conducted on the data as outlined in Section 2.6.

3.3 Results and discussion

3.3.1 Chemometric analysis of soils using complete MSP spectra

3.3.1.1 Principal component analysis

MSP spectra were collected from the quartz-recovered fine fractions of 45 soil samples, outlined in Chapter 2: Table 2.1. Most of the spectra (Figure 3.3) appeared as relatively straight lines with different slope angles, encompassing reflectance from all wavelengths of light, but especially the yellow – red regions (>550 nm). This was mirrored by the visible colour of the soil samples, ranging from various shades of grey and brown, through to yellow and orange. Some samples, such as those from Location 7, displayed a distinctive step increase in the reflectance of yellow – orange light (550 – 625 nm), characteristic of soils orange in colour. Whilst these more intensely coloured soils could be separated from the population based on visual examination of their spectra alone, others were too alike to confidently distinguish. PCA was therefore employed to identify and enhance any differences between them.

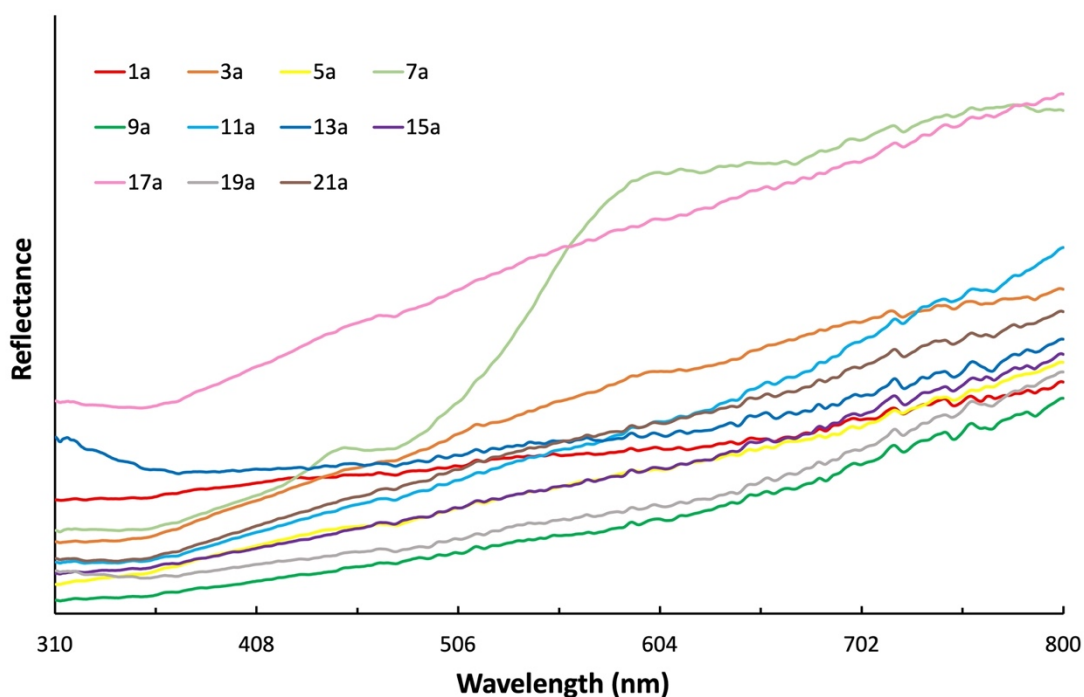


Figure 3.3 Raw MSP reflectance spectra showing the variability in composition of a selection of soil samples collected from differing locations within the Swan Coastal Plain.

PCA performed on these spectra revealed that 99.0% of the total variance in the dataset could be described by the first three principal components (PCs), as illustrated in the scree plot (Figure 3.4). Three-dimensional score plots generated using these PCs (Figure 3.5) revealed that most of the soil samples formed clusters based on the location from which they were collected, however, there was also a great deal of overlap throughout most of these clusters. As anticipated, the samples that achieved the greatest separation from the rest of the population were the red sands from Location 7, as they had more obvious differences in their MSP spectra due to their distinctive colour. Whilst commonly referred to as “red” sands throughout this study, this term was only adopted due to its regular use in Australia, and it must be noted that the appearance of Location 7’s soil could easily be described as orange (thus reinforcing the issues of subjective descriptions of colour). Location 7 was the only location that was able to be entirely individualised from the rest of the population within the model.

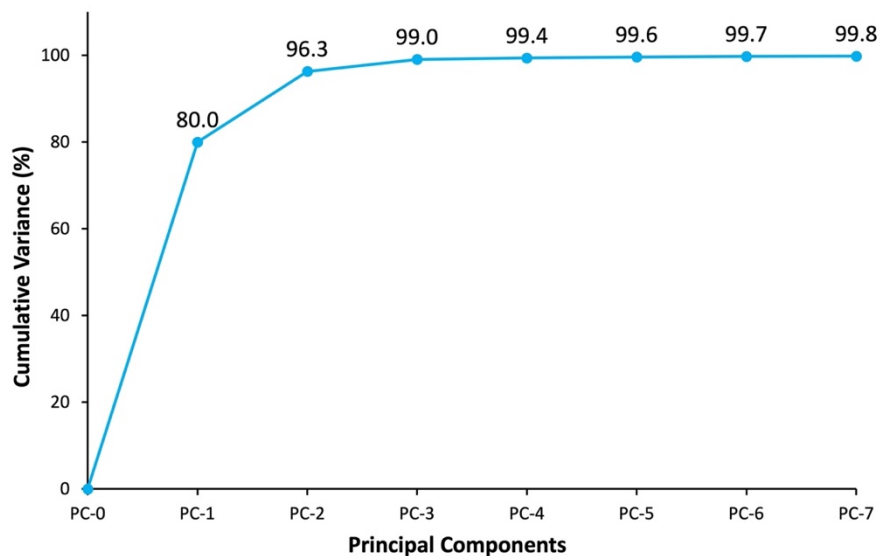


Figure 3.4 Scree plot depicting the cumulative variance in the MSP dataset retained by each PC.

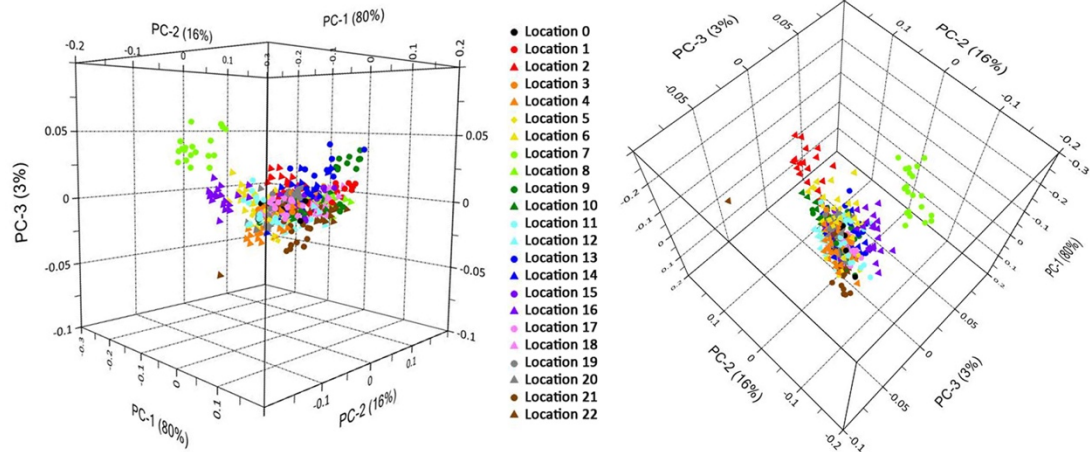


Figure 3.5 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from different locations based on their corresponding MSP spectra.

The factor loadings for the first three PCs (Figure 3.6) were studied to determine the regions in the MSP spectra that were associated with (and likely responsible for) the discrimination of samples along each component. The loadings plot for PC-1 showed a positive correlation with the reflectance of the purple/blue region of the visible light spectrum (<520 nm), and a negative correlation with the reflectance of orange/red light (>600 nm). Samples separated along this PC may therefore contain different proportions of yellow/orange/red components present in their soil. For example, all the red sands (Location 7) along with some yellow and brown sands (Locations 2, 6, 5 and 16) were positioned along PC-1 in a negative direction. This was due to their higher proportion of yellow, orange, or red components combined with the negative association of PC-1 with this spectral region; as well as the direct association of the negative region of PC-1 with orange/red light, the positive region of PC-1 was associated with purple/blue light, which indicates that the negative region was also associated with yellow/orange light due to the lack of purple/blue light being reflected (yellow/orange is opposite to purple/blue on the colour wheel).

The loadings plot for PC-2 revealed a negative correlation with the reflectance of yellow/orange light (550 – 650 nm). The red soils from Location 7 were best separated from the rest of the population negatively along PC-2, with the yellow and yellow/brown coloured sands from Locations 16 and 11 following closely behind. PC-3 displayed a negative correlation with the reflectance of blue/green light (450 – 550

nm), and a small positive association with the reflectance of yellow/orange light (575 – 650 nm). Soils separated along this PC were therefore again being skewed based on the proportion of yellow/orange/red components they contain. The soils from Locations 21, 4, 5, and 12 (all coloured variations of brown and yellow) were best separated in a negative direction along PC-3, while the red soils from Location 7 were best separated from the rest of the population positively along PC-3. The spread of the samples along PC-3 was skewed towards the positive side, with most samples sitting along the neutral baseline, indicating that the general population of soils all contained varying degrees of yellow, orange, and red components.

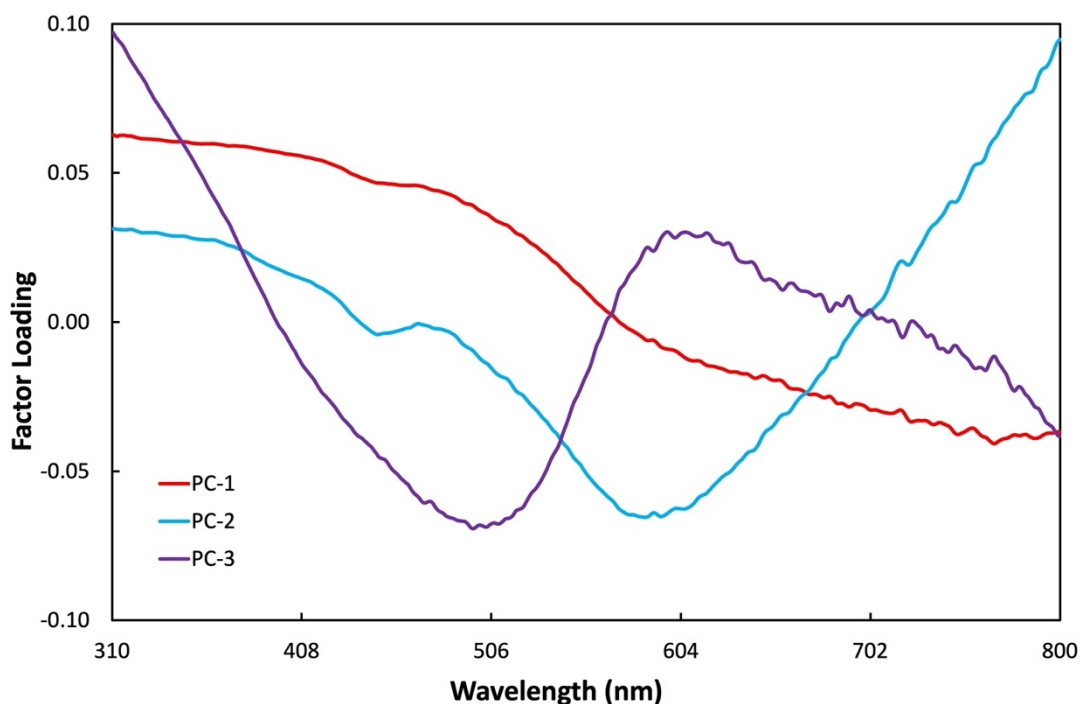


Figure 3.6 Factor loadings plot of PCs 1-3 for PCA of the soil MSP reflectance dataset.

Interestingly, not far behind the Location 7 soils on PC-3 were soils from Locations 9 and 13, which were both coloured grey. This may be due to variation within the non-visible region of the spectra, which has been picked up on some of the loadings. For example, PC-3 showed influence from variation attributed to very low wavelengths of light, or the ultraviolet (UV) region. The glass slides used to mount samples for MSP absorb light in the UV region (117), indicating that some samples may not have been sufficiently thick to block this absorbance. The PCA model also displayed one replicate from sample 22b as an outlier to the rest of the dataset, located at the most

extreme negative points of PC's 1 and 3. Using averaged MSP spectra to build the PCA model, rather than the entire replicate dataset, may assist in smoothing out the effects of the outlier and overcoming this anomaly, without completely removing it. In this instance, using the full MSP data for chemometric analysis highlighted the large degree of intra-variability within soil samples.

Colour coding of the soil samples in the PCA model based on visually assigned groupings (outlined in Chapter 2: Table 2.1) rather than sampling location, displayed the clustering based on their overall colour, as expected (Figure 3.7). The red sands were clearly discriminated from the rest of the population, however, there was a great deal of overlap between the variations of yellow and brown sands, and the brown and grey coloured sands or clays. The brown sands were widely dispersed throughout the plot, overlapping with all clusters except the red sands. While the MSP spectra were very closely correlated with the visual appearance of the soils, they did account for more of the micro-heterogeneity in colour compared with visual examination. This may become a disadvantage when using complex statistical methods on the data, as reflectance spectra can be noisy and not reproducible (118). This may lead to greater uncertainty and more difficulty when attempting to differentiate between samples, and ultimately can result in the overlapping of samples or significant separation within a sample class. This issue has previously been encountered for soil samples of larger particle sizes (4). Unfortunately, even after isolation of the quartz-recovered fine fraction (<20 μm), this micro-heterogeneity still posed difficulties with MSP, indicating that the evident solution may be to use a macro-spectrophotometric approach instead.

There was no additional structure uncovered within the dataset when visualising the PCA model grouped according to the dune system that the samples originated from nor the type of location that the soils were collected from (Figures 3.8 & 3.9), of which details are outlined in Chapter 2: Table 2.1. This indicates that the MSP spectra were most closely correlated with the colour and appearance of the soil, not the underlying mineralogy or environmental components of their surroundings.

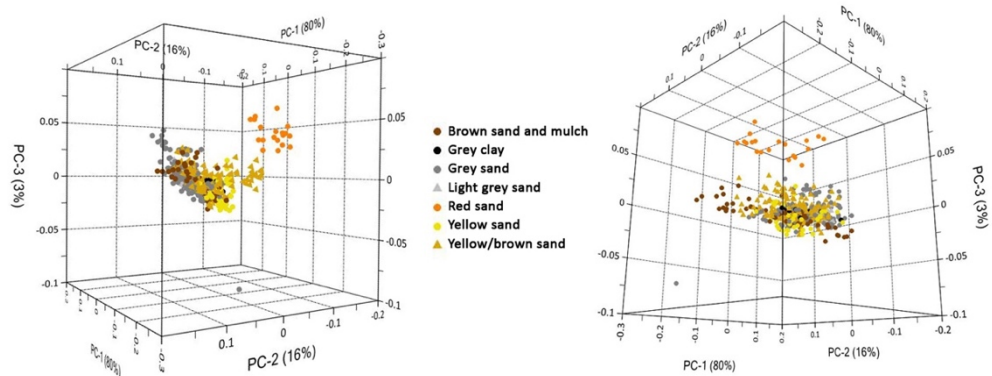


Figure 3.7 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding MSP spectra.

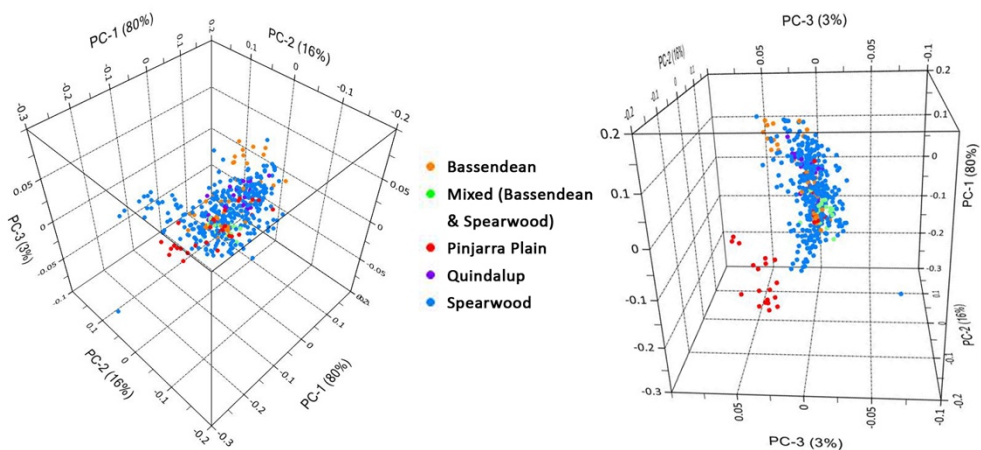


Figure 3.8 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding MSP spectra.

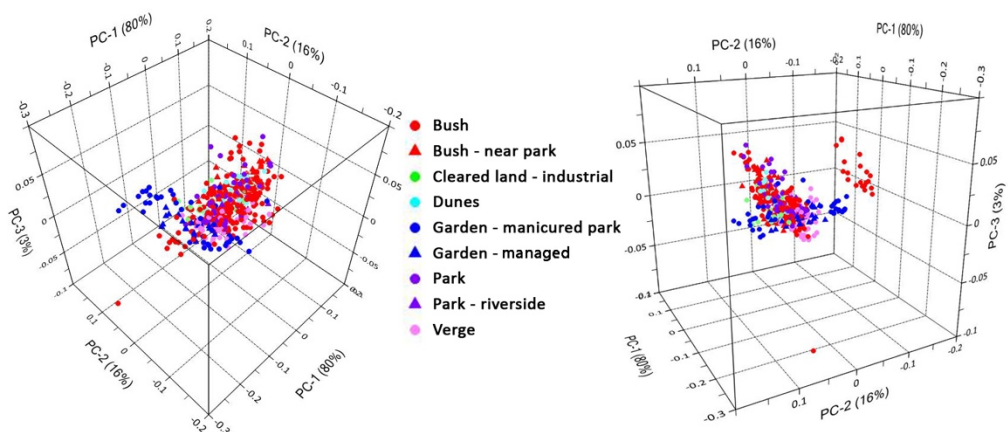


Figure 3.9 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding MSP spectra.

3.3.1.2 Linear discriminant analysis

The MSP data was then split into two datasets for LDA; nine out of the 10 replicates per sample were used to build the calibration model, and the remaining replicates were used as a validation set and projected onto the model to assess its performance. The validation set consisted of replicates that were not utilised within the calibration set, to ensure no overlap between models. Using the same replicates for the calibration and validation sets would lead to overestimation of the model's performance, and unrealistic results (105). Although maximum visual separation was achieved in the PCA model by utilising only the first three PCs, using additional PC's for LDA (up to seven) was found to improve discrimination between locations and increase the validation accuracy.

LDA was performed on the MSP dataset, with each location treated as an individual class. This produced a single discriminant function that was used to predict the source of replicates from the calibration dataset (nine out of the 10 replicates per sample). This discriminant model returned a calibration accuracy of 61.7% (Table 3.1). The two soils that were classified 100% correctly were those originating from Location 7, which was to be expected as they were the only location that achieved full discrimination within the PCA scores plot, and Location 17. There is some uncertainty surrounding the classification of Location 17 soils as there was almost no discrimination of these samples from other groups within the scores plot; it may be that their central position amongst the rest of the population density allowed them to be classified correctly by chance. This was supported by the discriminant values, which indicate the distance measured between a sample and the centroid of each class. A sample will be predicted as originating from the class with the best fit, i.e., the class that exhibits the lowest discriminant values (105). The discriminant values for Location 17's replicates (Table 3.2) showed that they were not well separated from several other location classes, so classifications to this location should not be made with confidence. The large degree of overlap between the rest of the locations observed within the scores plot was responsible for the misidentification of the majority of the remaining soil samples.

Table 3.1 Number of correct vs incorrect location classifications for samples in the MSP calibration set using a 7-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	% Correct
0	7	2	78
1	9	9	50
2	14	4	78
3	11	7	61
4	15	3	83
5	7	11	39
6	10	8	56
7	18	0	100
8	14	4	78
9	9	9	50
10	10	8	56
11	8	10	44
12	13	5	72
13	10	8	56
14	0	18	0
15	8	10	44
16	16	2	89
17	18	0	100
18	6	12	33
19	13	5	72
20	9	9	50
21	11	7	61
22	14	4	78
% Total Correct			62

Table 3.2 Discriminant values of Location 17's replicates from the MSP calibration dataset (rounded to three decimal places), with correct classifications shaded green. The last column demonstrates how far away the next closest classification was, as a percentage of the lowest discriminant value obtained.

Sample Replicate	Discriminant Values																						Classified Location	Closest Value Within (%)	
	Location 0	Location 1	Location 2	Location 3	Location 4	Location 5	Location 6	Location 7	Location 8	Location 9	Location 10	Location 11	Location 12	Location 13	Location 14	Location 15	Location 16	Location 17	Location 18	Location 19	Location 20	Location 21			Location 22
17a (1)	-13.915	-13.825	-20.323	-7.905	-10.514	-7.194	-12.471	-189.085	-19.974	-17.997	-8.422	-11.102	-10.962	-7.097	-5.573	-9.352	-37.696	-4.975	-10.280	-11.145	-7.157	-8.244	-15.597	17	12
17a (2)	-14.467	-20.242	-22.415	-8.273	-7.989	-7.369	-6.910	-177.429	-24.165	-21.918	-10.536	-7.737	-12.912	-8.864	-6.254	-13.291	-29.141	-4.547	-15.283	-8.850	-5.955	-8.194	-20.291	17	31
17a (3)	-10.750	-16.962	-19.288	-5.687	-8.268	-5.954	-7.354	-172.751	-21.935	-19.722	-8.695	-6.958	-11.637	-8.027	-5.611	-10.125	-28.705	-4.135	-13.506	-8.045	-4.765	-8.739	-19.565	17	15
17a (4)	-10.274	-13.646	-25.021	-7.440	-8.752	-7.806	-10.482	-151.297	-18.355	-16.892	-10.788	-6.184	-12.743	-9.261	-7.018	-9.682	-22.880	-4.693	-11.638	-9.341	-6.819	-8.693	-16.335	17	32
17a (5)	-12.059	-17.282	-22.226	-6.860	-9.087	-7.210	-7.919	-169.023	-22.034	-19.435	-9.908	-6.886	-12.980	-7.015	-5.439	-10.988	-27.374	-3.666	-13.692	-8.261	-5.881	-8.951	-19.470	17	48
17a (6)	-13.807	-17.868	-23.325	-8.591	-7.571	-6.236	-10.603	-178.924	-23.926	-23.221	-10.616	-8.488	-12.700	-9.844	-7.262	-12.070	-32.032	-4.207	-13.675	-12.320	-5.851	-8.473	-19.223	17	39
17a (7)	-13.122	-16.130	-22.163	-8.781	-6.065	-5.343	-9.714	-170.110	-21.079	-21.320	-9.354	-7.094	-12.080	-8.708	-6.362	-11.390	-29.554	-3.817	-12.930	-9.798	-5.567	-8.049	-18.312	17	40
17a (8)	-9.081	-15.769	-21.772	-7.436	-9.638	-6.837	-12.141	-151.016	-23.032	-22.693	-11.810	-7.257	-15.088	-12.697	-10.245	-10.618	-25.575	-6.482	-15.108	-13.216	-6.629	-13.094	-22.128	17	2
17a (9)	-14.462	-19.036	-23.771	-8.157	-8.966	-8.215	-7.886	-178.783	-22.513	-19.830	-10.376	-8.258	-12.394	-8.067	-5.653	-12.502	-29.723	-4.601	-14.065	-8.200	-6.566	-7.669	-18.687	17	23
17b (1)	-15.998	-24.273	-28.303	-11.000	-8.007	-8.952	-6.868	-156.988	-28.698	-26.879	-15.054	-5.797	-18.049	-10.580	-8.785	-17.255	-21.767	-4.197	-19.591	-11.541	-7.314	-10.960	-25.032	17	38
17b (2)	-11.581	-15.755	-22.031	-7.635	-5.850	-5.114	-8.816	-165.304	-20.424	-20.381	-9.062	-6.016	-11.710	-8.523	-5.997	-10.639	-27.131	-3.481	-12.777	-8.659	-4.934	-7.852	-18.265	17	42
17b (3)	-11.555	-20.308	-22.661	-8.683	-8.348	-7.167	-7.172	-144.655	-25.770	-24.590	-12.754	-4.759	-17.454	-9.976	-8.708	-13.959	-20.391	-4.631	-18.526	-10.096	-6.407	-13.185	-25.607	17	3
17b (4)	-17.834	-21.527	-32.173	-13.442	-8.586	-10.886	-9.214	-150.845	-23.880	-22.703	-15.240	-6.544	-17.778	-9.571	-8.250	-17.143	-21.139	-4.957	-17.476	-9.831	-9.711	-9.929	-21.287	17	32
17b (5)	-14.650	-20.914	-24.007	-10.608	-7.595	-7.951	-6.187	-149.021	-24.602	-23.358	-12.645	-5.146	-17.061	-8.616	-7.555	-15.386	-20.986	-4.395	-18.074	-8.501	-7.290	-11.353	-23.742	17	17
17b (6)	-16.197	-17.028	-27.411	-10.946	-9.803	-9.738	-9.798	-162.810	-20.314	-18.119	-11.806	-7.740	-14.859	-6.023	-5.579	-13.209	-26.405	-4.039	-13.421	-8.395	-8.940	-8.728	-17.646	17	38
17b (7)	-22.514	-23.055	-28.467	-14.117	-12.463	-11.667	-11.287	-191.308	-27.345	-24.286	-13.950	-12.190	-17.560	-6.324	-6.693	-17.591	-36.650	-5.087	-17.056	-12.337	-10.952	-10.584	-21.875	17	24
17b (8)	-19.148	-18.799	-25.393	-12.925	-9.138	-8.473	-12.417	-183.044	-24.315	-23.726	-12.087	-10.911	-15.023	-8.647	-7.607	-14.899	-35.542	-5.092	-14.374	-13.554	-9.030	-9.420	-19.115	17	49
17b (9)	-17.547	-18.711	-25.279	-12.110	-8.186	-8.221	-9.493	-168.091	-22.458	-21.685	-11.455	-8.015	-15.365	-6.608	-6.127	-14.624	-29.118	-4.031	-15.061	-9.284	-8.402	-9.489	-20.011	17	52

This LDA model was then used to predict the locations of 45 samples from the validation dataset (remaining one replicate per sample). Only 10 out of 45 samples, or 22.2%, had their location correctly predicted (Table 3.3). Soils from Locations 2, 7, and 22 were all classed 100% correctly. Whilst Location 7 also achieved 100% classification accuracy within the calibration model, Locations 2 and 22 still both achieved relatively high classification rates (78%) due to being situated at the outskirts of the population density in the scores plot. The other locations that achieved correct predictions for one of their validation samples (Locations 6, 9, 16, and 21) were also samples that were situated on the outskirts of the scores plot displaying minimal overlap with others, and as such, all achieved $\geq 50\%$ classification accuracy in the calibration model. It is therefore evident that samples do not necessarily need to be entirely discriminated within the PCA model to be predicted correctly by LDA, though it is desirable.

While Location 17 achieved 100% classification accuracy in the calibration model, both of its validation samples were predicted incorrectly, as suspected. Sample 17b showed some variation between its calibration and validation spectra (Figure 3.10), particularly in the shorter wavelengths, however, sample 17a did not display the same differences, indicating that the issue was most likely due to the uncertainty within the calibration model. Interestingly, most of the misclassified samples were predicted as originating from Locations 2, 4, 6, 16, or 22, presumably due to their high classification accuracy and hence the model's confidence in discriminating those groups more so than others. All the locations that were centrally situated in the scores plot had both samples predicted incorrectly, indicating that the LDA model was unable to discriminate between many of the locations that overlapped in areas of high density. Using averaged MSP spectra for future chemometric analysis, rather than the entire replicate dataset, may help to simplify the dataset and allow for more accurate predictions.

Table 3.3 Number of correct vs incorrect location predictions for samples in the MSP validation set using a 7-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Predicted	% Correct
0	0	1	4	0
1	0	2	2 & 22	0
2	2	0	-	100
3	0	2	6	0
4	0	2	6 & 16	0
5	0	2	6 & 16	0
6	1	1	2	50
7	2	0	-	100
8	0	2	22	0
9	1	1	2	50
10	0	2	2 & 6	0
11	0	2	6	0
12	0	2	21 & 22	0
13	0	2	6	0
14	0	2	6 & 22	0
15	0	2	2	0
16	1	1	7	50
17	0	2	6	0
18	0	2	2 & 22	0
19	0	2	6 & 16	0
20	0	2	6	0
21	1	1	4	50
22	2	0	-	100
% Total Correct				
22				

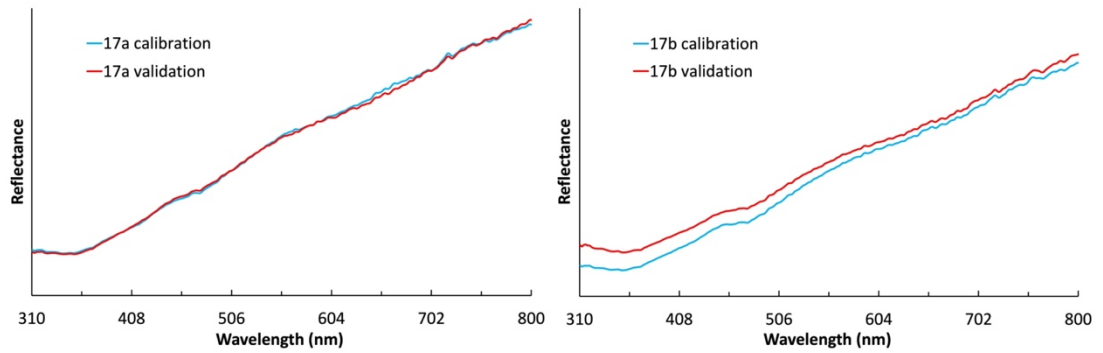


Figure 3.10 Raw calibration and validation MSP spectra obtained from Location 17's soils (most similar calibration sample pictured).

3.3.2 Chemometric analysis of soils using average MSP spectra

3.3.2.1 Principal component analysis

To minimise the heterogeneity observed within samples and obtain spectra that were more representative of the bulk soil colour by negating the influence of outliers, the MSP spectra previously collected from the quartz-recovered fine fractions of 45 soil samples were averaged to obtain one average spectrum per sample. PCA performed on these average spectra revealed that 97.7% of the total variance in the dataset could be described by the first three PCs (Figure 3.11).

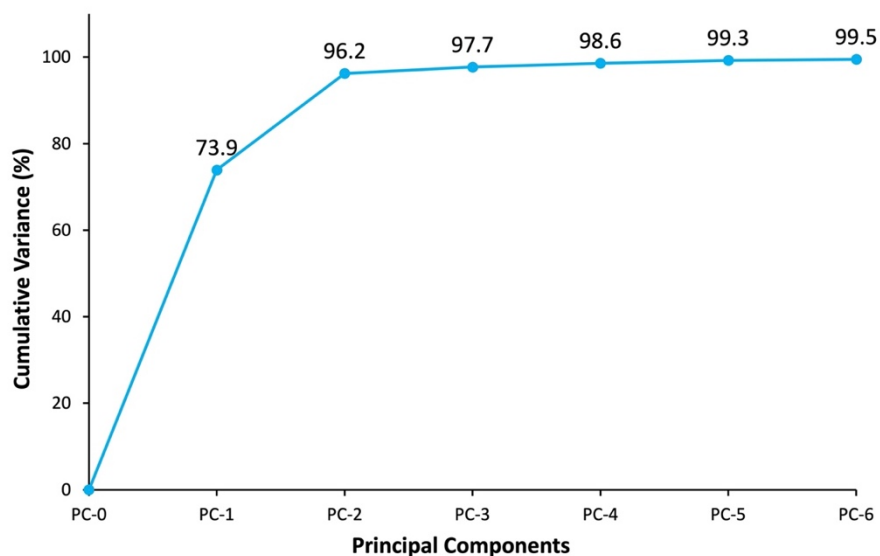


Figure 3.11 Scree plot depicting the cumulative variance in the MSP average spectra dataset retained by each PC.

Three-dimensional score plots generated using these PCs (Figure 3.12) showed improved separation in contrast to the previous model which utilised all replicate spectra; samples from Locations 2, 7, 16, and 22 all achieved clustering and full separation from the rest of the population with no overlap, whilst most remaining samples achieved clustering with minimal overlap, but in limited proximity to the rest of the population. Samples from Locations 9, 13, 14, and 21 all displayed significant separation between their two corresponding samples, with one sample for each location grouped centrally with the rest of the population, and the other sample distanced considerably at the extremes of the plot. This indicated higher levels of variability in the soils from these locations.

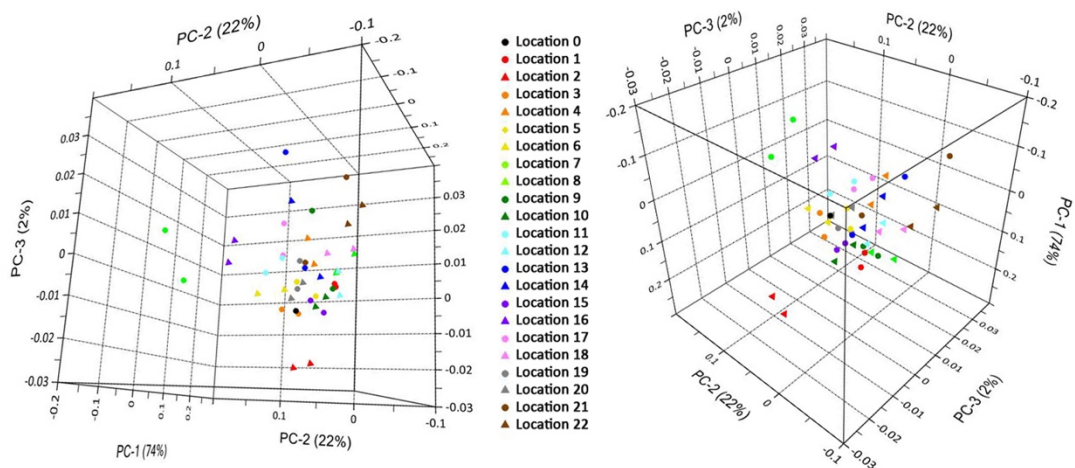


Figure 3.12 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from different locations based on their corresponding average MSP spectra.

The factor loadings for the first three PCs (Figure 3.13) were studied to determine the regions in the MSP spectra that were associated with the discrimination of samples along each component. The loadings plot for PC-1 displayed a global minimum associated with reflectance of the orange region of the visible light spectrum (550 – 650 nm). As such, the samples located at the most negative points along PC-1 were all red, yellow, and yellow/brown sands, except for the outlier from Location 21 (brown sand and mulch). The loadings plot for PC-2 revealed a negative correlation with the reflectance of blue/green light (450 – 525 nm), and positive correlation with the reflectance of orange light (575 – 625 nm). The red soils from Location 7 were best separated from the rest of the population along PC-2, due to

their characteristic orange colour. The samples located in the most negative region of PC-2 tended to be the grey coloured sands due to their lack of orange components, except for the outlier from Location 21.

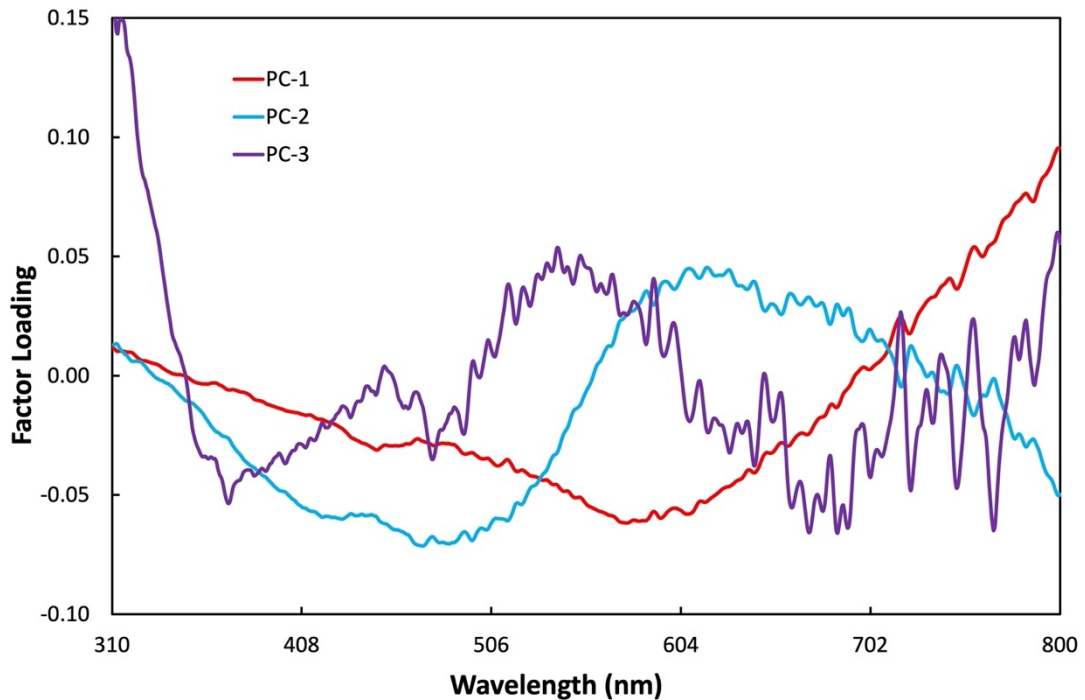


Figure 3.13 Factor loadings plot of PCs 1-3 for PCA of the soil average MSP reflectance dataset.

The loadings plot for PC-3 exhibited absorbance within the UV region, again indicating sample variation due to thickness of the soil sample. It also displayed a large degree of noise, so it was questionable as to whether PC-3 (responsible for just 1.5% of the total variance) represented any systematic variation between samples, or simply noise in the dataset. The samples that were best differentiated in a positive direction along PC-3 were from the locations that experienced significant separation between their two samples: Locations 9, 13, 14, and 21. When the scores plot was replotted utilising only PC-1 and PC-2 (Figure 3.14), comparable separation was achieved to that from the 3D model, with the exception of Location 22; this location overlapped with Location 18 on the 2D model, but was entirely separated from the rest of the population on the 3D model. This indicated that while PC-3 did not significantly contribute to the individualisation of the locations, it did provide additional separation when utilised alongside PCs 1 and 2. The inclusion of too many PCs when visualising the dataset may result in undesirable modelling of variation

attributed to noise within the data, whereas using less than the required number of PCs can cause valuable information to be overlooked (105).

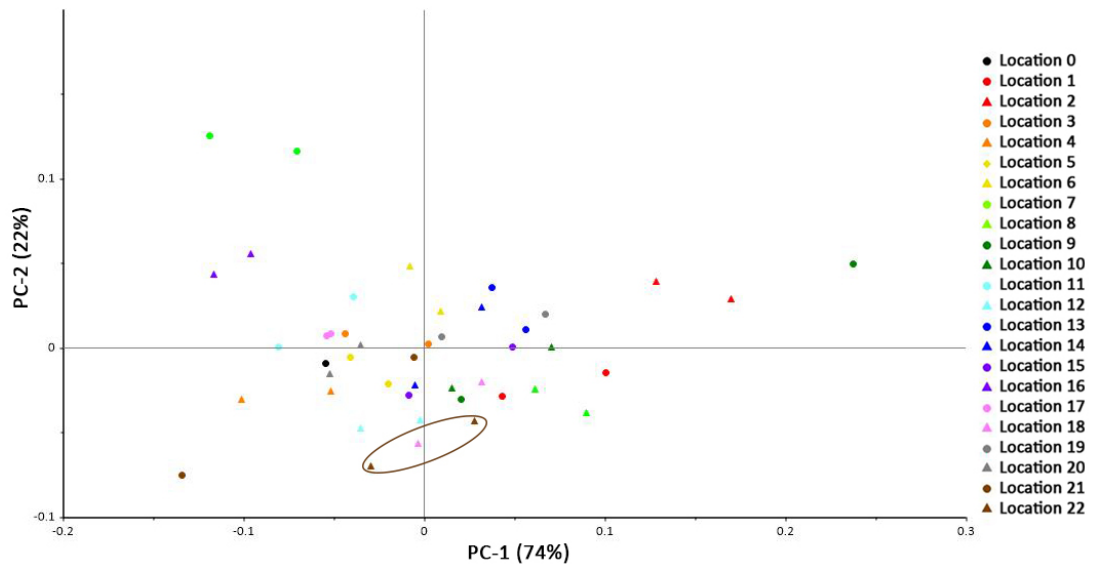


Figure 3.14 2-dimensional PCA scores plot showing the variability of soil samples from different locations based on their corresponding average MSP spectra. Circle indicates Location 22 which, upon removal of PC-3, now overlaps with Location 18.

Colour coding of the soil samples in the PCA model based on visually assigned groupings (outlined in Chapter 2: Table 2.1) better demonstrated the loose clustering based on their overall colour, and the large degree of overlap (Figure 3.15). This showed the red sands discriminated from the rest of the population, however, most of the other coloured groups were clumped together in the same region of the plot. There was no additional structure uncovered within the dataset when visualising the PCA model grouped according to the dune system that the samples originated from, nor the type of location that the soils were collected from (Figures 3.16 & 3.17). While the visible colour of the soils was the most discriminating factor within the MSP spectra, it was not the only feature contributing to the separation of these samples, as evident in Figure 3.15.

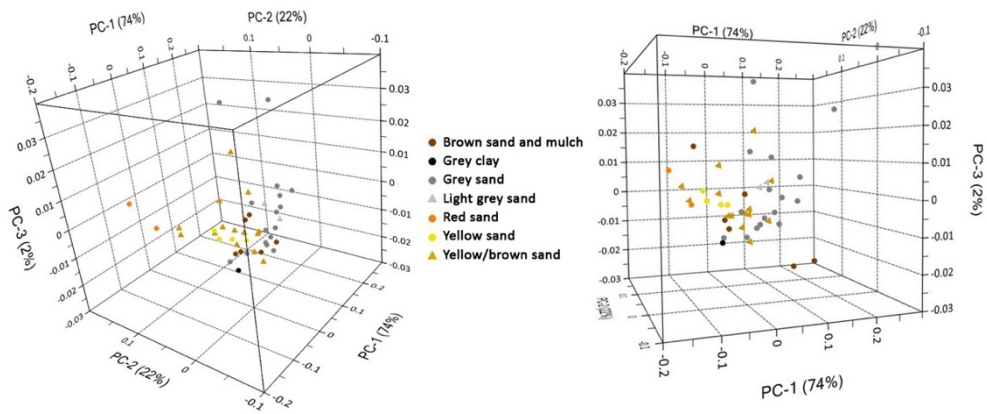


Figure 3.15 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding average MSP spectra.

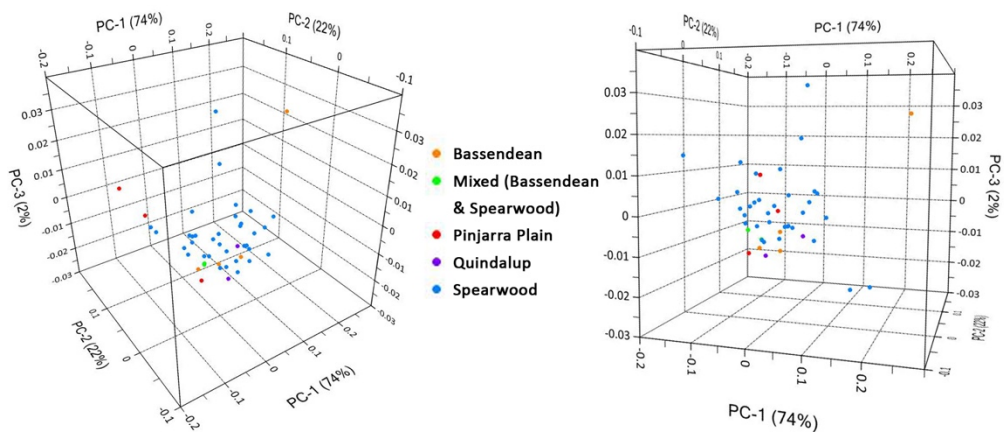


Figure 3.16 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding average MSP spectra.

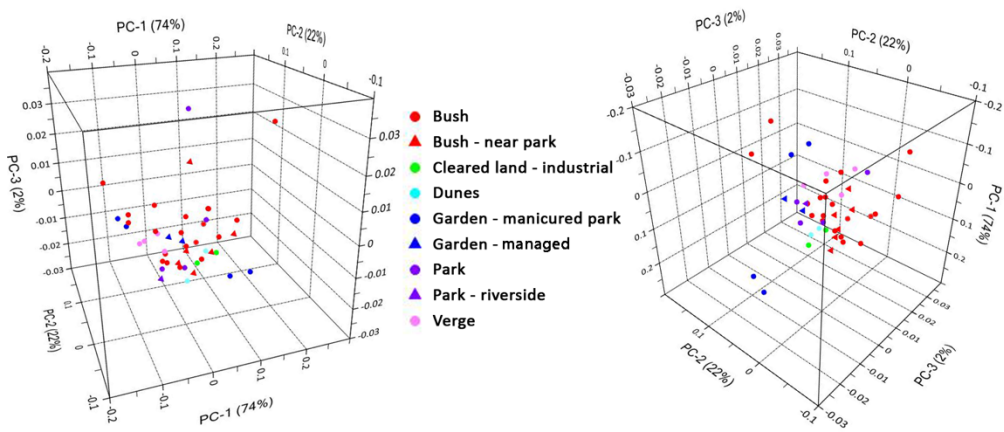


Figure 3.17 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding average MSP spectra.

3.3.2.2 Linear discriminant analysis

Due to only having one average MSP spectrum per sample, or two average spectra for each location, this data was unable to be split into two datasets for LDA; if one spectrum per location was utilised for the validation set, there would only be one spectrum per location left to build the model, which is not enough to display trends in the data and generate precise discriminate functions for the model. Instead, two independent datasets were utilised; the average MSP spectra were used as the calibration set to build the LDA model (excluding Location 0 as it only had one average spectra), and the same validation set from the previous LDA model (one replicate spectrum per sample) was predicted onto the model to assess its performance. The limited number of calibration samples however, did affect the total number of PCs that could be utilised for LDA, allowing a maximum of only two PCs to be used to build the model.

LDA was performed on the average MSP dataset with each location treated as an individual class, and the discriminant model returned a calibration accuracy of 54.6% (Table 3.4). This was slightly less than that obtained by the previous model (61.7%) which utilised all replicate spectra, rather than average spectra, however, the calibration accuracy of the previous model was proved to be overestimated in comparison to the validation accuracy. Of the six locations that were classified completely correctly, three of them (Locations 7, 16, and 22) had achieved full separation within the scores plot. Even though Location 2 was also entirely discriminated within the scores plot, only one of its samples was classified correctly, while the other was incorrectly classified as originating from Location 9. This is due to the large variation seen within Location 9 samples. The other three locations that were correctly classified (Locations 10, 15, and 17) were clustered appropriately in the scores plot, however, exhibited a fair amount of overlap with several other location clusters. This was supported by their discriminant values that showed that they were not well separated from other location classes within the LDA model (Table 3.5). Locations 9 and 14 both displayed significant intra-location separation within the scores plot, and as such, had their samples entirely misclassified within the LDA model. Location 18 exhibited minimal separation between its samples within

the scores plot but overlapped with several other clusters, while Location 3 samples were well-clustered together but in close proximity to several other samples; these locations also had both of their samples misclassified.

Table 3.4 Number of correct vs incorrect location classifications for samples in the average MSP calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	% Correct
1	1	1	50
2	1	1	50
3	0	2	0
4	1	1	50
5	1	1	50
6	1	1	50
7	2	0	100
8	1	1	50
9	0	2	0
10	2	0	100
11	1	1	50
12	1	1	50
13	1	1	50
14	0	2	0
15	2	0	100
16	2	0	100
17	2	0	100
18	0	2	0
19	1	1	50
20	1	1	50
21	1	1	50
22	2	0	100
% Total Correct			55

Table 3.5 Discriminant values of Location 10, 15, and 17's samples from the average MSP calibration dataset (rounded to three decimal places), with correct classifications shaded green. The last column demonstrates how far away the next closest classification was, as a percentage of the lowest discriminant value obtained.

Sample	Discriminant Values																						Classified Location	Closest Value Within (%)
	Location 1	Location 2	Location 3	Location 4	Location 5	Location 6	Location 7	Location 8	Location 9	Location 10	Location 11	Location 12	Location 13	Location 14	Location 15	Location 16	Location 17	Location 18	Location 19	Location 20	Location 21	Location 22		
10a	-4.699	-7.405	-8.365	-7.272	-5.618	-14.390	-99.502	-6.292	-6.192	-3.275	-16.595	-3.593	-6.852	-4.902	-3.307	-44.172	-12.920	-3.647	-5.533	-7.923	-5.345	-5.184	10	1
10b	-4.392	-4.580	-9.102	-9.712	-6.927	-14.139	-99.067	-6.152	-4.225	-3.275	-17.807	-5.384	-6.033	-5.150	-3.741	-45.471	-14.176	-4.722	-5.072	-9.333	-7.925	-6.915	10	14
15a	-5.991	-8.822	-7.034	-5.698	-4.584	-12.722	-93.586	-7.927	-7.834	-3.722	-14.192	-3.659	-6.440	-4.390	-3.320	-39.920	-10.864	-4.204	-5.208	-6.450	-4.215	-5.921	15	10
15b	-5.790	-5.430	-6.631	-7.359	-5.100	-10.839	-88.952	-8.145	-5.687	-3.417	-13.771	-5.450	-4.565	-3.839	-3.320	-38.517	-10.704	-5.358	-3.849	-6.837	-6.321	-8.030	15	3
17a	-21.770	-18.545	-3.852	-5.287	-5.504	-3.942	-48.835	-27.039	-21.851	-13.312	-3.403	-15.112	-6.546	-6.671	-10.504	-13.941	-3.092	-17.741	-7.054	-3.988	-7.909	-22.799	17	10
17b	-21.921	-18.518	-3.872	-5.414	-5.590	-3.861	-48.400	-27.230	-21.906	-13.419	-3.373	-15.322	-6.522	-6.719	-10.608	-13.768	-3.092	-17.937	-7.065	-4.047	-8.082	-23.058	17	9

This model was then used to predict the locations of 44 samples from an independent dataset (one replicate MSP spectra per sample). 18 out of 44 samples, or 40.9%, had their location correctly predicted (Table 3.6). This was a significant improvement in comparison to the previous LDA model that utilised individual replicate spectra (22.2%). Most of the locations that had all their samples correctly predicted were those that achieved 100% classification accuracy in the calibration model (Locations 7, 16, and 22). Discriminant values for Locations 7 and 16 showed that these groups were very well separated from all other classes (Table 3.7). However, Locations 4, 12, and 22, which also had both their validation samples correctly predicted, displayed very similar discriminant values to several other locations, so classifications to these groups should be treated with caution. As expected, Location 10, 15, and 17's validation samples were all predicted incorrectly despite achieving 100% classification accuracy in the calibration model, due to small differences in the magnitude of their calibration discriminant values for several location classes. Overall, using MSP data for chemometric analysis, even when using average spectra, still highlighted the innate intra-variability within soil samples. This prevented visually similar soils that originated from different locations from being fully characterised and discriminated.

Table 3.6 Number of correct vs incorrect location predictions for samples in the average MSP validation set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Predicted	% Correct
1	1	1	18	50
2	1	1	9	50
3	1	1	14	50
4	2	0	-	100
5	0	2	11 & 14	0
6	1	1	13	50
7	2	0	-	100
8	1	1	18	50
9	0	2	1 & 2	0
10	0	2	1 & 21	0
11	0	2	6 & 20	0
12	2	0	-	100
13	0	2	5 & 11	0
14	0	2	13 & 15	0
15	1	1	9	50
16	2	0	-	100
17	0	2	3 & 6	0
18	0	2	10 & 22	0
19	0	2	13 & 17	0
20	1	1	3	50
21	1	1	12	50
22	2	0	-	100
% Total Correct				41

Table 3.7 Discriminant values of Location 4, 7, 12, 16, and 22's samples from the average MSP validation dataset (rounded to three decimal places), with correct predictions shaded green. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

Sample	Discriminant Values																					Predicted Location	Closest Value Within (%)	
	Location 1	Location 2	Location 3	Location 4	Location 5	Location 6	Location 7	Location 8	Location 9	Location 10	Location 11	Location 12	Location 13	Location 14	Location 15	Location 16	Location 17	Location 18	Location 19	Location 20	Location 21			Location 22
4a	-12.186	-14.181	-4.511	-3.181	-3.402	-8.511	-74.314	-15.415	-14.420	-7.095	-8.134	-6.594	-6.368	-4.508	-5.381	-27.389	-6.049	-8.509	-5.626	-3.772	-3.336	-11.188	4	5
4b	-27.270	-30.235	-8.534	-5.890	-8.557	-11.316	-59.550	-31.890	-31.252	-18.622	-7.630	-16.218	-14.481	-12.015	-15.029	-19.206	-6.974	-20.297	-14.037	-7.089	-7.861	-23.601	4	18
7a	-198.482	-170.495	-115.326	-128.849	-132.188	-94.369	-12.515	-215.702	-190.082	-168.683	-88.780	-179.501	-127.437	-136.680	-158.009	-46.019	-99.135	-188.473	-134.562	-118.920	-146.097	-206.180	7	268
7b	-118.158	-96.504	-57.600	-68.397	-69.822	-42.913	-3.260	-131.514	-111.312	-95.725	-40.138	-104.910	-65.367	-72.385	-88.012	-15.250	-46.963	-111.211	-70.557	-60.558	-80.867	-125.197	7	368
12a	-7.218	-14.510	-11.833	-7.814	-7.681	-20.086	-110.246	-8.117	-11.438	-6.154	-20.360	-3.534	-12.221	-8.379	-5.855	-49.821	-16.219	-4.387	-10.132	-10.284	-5.131	-4.497	12	24
12b	-5.142	-8.788	-8.520	-6.798	-5.546	-14.929	-100.276	-6.659	-7.244	-3.636	-16.614	-3.297	-7.555	-5.208	-3.523	-44.231	-12.909	-3.601	-6.097	-7.831	-4.801	-4.886	12	7
16a	-66.471	-55.678	-23.354	-28.159	-30.322	-16.222	-13.670	-76.111	-64.506	-49.793	-12.785	-53.751	-30.648	-33.464	-43.736	-3.259	-16.447	-59.251	-33.394	-24.327	-36.174	-68.608	16	292
16b	-52.294	-40.987	-16.149	-21.718	-22.376	-9.830	-18.428	-61.045	-49.288	-37.859	-8.432	-42.952	-20.913	-24.069	-32.976	-3.772	-11.186	-47.108	-23.455	-17.542	-28.576	-56.042	16	124
22a	-4.582	-11.761	-15.755	-12.214	-10.803	-24.941	-125.870	-4.600	-8.052	-5.633	-26.957	-3.834	-13.942	-10.267	-6.399	-61.093	-21.935	-3.501	-11.602	-14.509	-8.395	-3.105	22	13
22b	-4.473	-11.937	-17.502	-13.890	-12.235	-27.096	-131.312	-4.202	-7.925	-6.126	-29.455	-4.350	-15.219	-11.458	-7.160	-64.997	-24.160	-3.748	-12.757	-16.262	-9.710	-3.115	22	20

3.3.3 Chemometric analysis of soils using MSP generated L*a*b* values

3.3.3.1 Principal component analysis

In an attempt to improve objective discrimination between soils from different locations, the 1931 L*a*b* colour identification system was applied to the MSP spectral data and the resulting chromaticity values subjected to chemometric analysis. The MSP average spectra were converted into L*a*b* chromaticity values, where L* is a measure of the lightness, a* is a measure of the green-red components, and b* is a measure of the blue-yellow components (Table 3.8). PCA performed on these L*a*b* values (without baseline correction or normalisation) revealed that 100% of the total variance in the dataset could be described by the first three PCs (Figure 3.18).

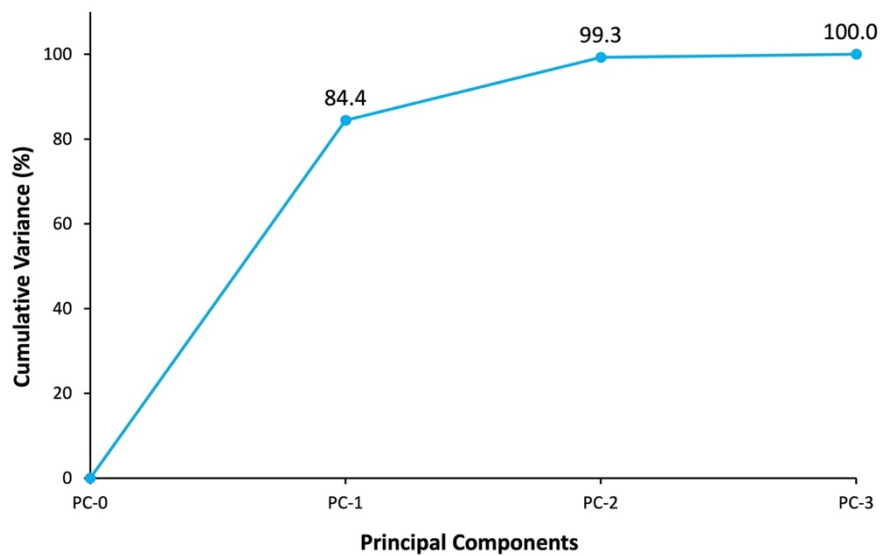


Figure 3.18 Scree plot depicting the cumulative variance in the L*a*b* dataset retained by each PC.

Three-dimensional score plots generated using these PCs (Figure 3.19) showed that most locations achieved clustering, however, there was overlap from nearby clusters. Some samples that achieved full discrimination from the rest of the population, for example Locations 2, 7, and 16, also exhibited a large degree of separation between their two samples, indicating significant differences in their L*a*b* values despite originating from the same location. The soils from Locations 2 and 16 were collected from gardens situated within manicured public parks, so were both easily accessible

and subject to a high degree of human interference. It is therefore logical to draw the conclusion that human interference may be contributing to the variation shown across a number of these sites. Location 7 soils were less subject to human activity as they were collected from dense bushland, indicating that insufficient sampling or natural variation within this soil may be the cause of its differentiation between samples.

Table 3.8 The soil samples analysed with MSP and their associated L*a*b* values generated from their average MSP spectra (rounded to four decimal places).

Soil sample	L* value	a* value	b* value	Soil sample	L* value	a* value	b* value
0	64.2791	2.8065	10.8863	12a	56.7341	0.8510	6.4930
1a	55.9059	1.1399	4.2602	12b	57.3398	1.4207	8.7911
1b	56.2500	1.3886	5.8750	13a	67.6367	1.7562	7.9304
2a	68.7063	6.2472	13.3930	13b	59.8091	2.0960	8.7805
2b	56.8339	4.5389	11.1808	14a	56.2367	1.4892	9.1630
3a	62.1019	2.5179	11.7694	14b	62.3351	1.7956	7.9813
3b	56.3380	2.5417	9.5831	15a	63.9001	1.8910	8.9114
4a	59.0213	2.0413	12.1935	15b	62.4857	1.7032	6.9949
4b	60.0170	1.4232	10.5839	16a	64.9556	3.3651	17.8349
5a	55.1989	2.2699	10.1966	16b	73.6729	4.8884	22.0461
5b	62.4657	2.1919	11.2158	17a	73.5557	1.9415	11.7408
6a	67.0394	4.1823	17.8365	17b	67.0131	1.8730	11.2222
6b	67.4678	3.4417	15.3133	18a	52.5049	0.4836	5.3095
7a	70.0773	7.1455	24.3205	18b	61.4045	1.0513	5.5890
7b	69.3384	9.4518	31.4252	19a	53.3802	2.0136	8.6523
8a	52.0707	1.0962	5.6762	19b	53.6458	2.4171	10.9857
8b	45.7070	0.8797	4.1592	20a	63.7990	2.2876	12.3336
9a	52.8953	1.3304	7.0526	20b	65.4104	2.6937	11.9284
9b	51.8791	1.4722	4.1714	21a	56.4992	1.3755	9.1620
10a	63.9778	1.6931	7.6101	21b	58.2570	-0.0501	7.0350
10b	59.1573	2.3984	9.2895	22a	51.0447	0.3322	4.8069
11a	65.6523	3.2408	15.1505	22b	54.8201	0.0600	4.4877
11b	66.9372	2.2640	12.8585				

Whilst the $L^*a^*b^*$ values should give similar results to those obtained with the average MSP spectra, only the visible region of the MSP spectra was utilised when calculating their associated $L^*a^*b^*$ values. The difference in separation therefore may be a result of eliminating noise or variation detected within the UV or infrared (IR) regions of the MSP spectra due to sample thickness. If this is the case, tighter clustering according to the visual appearance of the soil would be expected.

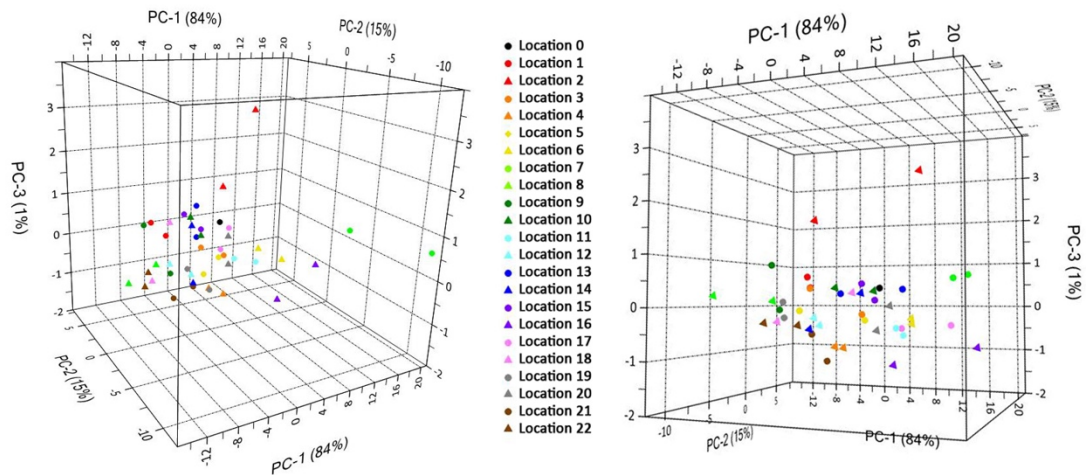


Figure 3.19 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from different locations based on their corresponding $L^*a^*b^*$ values, derived from their average MSP spectra.

To determine which values were associated with the discrimination of samples along each component, the factor loadings for the first three PCs were studied (Figure 3.20). The loadings plot for PC-1 showed strong positive correlation with both L^* and b^* values. Location 7 soils were the most positive along PC-1, primarily due to their high b^* values, which were associated with higher amounts of yellow within the sample. Location 16 soils were also positioned nearby, however, one sample was significantly more positive along PC-1 than the other, due to the variation seen in their L^* and b^* values. Samples from Locations 8, 9, and 22 (all grey coloured) were positioned at the most negative point on PC-1 due to their low L^* and b^* values, indicating that their soils were darker in colour and less yellow compared to the rest of the population. PC-2 revealed a positive correlation with L^* values, and strong negative correlation with b^* values. Location 7 soils were the most separated along PC-2 in a negative direction, due to their significantly higher b^* values resulting from

their distinct colour. PC-3 exhibited a significant positive correlation with a^* values. Soils from Location 2 (brown) were the most influenced by PC-3, followed by Location 7 (red), due to elevated a^* values from higher proportions of red within their samples.

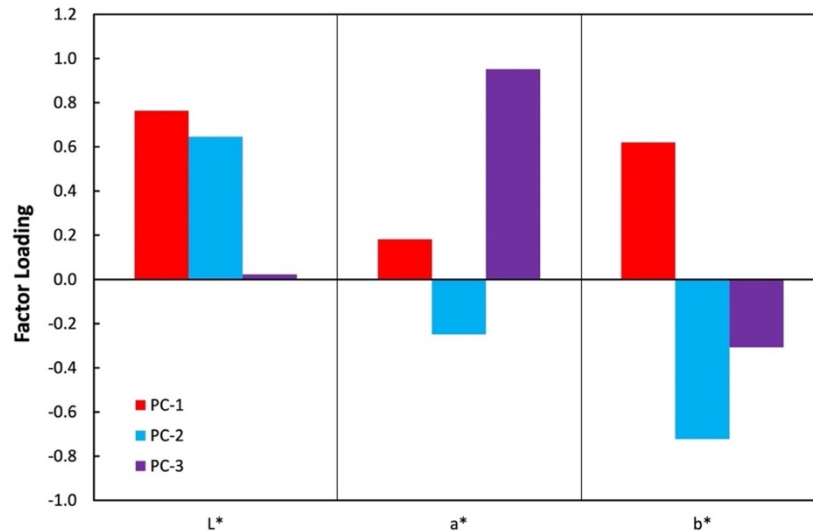


Figure 3.20 Factor loadings plot of PCs 1-3 for PCA of the soil $L^*a^*b^*$ values, generated from average MSP reflectance spectra.

Colour coding of the PCA model based on visual appearance (outlined in Chapter 2: Table 2.1) illustrated the loose clustering based on overall colour, but still revealed the large degree of overlap that was seen in previous models (Figure 3.21). The red soils were again the only group entirely discriminated from the rest of the population. As expected, the positive region of PC-1 was mostly occupied by red, yellow, and yellow/brown soils due to their higher b^* values, however, as the L^* values also contributed to the variation across PC-1, some of these yellow based soils were positioned in the negative, representative of their darker shading. There was no additional structure uncovered in the model when visualising the soils grouped according to the dune system they originated from, nor the type of location they were collected from (Figures 3.22 & 3.23).

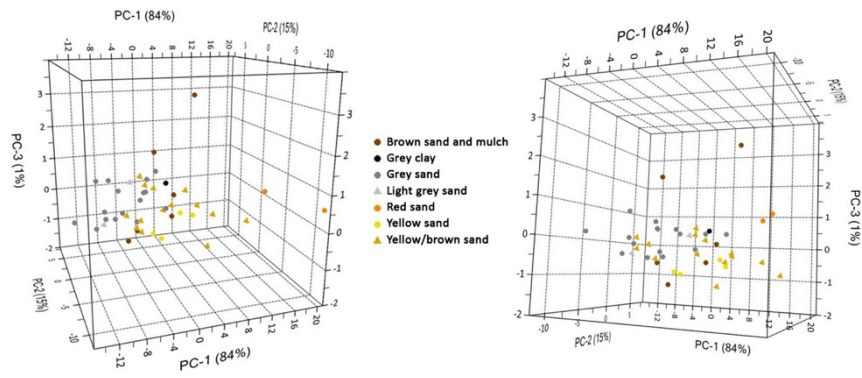


Figure 3.21 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding $L^*a^*b^*$ values determined from their average MSP spectra.

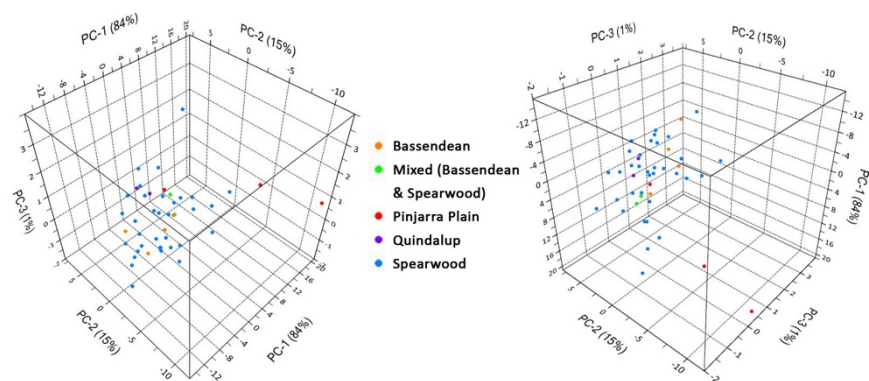


Figure 3.22 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding $L^*a^*b^*$ values determined from their average MSP spectra.

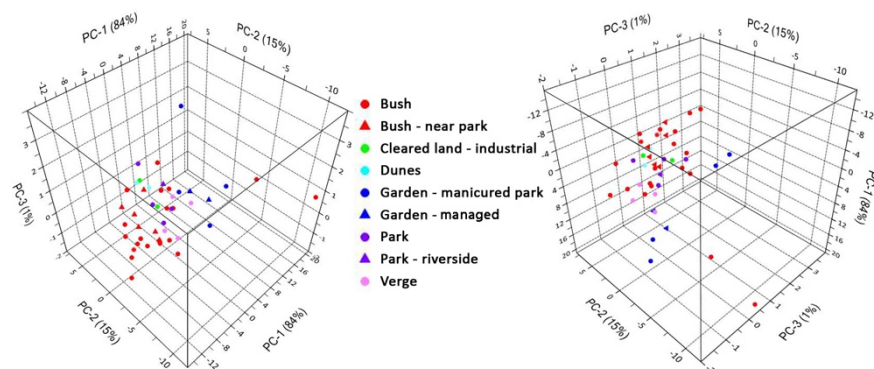


Figure 3.23 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding $L^*a^*b^*$ values determined from their average MSP spectra.

3.3.3.2 Linear discriminant analysis

As previously outlined in Section 3.3.2.2, only one average MSP spectrum per sample was generated, and hence only one set of $L^*a^*b^*$ values per sample. For this reason, the $L^*a^*b^*$ values obtained from the average MSP spectra were used as the calibration set to build an LDA model (excluding Location 0), and the 10th replicate spectra (one per sample) from the full MSP dataset were converted into $L^*a^*b^*$ values (Table 3.9) and predicted onto the model to assess its performance. It is important to note that due to the limited number of calibration samples, a maximum of only two PCs could be utilised to build the model.

LDA was performed on the average $L^*a^*b^*$ dataset, with each location treated as an individual class. This discriminant model returned a calibration accuracy of 47.7% (Table 3.10), which was lower than both the previous MSP models. All five locations that were classified 100% correctly (Locations 6, 7, 17, 19, and 20) were clustered based on their groupings within the scores plot and achieved separation from nearby locations. The locations that had just one of their samples classified correctly showed appropriate clustering with minimal overlap in the 3-PC scores plot, however, as the LDA model was only able to utilise two PCs, the separation attributed to PC-3 was lost. This caused a lot of these locations to subsequently overlap with each other, as demonstrated in Figure 3.24. Locations 3, 5, 10, and 14 all displayed separation between their individual samples within the scores plot and a high degree of overlap with other clusters, and hence had their samples entirely misclassified within the LDA model. Although Locations 2 and 12 did achieve some discrimination within the 3D scores plot, the removal of PC-3 and its attributed variation caused them to overlap with many other groups in the 2D scores plot, causing all their samples to be misclassified within the LDA model. This demonstrates how important it is to utilise an appropriate number of PCs when building the LDA calibration model, and in turn, ensure enough samples have been collected for analysis. Unfortunately, this is not always possible when conducting analysis of forensic trace evidence, as the number of collected samples may be limited.

Table 3.9 The soil sample replicates utilised for the LDA validation set and their associated L*a*b* values generated from their MSP spectra (rounded to four decimal places).

Soil sample (10 th replicate)	L* value	a* value	b* value	Soil sample (10 th replicate)	L* value	a* value	b* value
1a	59.1183	1.3574	4.3157	12a	57.8061	0.9632	6.5447
1b	59.5282	0.9172	5.9220	12b	57.2317	1.7137	9.4209
2a	69.7178	5.7035	11.5173	13a	69.5672	1.2959	6.7269
2b	52.1190	5.4198	12.3133	13b	58.0038	1.3991	7.0395
3a	57.2933	2.6324	10.6697	14a	56.4230	1.4673	9.0910
3b	51.5144	2.4748	9.6603	14b	54.1626	1.8653	8.7516
4a	58.4932	2.0310	12.1207	15a	62.3154	2.6227	10.6758
4b	63.7901	0.2095	8.9203	15b	59.0960	1.8009	6.7086
5a	56.1023	3.5198	14.1750	16a	68.6653	3.9915	21.4552
5b	64.6478	1.3416	11.3023	16b	73.8167	6.2242	23.5438
6a	65.1538	3.6523	14.9838	17a	70.8600	1.8541	12.9943
6b	64.2896	4.1962	13.8927	17b	68.0675	2.2127	11.6651
7a	66.3696	7.6883	20.7809	18a	52.2895	0.3844	4.7400
7b	71.8725	9.9805	34.2156	18b	62.1228	1.2048	5.5036
8a	54.9322	1.4039	6.1755	19a	53.6100	2.1700	9.7116
8b	44.4803	1.2914	5.3426	19b	57.9618	2.2315	11.3623
9a	53.3229	2.2343	9.0512	20a	60.7616	2.7242	12.2753
9b	56.0878	1.7210	4.4668	20b	67.3697	3.0449	12.0730
10a	68.3154	1.4674	9.2595	21a	57.6761	0.2649	7.0031
10b	65.0541	1.7991	7.7829	21b	53.2545	-0.1012	5.7257
11a	69.1661	3.3941	16.0846	22a	53.5115	0.3596	4.9407
11b	65.3382	2.2883	13.2274				

Table 3.10 Number of correct vs incorrect location classifications for samples in the L*a*b* calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Classified	% Correct
1	1	1	18	50
2	0	2	4 & 11	0
3	0	2	5 & 20	0
4	1	1	3	50
5	0	2	19 & 20	0
6	2	0	-	100
7	2	0	-	100
8	1	1	9	50
9	1	1	22	50
10	0	2	14 & 15	0
11	1	1	20	50
12	0	2	18 & 21	0
13	1	1	14	50
14	0	2	15 & 21	0
15	1	1	13	50
16	1	1	6	50
17	2	0	-	100
18	1	1	9	50
19	2	0	-	100
20	2	0	-	100
21	1	1	12	50
22	1	1	1	50
% Total Correct				48

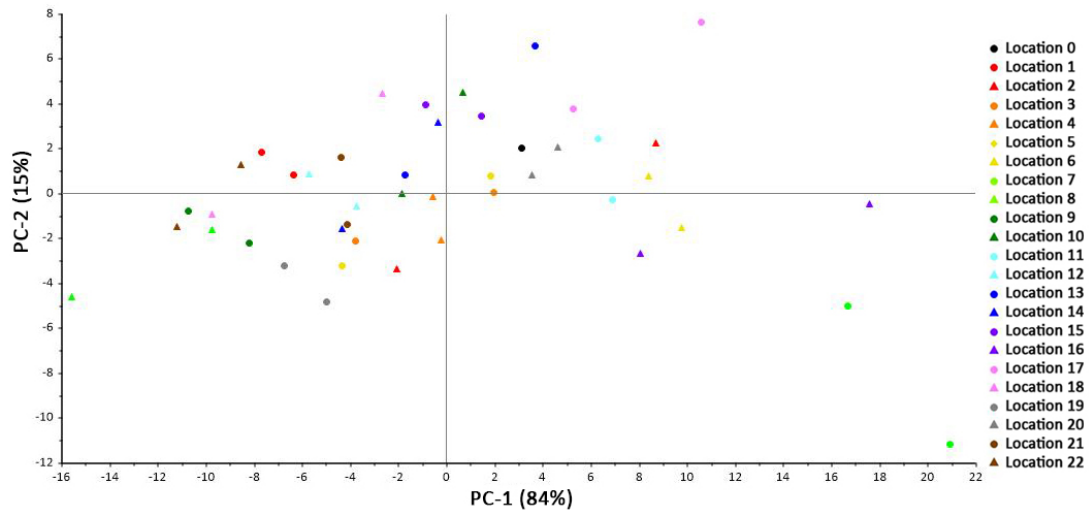


Figure 3.24 2-dimensional PCA scores plot showing the variability of soil samples from different locations based on their corresponding $L^*a^*b^*$ values, derived from their average MSP spectra.

This model was then used to predict the locations of 44 independent samples based on their MSP generated $L^*a^*b^*$ values (one replicate per sample). 12 out of 44 samples, or 27.3%, had their location correctly predicted (Table 3.11). This was a worse result than achieved by the average MSP model (40.9%), but minimal improvement compared to the original full MSP model (22.2%). Out of the three locations that attained 100% correct predictions, only Location 17 achieved 100% classification accuracy within the calibration model. All these locations however, were not very well separated based on their discriminant values (Table 3.12), and as such there was not a considerable amount of confidence surrounding these predictions.

Additionally, all other locations with 100% calibration accuracy had $\leq 50\%$ of their samples predicted correctly in the validation model, indicating possible discrepancies between the averaged and the unaveraged $L^*a^*b^*$ values. Comparing the calibration and validation values for Location 6 illustrated the minor differences across all three chromaticity coordinates (Figure 3.25). Whilst the $L^*a^*b^*$ values were generated from the average MSP spectra, the average MSP LDA model significantly outperformed this $L^*a^*b^*$ model. Both models were only able to utilise the first two PCs, however, this only appeared to negatively affect the $L^*a^*b^*$ model. PC-3 from the average MSP PCA represented a large amount of noise, and by excluding it when

constructing the LDA model, this improved discrimination based on variation in soil characteristics. However, when PC-3 from the L*a*b* PCA was excluded from its LDA model, this was at a disadvantage as it encompassed variation detected in the a* values, which was omitted. Therefore, the average MSP LDA model performed better and achieved a higher validation accuracy, even though improved separation within the scores plot was seen in the L*a*b* PCA model.

Table 3.11 Number of correct vs incorrect location predictions for samples in the L*a*b* validation set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Predicted	% Correct
1	0	2	18	0
2	0	2	17 & 19	0
3	0	2	5 & 19	0
4	1	1	13	50
5	0	2	2 & 20	0
6	0	2	11	0
7	1	1	16	50
8	1	1	9	50
9	0	2	1 & 19	0
10	0	2	13 & 15	0
11	1	1	6	50
12	0	2	5 & 18	0
13	0	2	12 & 15	0
14	0	2	19 & 21	0
15	0	2	3 & 12	0
16	2	0	-	100
17	2	0	-	100
18	0	2	15 & 22	0
19	1	1	4	50
20	1	1	2	50
21	0	2	9 & 12	0
22	2	0	-	100
% Total Correct				27

Table 3.12 Discriminant values of Location 16, 17, and 22's samples from the L*a*b* validation dataset (rounded to three decimal places), with correct predictions shaded green. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

Sample	Discriminant Values																				Predicted Location	Closest Value Within (%)		
	Location 1	Location 2	Location 3	Location 4	Location 5	Location 6	Location 7	Location 8	Location 9	Location 10	Location 11	Location 12	Location 13	Location 14	Location 15	Location 16	Location 17	Location 18	Location 19	Location 20			Location 21	Location 22
16a	-45.367	-13.799	-21.353	-19.926	-21.673	-6.533	-11.924	-50.662	-44.437	-29.046	-11.780	-34.140	-29.674	-29.346	-31.369	-3.464	-20.713	-44.087	-27.115	-16.544	-32.839	-50.296	16	89
16b	-61.729	-22.755	-32.940	-31.128	-33.411	-11.759	-8.245	-69.225	-61.470	-41.500	-19.154	-48.591	-41.598	-42.452	-43.730	-5.865	-28.751	-60.052	-40.945	-25.664	-46.992	-68.041	16	41
17a	-16.763	-5.812	-8.032	-7.715	-8.359	-6.816	-46.765	-25.068	-19.377	-7.832	-4.370	-12.145	-6.980	-9.273	-7.700	-12.418	-3.404	-15.646	-14.026	-4.506	-11.513	-21.141	17	28
17b	-12.577	-4.934	-5.915	-5.768	-6.161	-7.670	-51.283	-19.606	-14.717	-5.503	-4.320	-8.771	-4.982	-6.558	-5.499	-14.509	-3.316	-11.667	-10.699	-3.669	-8.276	-16.231	17	11
22a	-3.325	-15.203	-9.119	-9.984	-8.937	-27.909	-94.356	-3.950	-3.336	-6.781	-19.009	-4.472	-7.849	-5.921	-7.198	-42.143	-16.784	-3.521	-7.562	-13.743	-4.768	-3.114	22	7
22b	-3.333	-16.563	-10.173	-11.111	-10.001	-29.764	-98.692	-4.358	-3.697	-7.197	-20.322	-4.942	-8.103	-6.477	-7.393	-44.584	-17.271	-3.519	-8.775	-14.758	-5.279	-3.159	22	6

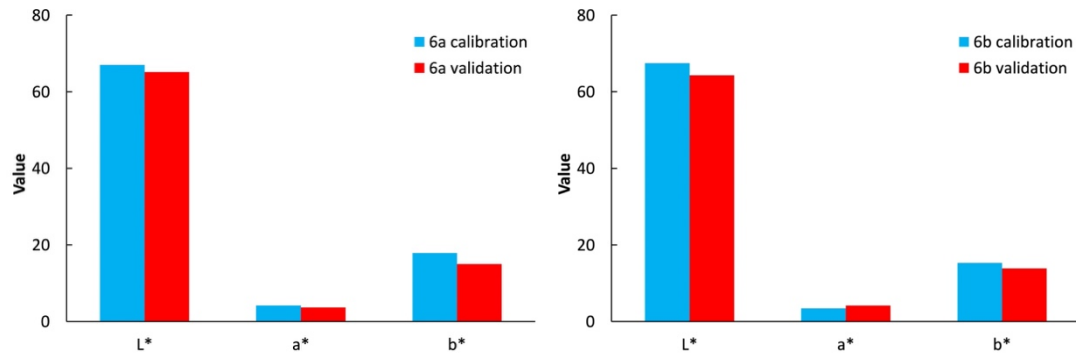


Figure 3.25 Calibration (average values) and validation L*a*b* values obtained from Location 6's soils.

3.4 Conclusions

This chapter explored the use of MSP and L*a*b* colour values in combination with chemometrics to differentiate between soil samples from different locations. This was the first demonstrated use of MSP on the quartz-recovered fine fraction of soils, and hence also the first use of chemometric methods to interpret MSP data obtained from this fine fraction. MSP analysis was conducted on the isolated quartz-recovered fine fractions of these soils, to allow for detection of the variation that could not be identified by analysis of the bulk material. PCA and LDA conducted on this data allowed for some discrimination between samples that produced visually similar MSP spectra.

The use of the entire MSP dataset for PCA resulted in most of the soil samples forming clusters based on the location from which they were collected, however, there was also a great deal of overlap throughout most of these clusters. Location 7 was the only class that was able to be entirely discriminated within the scores plots, due to its distinctive colour. Using the entire replicate MSP dataset for LDA caused the calibration accuracy to be greatly overestimated, with the validation dataset achieving a much lower accuracy. This model was able to correctly predict the locations of samples from Locations 7, 2, and 22, illustrating that it is possible to correctly classify groups that were not entirely discriminated within the PCA scores plot, however, the confidence surrounding these predictions was low.

Utilising the average MSP spectra for PCA improved the separation achieved within the scores plot, allowing for the complete discrimination of Locations 2, 7, 16, and 22. LDA conducted on this dataset showed a significant improvement in the validation accuracy, with the origin of samples from Locations 4, 7, 12, 16, and 22 all being predicted correctly. While some of these locations were well separated from the rest of the population, others were close to being assigned to incorrect classes, and the accuracy of those predictions should not be relied upon.

Converting the MSP spectra into 1931 L*a*b* values before PCA resulted in most locations achieving clustering, but with minimal separation from other classes. This may have been due to the reduction in noise, or exclusion of variation detected within the UV or IR regions of the MSP spectra. While these results indicated potential for a higher classification accuracy within LDA, this was not the case, and the validation accuracy of the model was poor. The omission of PC-3 when conducting LDA meant that important variation was being lost, highlighting the importance of utilising a sufficient number of samples for chemometrics. It is believed that the inclusion of more PCs would have resulted in greater classification accuracy.

Overall, utilising MSP data for chemometric analysis emphasised the large degree of intra-variability within soil samples, due to their naturally occurring environments. Whilst the dataset for this study was limited by the number of samples per location, the intention was to imitate sample size requirements from a forensic case work scenario as closely as possible. In these situations, there is usually minimal sample available for analysis. The MSP spectra, and resulting L*a*b* values, were most closely correlated to the colour and appearance of the soil, as expected. Therefore, the PCA and LDA models were able to discriminate between soils that were dissimilar in appearance but had a harder time differentiating between soils similar in colour. Additionally, reflectance spectra can often be noisy and not reproducible, and this was accentuated even more by performing complex statistical analysis on the data.

It was also shown that human interference with the soil's environment may be influencing the composition of the soil and should be explored further. Similarly, this

chapter highlighted the importance of sample preparation when using MSP for analysis, as samples that are not sufficiently thick will cause UV absorption that can interfere with the chemometric analysis; these results can be used to educate forensic examiners on how to account for these issues if encountered when analysing trace samples in casework. It is recommended that MSP analysis of soils be utilised for pairwise comparisons, or to quickly rule out samples that appear visually distinctive using a statistical measure of differentiation. To maximise the information obtained and decrease the uncertainty surrounding source determination, it should also be used in sequence with other analysis techniques, such as those explored in subsequent chapters of this thesis.

Chapter 4. Analysis of the Quartz-Recovered Fine Fraction of Western Australian Soils using Vibrational Spectroscopy Paired with Chemometrics

Portions of this chapter have been published, or submitted for publication, in the following articles:

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics for the forensic analysis of Western Australian sandy soils." *Forensic Chemistry*, 2022. 28: 100412.

T. G. Newland, K. Pitts, and S. W. Lewis. "Negative result: Application of Raman spectroscopy to the forensic analysis of an arid, sandy, soil." *Forensic Science International: Reports*, 2022. Submitted for publication.

4.1 Introduction

Soils submitted for forensic analysis are complex mixtures consisting of varying proportions of inorganic minerals, organic biological matter, living microbiomes, and other artificial materials (4). While many techniques can be utilised for their chemical characterisation, most focus on just one aspect of their composition. For example, X-ray diffraction (XRD) and laser-induced breakdown spectroscopy (LIBS) are commonly used to detect the inorganic components within soil samples (7, 11, 62), while high-performance liquid chromatography (HPLC) and gas chromatography (GC) examine organic components (6, 32, 59). Vibrational spectroscopy is a field that can be utilised to characterise both inorganic and organic components within soil, as well as contaminants such as petroleum and pesticides, providing a more detailed examination of its contents (4, 17, 29, 77, 91, 119). Raman and infrared (IR) spectroscopy can provide qualitative and semi-quantitative analysis of the composition of soil samples. They also have the advantages of being non-destructive, rapid, relatively sensitive, and only requiring small amounts of sample, all of which are advantageous for the examination of forensic trace evidence (4, 29, 120).

Raman spectroscopy has previously been investigated for its application to the analysis of other forms of forensic trace evidence, including drugs (78), fibres (82), and paints (80). However, its use for forensic soil analyses has been limited, for which there are two main reasons. The first is that Raman spectrometers tend to not be as widely available, and many forensic laboratories are not currently equipped with the appropriate instrumentation. The second is that Raman spectra of soils suffer from a large amount of fluorescence interference, usually caused by humus, that masks the characteristic peaks of the underlying soil components (77, 121-123). Therefore, it is important to consider alternative sample preparation methods that may help decrease sample fluorescence (77, 122). Generally, sample preparation methods with multiple steps are not preferred for forensic applications as they can further change the sample and potentially lead to contamination and loss of sample, which is commonly only present in trace amounts. Additionally, the sample preparation

may not be necessary or successful at reducing fluorescence in soils already low in organic material, such as those found on the Swan Coastal Plain in Western Australia.

An alternative approach is the use of IR spectroscopy, which provides complementary information to Raman spectroscopy. Not only is it widely available, with many forensic laboratories already owning IR spectrometers, but it is also simple to use (87). Limited studies have been published that utilise IR spectroscopy to analyse forensic soil samples. Of these studies, most have also used chemometrics to interpret their results, allowing for discrimination between different soil types (4, 64, 65, 76). Even with the limited applications presented in publications, the technique is still considered useful in certain cases, especially as part of a forensic questioned versus known comparison. However, all of these studies have involved very different soil types to those found in the dry, arid climate of Western Australia (4, 33, 64, 65, 76). The majority of them also analyse the bulk soil, which is impractical for Swan Coastal Plain soils, which are rich in quartz grains and have minimal variation in their bulk chemistry (54, 94). Spectroscopic analysis of the quartz-recovered fine fraction of soils has never been demonstrated in the open literature but holds potential for revealing variations in soil samples that are limited in size and composition (94).

In this chapter, Raman spectroscopy and attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy were explored for the forensic analysis of the quartz-recovered fine fraction of soils from the Swan Coastal Plain in Perth, Western Australia. Principal component analysis (PCA) was performed on the baseline corrected and normalised ATR-FTIR absorbance spectra to assess whether soils could be distinguished based on their location or other attributes of the soil, such as visual appearance, dune system, or location type. Linear discriminant analysis (LDA) was then used to classify and predict replicate samples using the test set validation method, and the accuracy of the model was assessed.

4.2 Experimental

Soil samples were collected as outlined in Sections 2.2 and 2.3.1. The quartz-recovered fine fractions were prepared from each of the soil samples, as described in Section 2.4. Raman spectra were collected from the extracted quartz fine fractions as described in Section 2.5.2. ATR-FTIR spectra were collected from the extracted quartz fine fractions as described in Section 2.5.3. PCA and LDA were then conducted on the resulting data as outlined in Section 2.6.

4.3 Results and discussion

4.3.1 Analysis of soils using Raman spectroscopy

Raman spectra were collected from the quartz-recovered fine fractions of Swan Coastal Plain soil samples (outlined in Chapter 2: Table 2.1), mounted on glass slides (Figure 4.1). These spectra all exhibited strong levels of fluorescence, which masked all peaks characteristic of the soil components. This is a common occurrence that has been encountered previously with Raman analysis of soil samples despite using a range of different excitation wavelengths and experimental conditions (77, 98, 121, 122). A small sharp peak was visible above the fluorescence in Sample 5b (460 cm^{-1}) indicating the possible presence of quartz within the fine fraction (30, 77), however, this peak was masked on all other samples. There was also a broad hump evident in Sample 4b ($\sim 1300 - 1600\text{ cm}^{-1}$) that may be due to the presence of carboxyl groups and humic substances within the sample (123, 124), but this could not be ascertained due to the imperceptible structure and breadth of the peak.

Reduction of the fluorescence was attempted by experimenting with the laser intensity and integration time; however, this had no beneficial effect on the spectra (Figure 4.2). The data was also subjected to baseline correction methods, but this too was unsuccessful at eradicating the interference from fluorescence (Figure 4.3). Therefore, Raman spectroscopy was deemed unsuitable for analysis of the quartz-recovered fine fraction of soils, and the spectra obtained were not utilised for any further chemometric analyses.

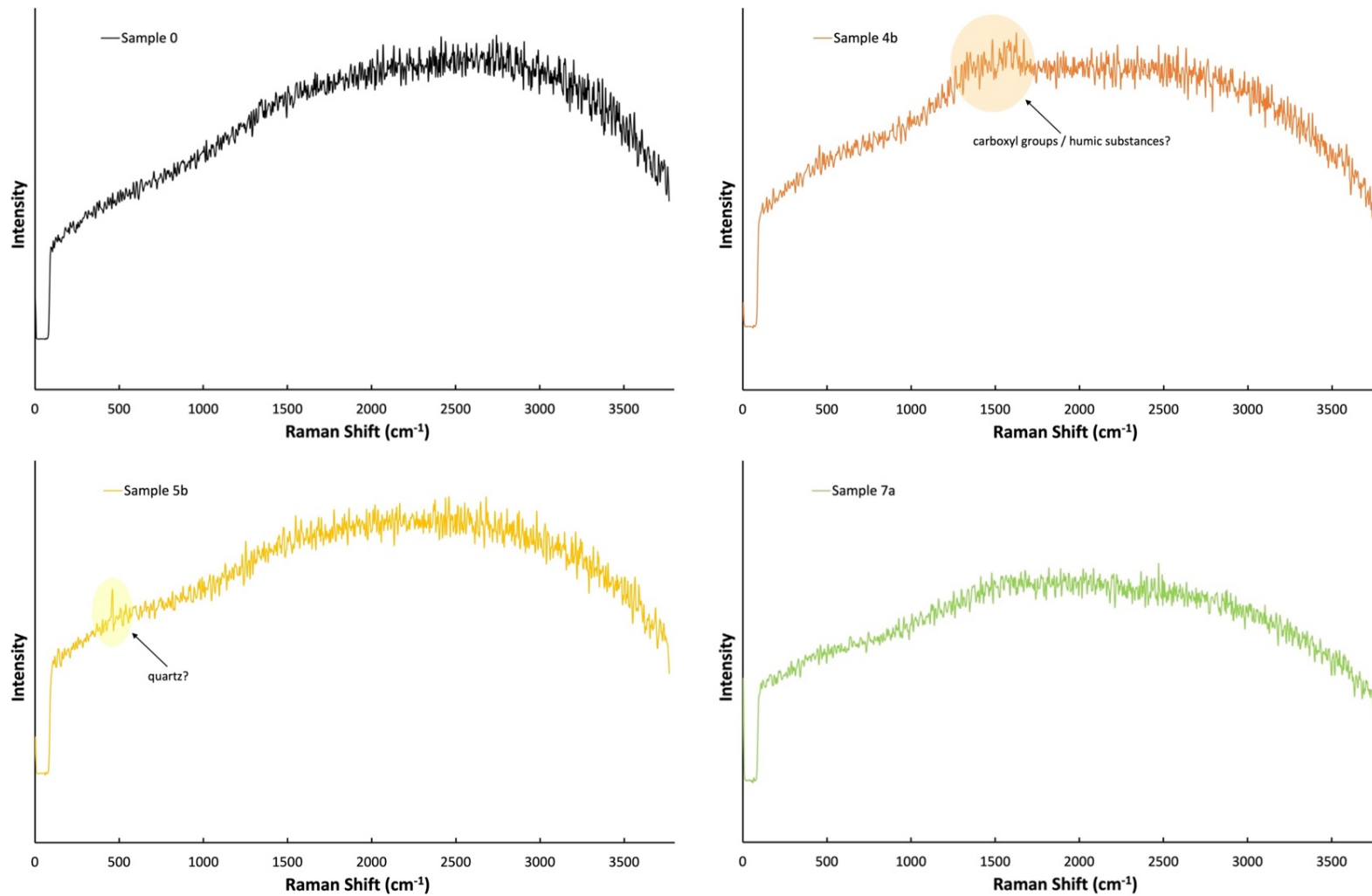


Figure 4.1 Raman spectra (unprocessed) of four different soil samples from the Swan Coastal Plain, illustrating the high degree of fluorescence that is masking potential characteristic peaks.

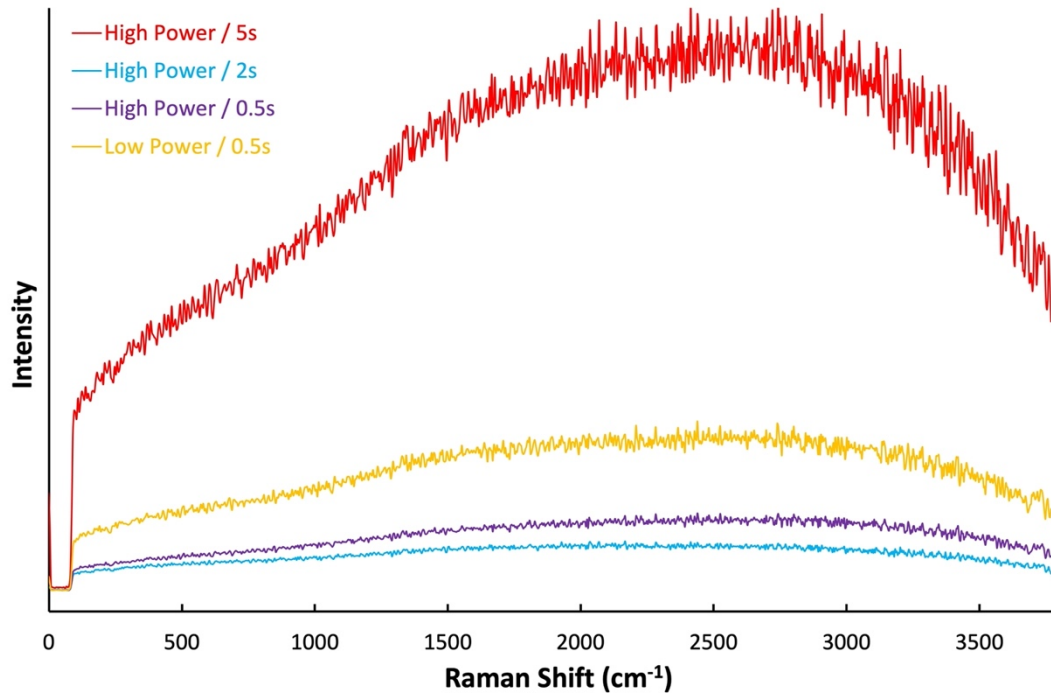


Figure 4.2 The effects of laser intensity (high vs. low power) and integration time (in seconds) on Sample 0's Raman spectra.

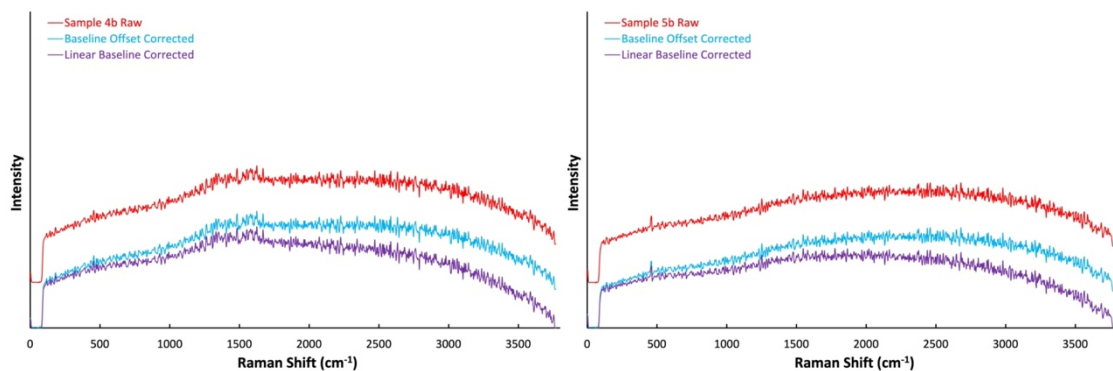


Figure 4.3 The effects of baseline correction (baseline offset vs linear baseline) on Sample 4b and 5b's Raman spectra, with visible peaks remaining unchanged and unable to be enhanced across all spectra.

4.3.2 Analysis of soils using ATR-FTIR spectroscopy

ATR-FTIR absorbance spectra were collected from the quartz-recovered fine fractions of 45 soil samples (outlined in Chapter 2: Table 2.1), a selection of which are illustrated in Figure 4.4. Despite these soils originating from different regions and dune systems within Perth, the majority of the spectra were indistinguishable from one another in terms of the main compounds that make up the soil samples. Peaks

were tentatively assigned based on reference spectra of common minerals and compounds known to be found within Swan Coastal Plain soils (Table 4.1). Most of the soil samples were dominated by quartz and kaolinite, with some variation seen in the peak heights located in the fingerprint region ($400 - 1200 \text{ cm}^{-1}$) as well as the peak heights and ratios of the characteristic kaolinite peaks at $3630/3695 \text{ cm}^{-1}$. Other apparent minerals within the samples included gibbsite, goethite, and hematite, with some additional contribution from organics (humus and $\text{H}_2\text{O}/\text{CO}_2$), all present in varying intensities. Whilst some subtle differences could be perceived between sample spectra, identifying all possible variations manually would be a labour-intensive process that involves careful human examination and judgement. Therefore, chemometric methods were applied to efficiently and objectively identify and maximise the differentiation between samples.

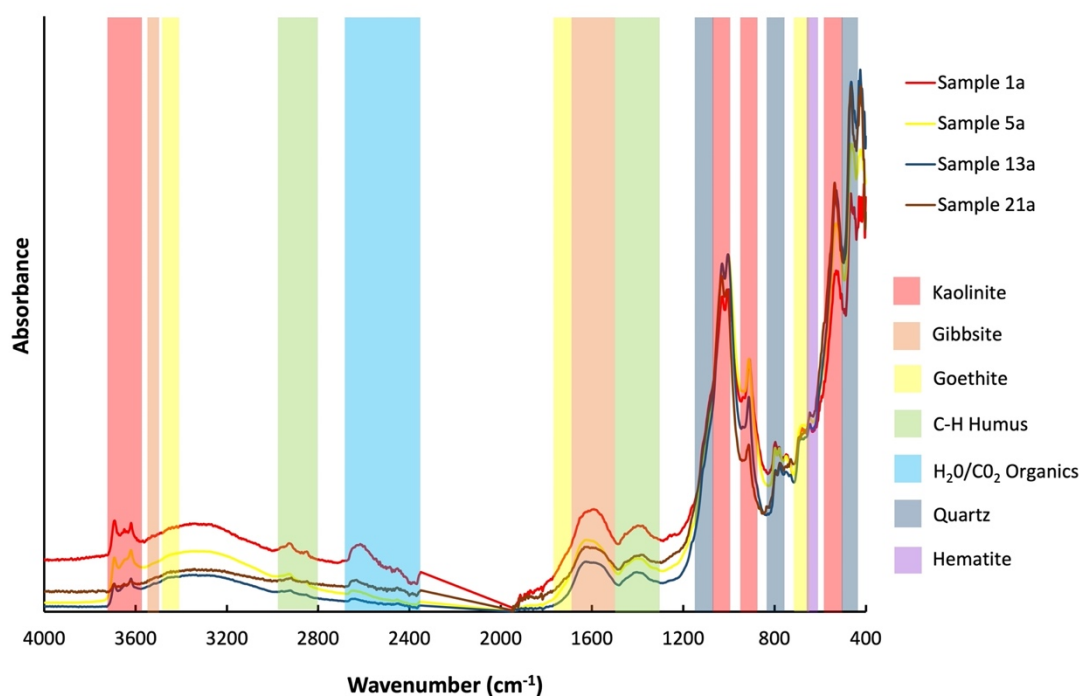


Figure 4.4 Baseline corrected and normalised ATR-FTIR absorbance spectra showing the similarities in composition of a selection of soil samples collected from differing locations within the Swan Coastal Plain. Annotations were based off peak assignments outlined in Table 4.1.

Table 4.1 Common minerals found in Swan Coastal Plain soils and their associated IR peaks.

Mineral/Compound	IR Peaks (cm ⁻¹)
Kaolinite ²	3695, 3620 , 1113, 1031, 1008, 912 , 754, 698, 538, 470, 430
Quartz ²	1082 , 790/778, 692, 459
Hematite ²	1179, 1117, 1084, 610, 548, 470
Gibbsite ²	3454 (broad), 1585 , 1425, 975, 578
Vermiculite ³	3330 (broad), 1640, 944 (skewed), 815, 719
Goethite ²	3457 (broad), 3102 , 1641, 1425, 900, 803, 664, 568, 461/426/402
Microcline feldspar ¹	1129, 985 , 768, 725
Humic acid (sodium salt) ³	3277 (broad), 2919/2850, 1563, 1380, 1090- (broad)
Calcite ²	1416 , 874, 712

¹Thermo Fisher Scientific for Nicolet FTIR, "HR comprehensive Forensic FT-IR".

²Thermo Fisher Scientific, 2008, "HR Inorganics".

³SensIR technologies, 2001, "Common chemicals by Diamond ATR".

4.3.3 Principal component analysis

4.3.3.1 Discrimination using location-based chemical data

PCA performed on these spectra revealed that 95.8% of the total variance in the dataset could be described by the first five principal components (PCs), as illustrated in the scree plot (Figure 4.5).

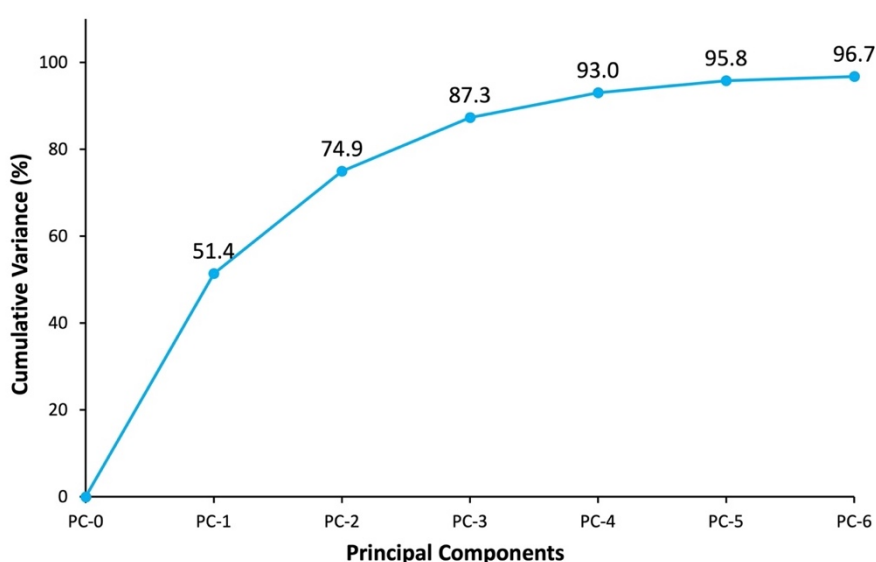


Figure 4.5 Scree plot depicting the cumulative variance in the ATR-FTIR dataset retained by each PC.

Three-dimensional score plots generated using the first three PCs (Figure 4.6) resulted in the discrimination of nine different locations from the rest of the population; Locations 0, 3, 8, 9, 10, 11, 15, 18, and 22 all achieved visual separation across PCs 1-3. The inclusion of PCs 4 and 5 allowed for further discrimination of Locations 2, 5, 12, 16, and 17, leaving an additional nine locations unable to be individualised. PC-6 did not allow for any further discrimination between samples, so was not incorporated into the visual model to avoid overfitting of the data. All samples exhibited varying degrees of intra-location separation, displaying clustering based on individual samples instead of overall location. Locations within which this was most prominent, e.g., 1, 8, 9, and 18, were all minimally accessible bushland sites with grey-coloured sands, indicating higher variability within these soils.

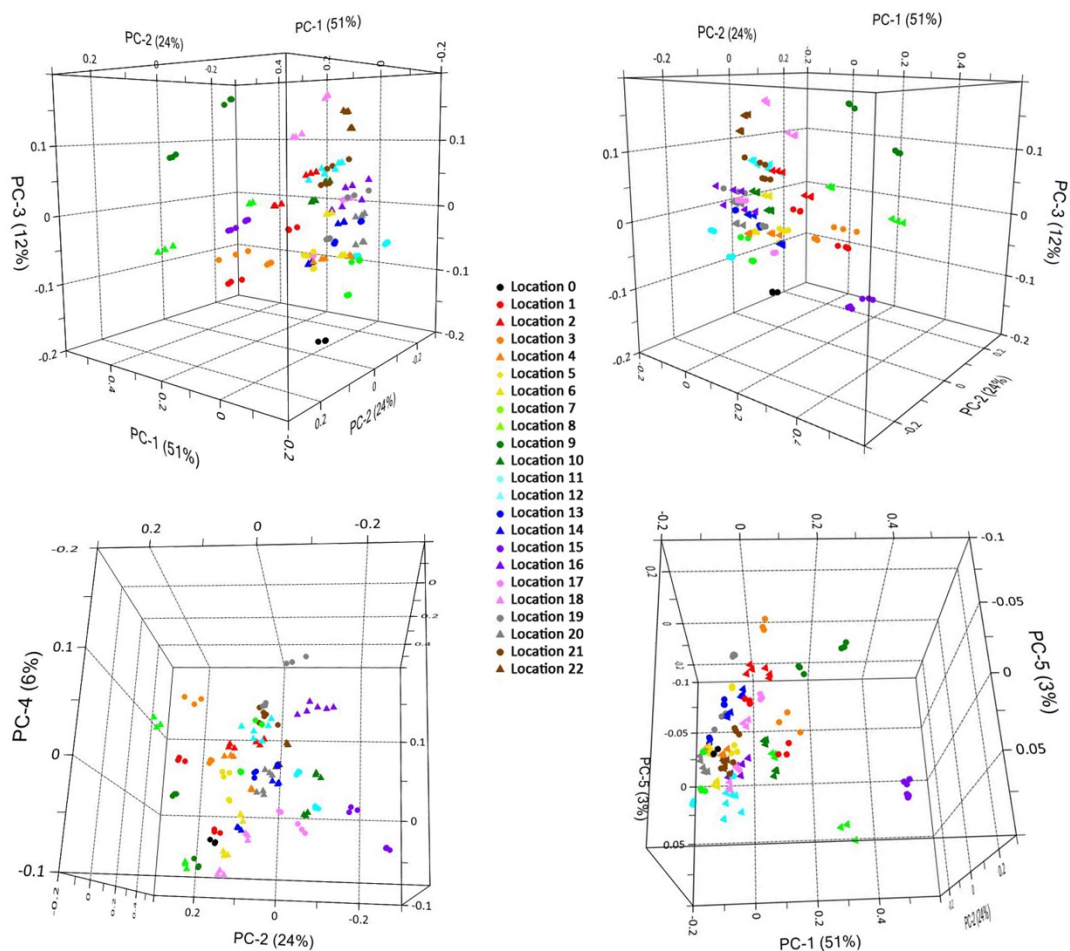


Figure 4.6 3-dimensional PCA scores plots generated using the first five PCs, showing the variability of soil samples from different locations based on their corresponding ATR-FTIR spectra.

The factor loadings for the first five PCs were studied to determine the regions in the ATR-FTIR spectra that were associated with, and likely responsible for, the discrimination of samples along each component. Many of the peaks within the loadings, however, were distorted due to a high degree of interference and overlap between positive and negative correlations, making mineral identification challenging. Identification was attempted based on the known IR spectra of common minerals and compounds found within Swan Coastal Plain soils, found in Table 4.1. Minor variations in mineral content would be expected to create differences in the overall spectra, but it was difficult to completely identify from the loadings alone, so ATR-FTIR spectra obtained from representative samples situated at the extremes of each PC were also used to inform the mineral associations made using the loadings plots (Figure 4.7).

Variation along PC-1 was attributed to negative correlations with kaolinite and quartz, and positive contributions from humus and/or calcite. Samples from Location 15 were best separated in a positive direction along PC-1, therefore likely containing higher relative concentrations of humus or calcite, and lower concentrations of kaolinite and quartz. This was confirmed by inspection of spectra from Location 15 and Location 20 (situated at the opposite negative end of PC-1); Location 20's spectra were dominated by kaolinite and quartz, while Location 15's spectra were lacking these peaks, instead containing other clay-type minerals and higher relative concentrations of calcite. Location 15 soils originated from the Quindalup dune system, nearest to the coast, which are known to contain lower concentrations of quartz (36, 94). Quindalup soils also commonly contain higher levels of calcite from the limestone bedrock and shell fragments from marine life (43, 47).

PC-2 suggested a negative correlation with humus and/or calcite, and a positive correlation with quartz, with possible smaller contributions from gibbsite and goethite. Samples from Location 15 were again the best separated from the rest of the population due to negative scores along PC-2, confirming that they contained higher relative levels of calcite in their quartz-recovered fine fraction. Their low amounts of quartz and gibbsite were again expected due to their coastal positioning; Swan Coastal Plain soils tend to show a decrease in both quartz and gibbsite nearer

the coast (36, 94). Samples from Locations 8 and 9 were best separated in a positive direction along PC-2, due to relatively higher concentrations of quartz and gibbsite/goethite, which was confirmed by inspection of their spectra.

PC-3 showed a negative correlation with kaolinite and humus/calcite, and positive correlation with quartz, as well as a smaller contribution from gibbsite. Samples from Location 0, followed by Location 7, were situated at the most negative point along PC-3, indicating that they may contain the highest levels of kaolinite and humus/calcite in their soil. This was reinforced by their spectra which were dominated by kaolinite peaks. These soils all originated from the Pinjarra Plain, which represents an accumulation of fluvial deposits, and hence these soils can contain relatively high concentrations of this mineral (43, 94). Samples from Locations 9, 18, and 22 were positioned positively along PC-3 due to containing relatively higher concentrations of quartz and gibbsite, as evident in their spectra. These soils were all grey-coloured and originated from bushland sites that were close to the border of the Spearwood / Bassendean dunes.

While PC-4 and PC-5 were utilised to visualise the dataset, they together contributed to the separation of only five additional groups. PC-4 showed association with several different compounds, such as kaolinite, quartz, goethite, gibbsite, humus, and calcite, and hence exhibited a high degree of interference. The spectra for Locations 5, 16, and 17 (locations that achieved discrimination through the inclusion of PC-4) were compared to the loadings plot to identify which compounds specifically were responsible for their separation (Figure 4.8). Location 5 samples appeared to be lacking in both humus and calcite compared to the other soils, while Location 16 soils showed lower concentrations of kaolinite and gibbsite, and higher concentrations of quartz.

The loadings for PC-5 were also quite complicated, suggesting negative correlations with goethite, gibbsite, and hematite, as well as a smaller contribution from quartz, and positive correlations with carbon-based organics, goethite, and calcite. Soils from Locations 2 and 12 were able to be individualised by variation attributed to PC-5; Location 2 soils contained higher concentrations of gibbsite and humus, while

Location 12 soils showed a high concentration of materials within the fingerprint region. Location 2 soils originated from a manicured garden bed within a public park, so the addition of organic mulch to the soil may be responsible for the higher amounts of humus present. While tentative assessments could be made regarding the identification of minerals contributing to the loadings, it is necessary to allow for the possibility that compounds other than the common minerals listed in Table 4.1 were detected through ATR-FTIR analysis. In the subsequent chapter, XRD was used post-ATR-FTIR for cross-referencing of the minerals detected to confirm the above associations.

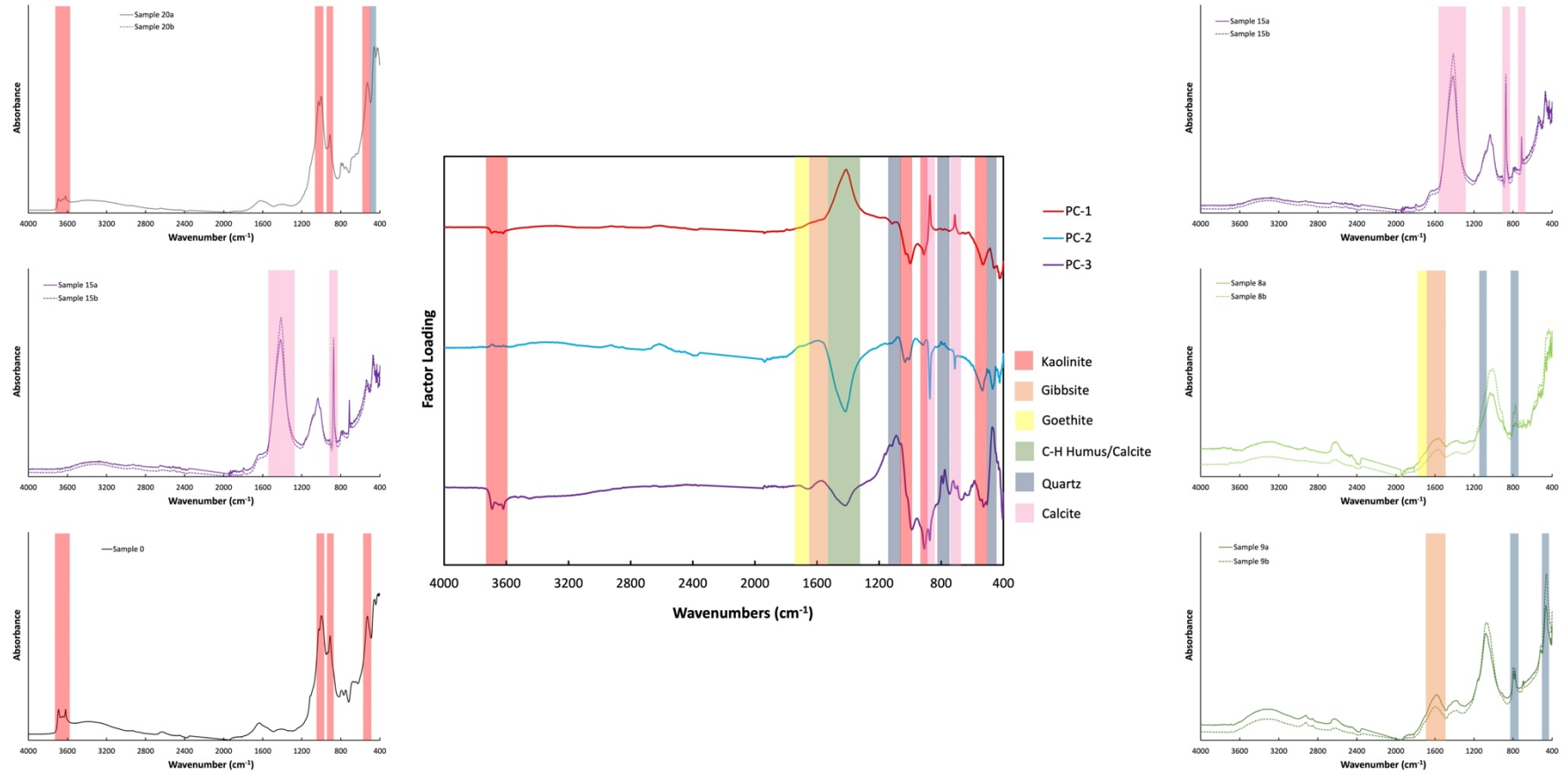


Figure 4.7 Factor loadings plot of PCs 1-3 for PCA of the soil ATR-FTIR dataset, along with representative spectra obtained from samples situated at the extreme points of the scores plot. The main peaks of interest have been highlighted and annotated with their contributing compounds.

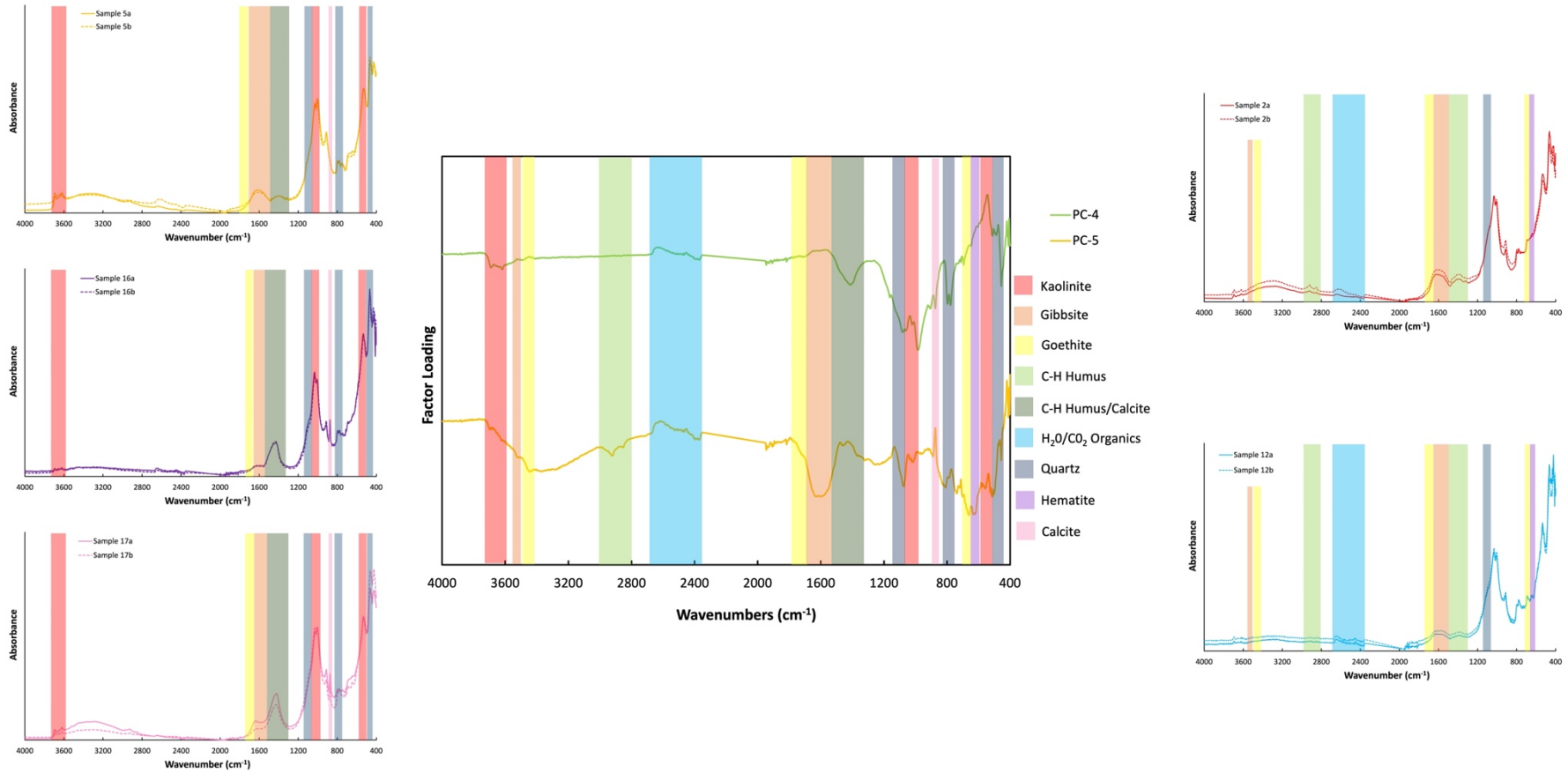


Figure 4.8 Factor loadings plot of PCs 4-5 for PCA of the soil ATR-FTIR dataset, along with representative spectra obtained from samples that achieved discrimination within the scores plot based on the variation attributed to PCs 4 and 5. The main peaks responsible for the separation between samples have been highlighted and annotated with the contributing compounds.

4.3.3.2 Discrimination using feature-based data

Improved separation was investigated by attempting to reveal greater structure in the ATR-FTIR data based on other attributes of the soil, such as their visual appearance. Further interpretation of chemical-based PCA data using features such as soil colour has previously not been demonstrated within the open literature. Colour coding of the soil samples in the PCA model based on visually assigned groupings (outlined in Chapter 2: Table 2.1) did display clustering, however, these clusters were widespread and overlapping with each other (Figure 4.9).

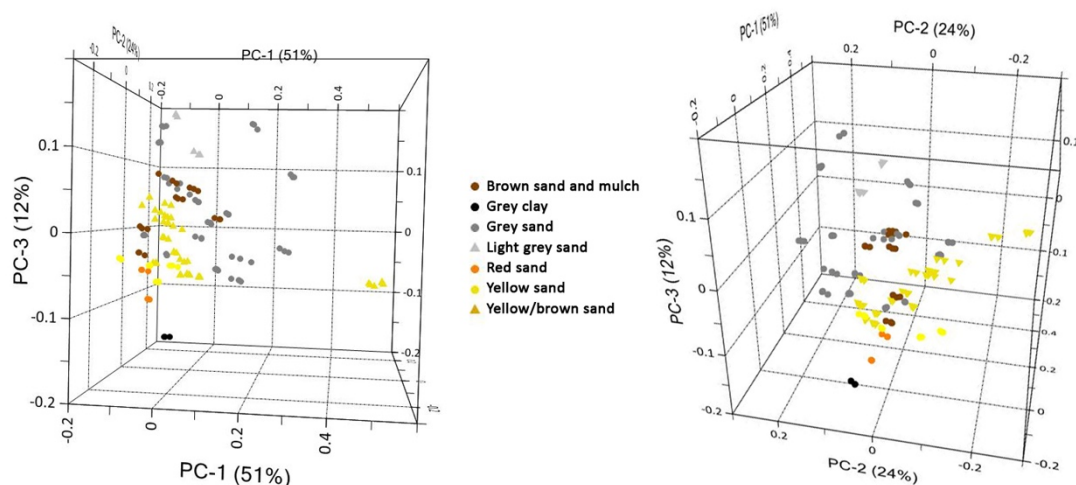


Figure 4.9 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding ATR-FTIR spectra.

The grey sands again showed the most variation, being spread extensively throughout the scores plot. The grey clays, red sands, and light grey sands were able to be discriminated from the rest of the population, however, this may be due to their smaller sample sizes. Significant overlap was observed in clusters of similar colours, for example, yellow, yellow/brown, brown, and red sands, as a lot of the variation across the PCs was potentially due to minerals responsible for the red and yellow hues of the soils, such as goethite, hematite, and kaolinite (often stained red brown from iron oxides (125)). The clustering and separation achieved within this model was comparable to that achieved by the PCA models built using microspectrophotometry (MSP) data in Chapter 3, indicating that the minerals seemingly responsible for most of the variance within the ATR-FTIR model are also

associated with the apparent colour of the soils. Whilst this method of interpretation was useful for visualising the dataset, it did not add further discrimination between soils similar in appearance.

The PCA model was then grouped according to the dune system from which the soil sample originated, outlined in Chapter 2: Table 2.1 (Figure 4.10). Well-defined clusters were formed for each dune system, however, the soils from Spearwood, Mixed (Spearwood & Bassendean), and Pinjarra Plain were grouped in very close proximity to each other, overlapping across most of the PCs. The degree of clustering achieved indicates that ATR-FTIR analysis can detect compounds within the soil samples in quantitative ratios characteristic of their dune system. Therefore, ATR-FTIR analysis could potentially provide forensic intelligence to investigators concerning specific regions where a soil sample may have originated. No additional structure in the dataset was visualised when grouping the soils according to the type of location they were collected from (Figure 4.11).

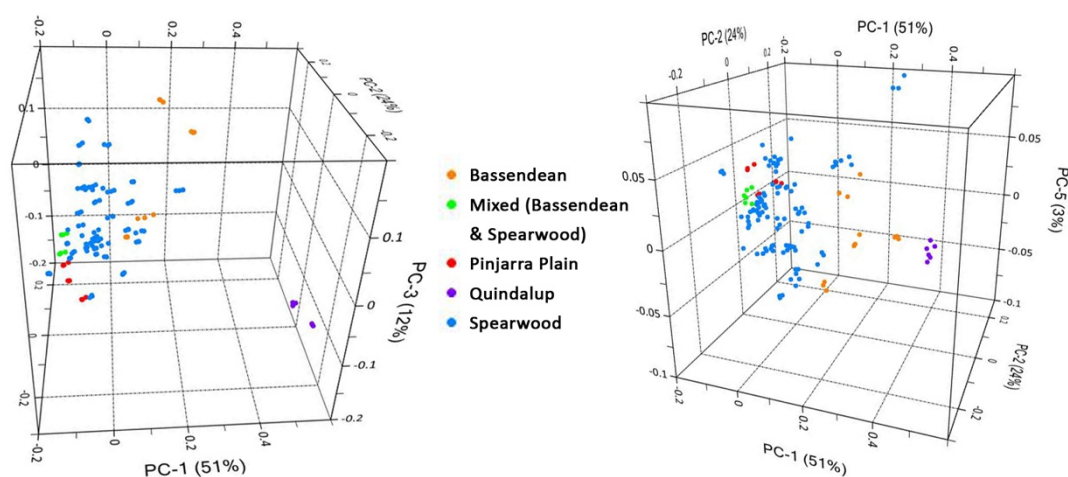


Figure 4.10 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding ATR-FTIR spectra.

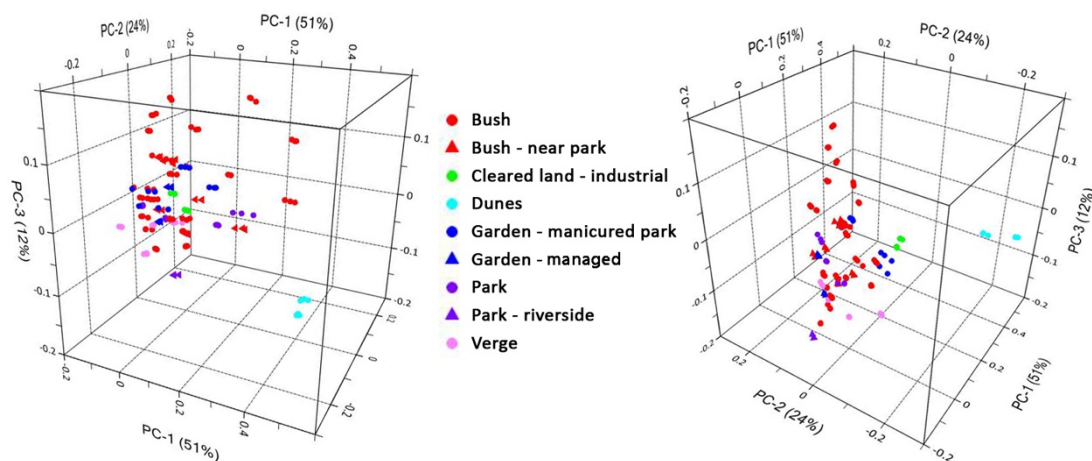


Figure 4.11 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding ATR-FTIR spectra.

4.3.3.3 Repeated attempts at discrimination using location-based chemical data

PCA was repeated on the soil ATR-FTIR data, this time omitting the locations that were previously able to be discriminated. PCA performed on only these nine locations revealed that 93.8% of the total variance in the dataset could be described by the first five PCs, as illustrated in the scree plot (Figure 4.12).

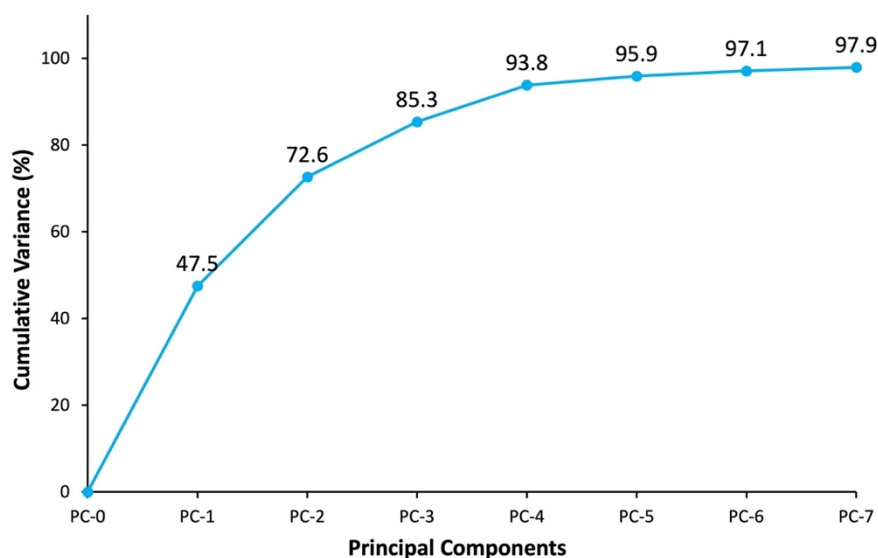


Figure 4.12 Scree plot depicting the cumulative variance retained by each PC in the reduced-sample soil ATR-FTIR dataset.

Three-dimensional score plots generated using the first three PCs (Figure 4.13) resulted in the discrimination of the remaining nine locations. While some clusters were situated in close proximity to other locations, there was no overlap between them. However, all locations again exhibited separation between their individual samples, with two apparent clusters created by each sample's replicates. The inclusion of PCs 4 and 5 when visualising the scores plot did not enhance the separation or clustering of any groups, and as a result, the variation attributed to these PCs was disregarded.

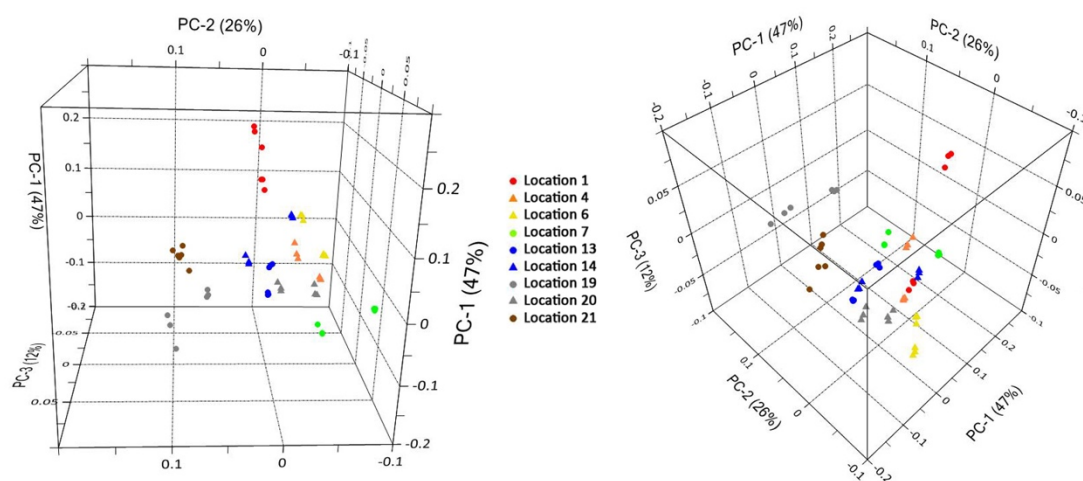


Figure 4.13 3-dimensional PCA scores plot (shown from two perspectives) generated using the first three PCs, displaying the variability of a reduced number of soil samples from different locations based on their corresponding ATR-FTIR spectra.

Most of the separation within the reduced-sample PCA model was due to variation within the peaks attributed to kaolinite and quartz, with differences in gibbsite concentration also apparent across the loadings from PCs 1-3 (Figure 4.14). PCs 1 & 2 showed additional influence from peaks produced by organic components within the soil samples (mainly humus $\sim 1400\text{ cm}^{-1}$), while PC-3 showed positive contribution from several peaks identified as goethite ($670, 3100\text{ cm}^{-1}$) and hematite (615 cm^{-1}). While the compounds attributed to variation within the reduced-sample PCA model are the same ones that accounted for the variation within the previous PCA model, subtle differences in their concentrations could be detected once the other samples were removed. Many of the original samples had more significant differences in their spectra that overshadowed the subtle variations seen amongst similar soils, thus,

dominating the loadings. These samples had to be removed from the dataset once separated to allow for detection of the more subtle variations shown between similar samples and utilisation of these differences for their discrimination.

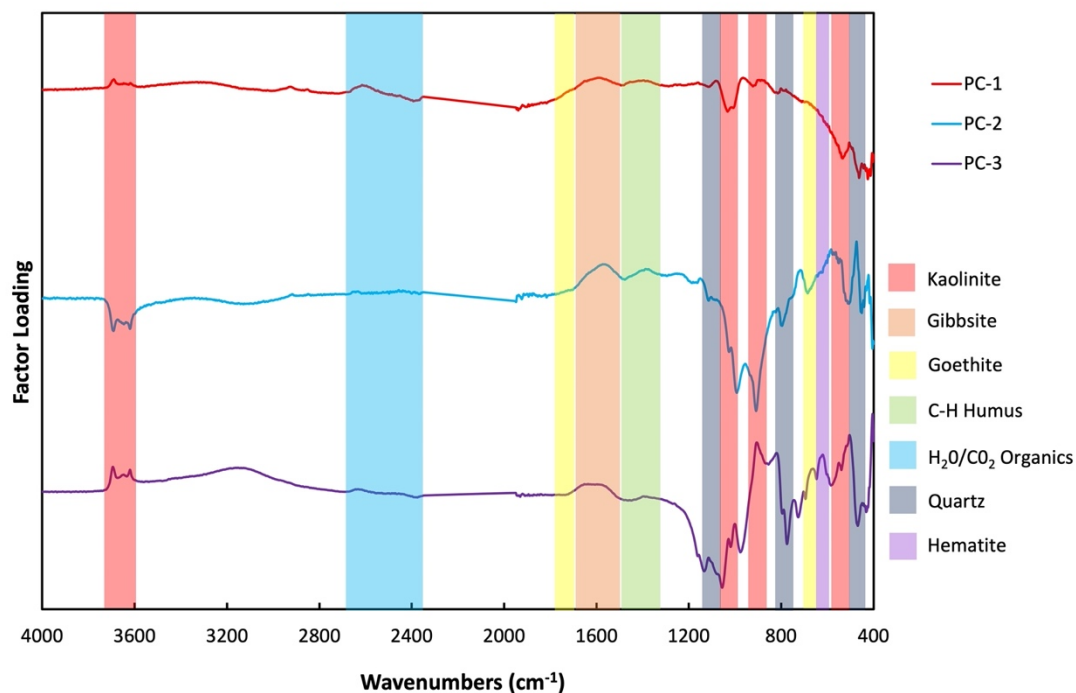


Figure 4.14 Factor loadings plot of PCs 1-3 for PCA of the reduced-sample soil ATR-FTIR dataset.

4.3.4 Linear discriminant analysis

LDA was used as a classification technique to predict the original locations of soil samples based on their chemical data. The ATR-FTIR data was split into two datasets for LDA; two of the three replicates per sample were used to build the calibration model, and the remaining one replicate per sample was used as a validation set and predicted onto the model to assess its performance. The validation set consisted of replicates that were not utilised within the calibration set to ensure no overlap between models and a realistic estimation of the model's performance (105). Location 0 was excluded from LDA, due to its limited sample size (only one sample). LDA was performed on the ATR-FTIR dataset post-PCA, with each location treated as an individual class. This produced a single discriminant function that was used to classify replicates from the calibration dataset.

This discriminant model returned a calibration accuracy of 96.6% (Table 4.2). While five PCs were shown to achieve maximum separation between locations through PCA, it was only possible to utilise the first four PCs for LDA, due to the available number of samples per location; as LDA involves matrix inversion, the number of variables must be smaller than the number of samples involved (30, 105). As Location 12 was only separated from the population across PC-5, these samples were no longer able to be fully discriminated, and as such, one of them was misclassified. Half of Location 6's soils were also misclassified, which was to be expected as they overlapped with several other clusters in the comprehensive PCA scores plot due to some inter-sample separation and were unable to be discriminated without further manipulation.

This LDA model was then used to predict the locations of 44 samples from the validation dataset. 41 samples, or 93.2%, had their location correctly predicted (Table 4.3). This shows that the model was able to accurately identify some groups that overlapped within the scores plot, though the certainty of these predictions is unreliable. Location 6 and 12 had samples predicted incorrectly due to some of their replicates being misclassified within the calibration dataset. There were also clear similarities between soils from Locations 12 and 21, as both these groups had their samples incorrectly predicted as originating from each other. Apart from both being collected from dense bushland locations within the Spearwood dune system, no further commonalities between these groups could be identified. Despite the degree of overlap between Locations 1, 4, 6, 7, 13, 14, 19, 20, and 21 within the scores plot and the resulting requirement for further discrimination using an isolated PCA model, the confidence surrounding these predictions was moderate. The associated discriminant values (Table 4.4), measuring the distance between each sample and the centroid of each class, showed that approximately half of these samples were close to being predicted as originating from at least one other location, while the other half were well separated from all other location classes. The discriminant values could have potentially been improved by incorporating PC-5 into the analysis or utilising a larger number of samples within each class.

Table 4.2 Number of correct vs incorrect location classifications for samples in the ATR-FTIR calibration set using a 4-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Classified	% Correct
1	4	0	-	100
2	4	0	-	100
3	4	0	-	100
4	4	0	-	100
5	4	0	-	100
6	2	2	4	50
7	4	0	-	100
8	4	0	-	100
9	4	0	-	100
10	4	0	-	100
11	4	0	-	100
12	3	1	21	75
13	4	0	-	100
14	4	0	-	100
15	4	0	-	100
16	4	0	-	100
17	4	0	-	100
18	4	0	-	100
19	4	0	-	100
20	4	0	-	100
21	4	0	-	100
22	4	0	-	100
			% Total Correct	97

Table 4.3 Number of correct vs incorrect location predictions for samples in the ATR-FTIR validation set using a 4-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Predicted	% Correct
1	2	0	-	100
2	2	0	-	100
3	2	0	-	100
4	2	0	-	100
5	2	0	-	100
6	1	1	4	50
7	2	0	-	100
8	2	0	-	100
9	2	0	-	100
10	2	0	-	100
11	2	0	-	100
12	1	1	21	50
13	2	0	-	100
14	2	0	-	100
15	2	0	-	100
16	2	0	-	100
17	2	0	-	100
18	2	0	-	100
19	2	0	-	100
20	2	0	-	100
21	1	1	12	50
22	2	0	-	100
			% Total Correct	
				93

Table 4.4 Discriminant values of replicates from the ATR-FTIR validation dataset (rounded to three decimal places), with correct predictions shaded green and incorrect predictions shaded red. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

Sample Replicate	Discriminant Values																						Predicted Location	Closest Value Within (%)
	Location 1	Location 2	Location 3	Location 4	Location 5	Location 6	Location 7	Location 8	Location 9	Location 10	Location 11	Location 12	Location 13	Location 14	Location 15	Location 16	Location 17	Location 18	Location 19	Location 20	Location 21	Location 22		
1a	-4.475	-22.093	-5.212	-26.287	-13.373	-15.566	-54.611	-31.192	-148.690	-75.951	-64.881	-23.981	-29.176	-19.411	-622.493	-67.059	-39.664	-56.996	-48.640	-31.569	-29.007	-44.895	1	16
1b	-4.546	-16.833	-10.527	-33.021	-17.674	-14.268	-67.386	-24.845	-122.463	-63.739	-68.717	-20.027	-32.084	-19.022	-592.916	-66.681	-33.823	-34.985	-55.386	-34.388	-26.514	-31.748	1	132
4a	-36.558	-50.995	-33.550	-3.632	-8.393	-12.019	-7.681	-110.783	-282.884	-94.923	-18.218	-39.450	-6.674	-13.505	-702.339	-47.785	-38.831	-122.585	-21.093	-4.543	-34.844	-69.493	4	25
4b	-27.230	-40.506	-23.118	-3.353	-5.581	-9.648	-11.187	-92.246	-255.293	-89.444	-23.807	-30.148	-6.366	-11.830	-693.069	-45.665	-37.335	-107.103	-19.174	-5.056	-27.100	-58.907	4	51
6a	-27.037	-45.002	-27.352	-4.127	-5.612	-6.794	-13.191	-93.066	-254.962	-92.775	-25.000	-35.368	-8.136	-11.585	-700.096	-53.547	-37.633	-104.821	-27.062	-5.718	-33.256	-63.340	4	36
6b	-10.643	-27.020	-17.866	-21.207	-12.438	-6.057	-48.148	-44.941	-162.004	-84.896	-59.283	-22.131	-23.617	-18.308	-679.072	-70.034	-42.239	-48.043	-48.935	-21.721	-27.301	-37.612	6	76
7a	-64.570	-88.429	-59.634	-11.449	-22.681	-29.351	-4.208	-158.569	-364.198	-136.962	-23.974	-72.392	-19.634	-32.472	-785.371	-71.604	-65.822	-179.459	-34.902	-15.130	-64.821	-113.482	7	172
7b	-64.520	-73.096	-54.388	-10.711	-22.747	-30.964	-4.162	-154.816	-349.692	-115.778	-16.847	-55.927	-13.133	-28.445	-741.934	-48.647	-55.565	-167.449	-18.199	-10.753	-46.710	-91.232	7	157
13a	-36.561	-35.074	-29.819	-7.554	-11.088	-15.955	-14.733	-104.435	-259.152	-66.522	-12.788	-25.283	-3.429	-9.520	-628.052	-25.363	-24.751	-107.196	-9.219	-4.438	-19.386	-47.408	13	29
13b	-24.990	-30.173	-20.593	-5.669	-5.699	-10.413	-16.499	-86.633	-236.989	-63.889	-16.384	-23.802	-3.668	-5.491	-617.795	-30.658	-21.563	-96.514	-14.100	-4.986	-19.970	-47.102	13	36
14a	-28.679	-22.517	-22.636	-13.994	-12.060	-17.732	-28.023	-85.577	-220.928	-41.479	-14.798	-19.497	-5.983	-5.078	-544.880	-17.542	-11.088	-88.943	-11.544	-10.258	-14.672	-36.234	14	18
14b	-15.477	-26.800	-16.096	-13.119	-6.616	-10.009	-31.615	-66.239	-201.595	-54.946	-26.467	-27.621	-11.472	-4.490	-572.209	-40.820	-16.566	-81.101	-29.450	-14.222	-26.736	-48.622	14	47
19a	-38.605	-31.711	-26.519	-12.495	-14.965	-24.917	-19.518	-103.680	-256.951	-57.262	-13.666	-24.617	-6.247	-11.499	-585.869	-17.334	-22.205	-114.151	-5.071	-10.159	-17.483	-47.633	19	23
19b	-55.722	-33.328	-38.824	-27.052	-31.207	-41.593	-34.334	-117.837	-261.021	-56.044	-23.255	-23.435	-15.332	-24.463	-575.452	-11.684	-30.554	-117.261	-4.132	-20.562	-14.754	-40.286	19	183
20a	-39.487	-49.939	-35.853	-4.336	-9.970	-13.151	-8.022	-114.470	-284.709	-92.136	-16.541	-37.341	-5.825	-13.862	-698.533	-43.624	-37.872	-121.652	-18.307	-3.761	-32.146	-65.502	20	15
20b	-32.029	-33.572	-27.159	-6.307	-9.212	-11.809	-15.351	-96.198	-248.122	-73.395	-18.605	-22.497	-3.948	-10.281	-656.630	-31.842	-29.308	-97.796	-12.357	-3.640	-18.245	-44.250	20	8
21a	-26.128	-9.291	-17.038	-27.663	-22.098	-26.225	-50.162	-62.236	-171.582	-41.241	-40.467	-5.114	-16.955	-16.570	-550.199	-20.190	-22.533	-60.134	-15.476	-22.280	-3.297	-14.228	21	55
21b	-23.509	-7.222	-16.814	-33.859	-25.837	-27.004	-61.264	-50.686	-148.251	-45.666	-53.894	-3.213	-23.474	-21.506	-563.260	-29.527	-28.823	-45.258	-24.437	-28.223	-3.792	-9.031	12	18

4.4 Conclusions

This chapter explored the use of Raman and ATR-FTIR spectroscopy for analysing the quartz-recovered fine fraction of Swan Coastal Plain soils, which has previously not been demonstrated in the open literature. The quartz-recovered fine fraction was used to allow for the detection of subtle variations in composition that could not otherwise be identified due to the quartz-dominated nature of the bulk material. This data was analysed using chemometric methods in an attempt to differentiate between soil samples from different locations. Whilst Raman analysis was unsuccessful, PCA and LDA conducted on ATR-FTIR data allowed for full discrimination of all samples, including those that produced visually similar spectra.

Raman spectroscopy has previously been demonstrated for the analysis of soils, however, in these reports, extensive sample preparation was required for success due to excessive fluorescence (77, 122). These studies were also conducted on soil types that were not comparable to the sandy soils found in Perth. Unfortunately, Raman analysis of the quartz-recovered fine fraction of sandy soils showed high levels of fluorescence, which masked all potential characterisation data captured within the spectra. No improvement was seen through further investigation with experimental conditions or corrections to the data. As a result, Raman spectroscopy was considered unsuitable for the analysis of the quartz-recovered fine fraction of soils.

The use of PCA on ATR-FTIR spectroscopic data resulted in the discrimination of all soils based on the locations from which they originated. Factor loadings showed that variation was seen within the mineral content of the soil, as well as the organic components. The initial PCA model showed appropriate clustering of locations, with more than half of the dataset able to be entirely separated from the rest of the population. The remaining indistinguishable cluster was extracted and again analysed by PCA without the separated samples, and all these overlapping locations were then able to be entirely discriminated. Both models did however, display some significant separation within location groups, with many samples clustering based on

individual samples rather than overall locations, indicating potential issues with sampling or a high level of variation within the soil across several sites.

LDA post-PCA was successful, achieving high classification and validation accuracy (96.6% vs. 93.2%). The model was able to correctly predict soils from locations that were overlapping within the initial PCA scores plot, though some were close to being assigned to incorrect classes; hence, there was not a high level of confidence surrounding their classification. The inclusion of additional PCs when conducting LDA may account for more of the variation between soils and result in higher-confidence predictions.

Overall, chemometric analysis of ATR-FTIR data was successful in discriminating between soils from different locations and predicting the locations of similar samples. The ATR-FTIR spectra indicated the presence of both mineralogical and organic components within the soil, and these showed correlation with the appearance of the soil and the dune system from which the soil originated. This meant that unlike the MSP methods demonstrated in Chapter 3, the ATR-FTIR PCA and LDA models were able to discriminate between soils that were similar in appearance. This is built upon in Chapter 5 with the exploration of XRD to see if further discrimination can be achieved through the use of additional methods post-ATR-FTIR analysis.

ATR-FTIR shows great promise for efficient discrimination of visually similar sandy soils and can be used in sequence with other techniques to separate soils that were previously unable to be individualised. This method can easily be incorporated into forensic casework to provide further discrimination between similar samples in an objective manner, with the conclusions reinforced by a scientific measure of similarity that can be presented as evidence in court. It was apparent that sampling methods may have influenced the variability of the soil and should be explored further. This technique could potentially also be developed for use as a predictive screening tool that could provide forensic intelligence throughout the investigative process, allowing detectives to focus on areas of interest. However, this would require more in-depth validation testing and expansion of the sample population.

Whilst the dataset for this study was intended to imitate a forensic casework scenario as closely as possible with limitations on sample size, this study should be expanded to include a more complex range of soils and a greater number of samples analysed per site.

Chapter 5. Analysis of the Quartz-Recovered Fine Fraction of Western Australian Soils using X-Ray Diffraction Paired with Chemometrics

Portions of this chapter have been published in the following article:

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics for the forensic analysis of Western Australian sandy soils." *Forensic Chemistry*, 2022. 28: 100412.

5.1 Introduction

Whilst Chapter 4 demonstrated effective discrimination of soils using attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy, it was challenging to confidently ascertain from the complicated, overlapping loadings which components within the soils were responsible for their discrimination. Additionally, there were several samples that were not well-separated from the rest of the population in first pass principal component analysis (PCA) models, and as such, were not able to have their location predicted correctly using linear discriminant analysis (LDA). The use of another reputable technique in sequence with ATR-FTIR could improve the confidence of the results, confirm the associations made, and potentially even provide additional discrimination of some soils, using complementary information to that already obtained through previous analysis.

X-ray diffraction (XRD) is commonly applied as a non-destructive method that can provide both qualitative and semi-quantitative information on the inorganic material present within soil samples (7, 24, 54, 72). It is also a sensitive technique with small sample size requirements; it can successfully analyse milligram quantities of soil, making it suitable for analysis of forensic trace evidence (7, 24, 72). There are several studies published that make use of XRD for forensic analysis of soils, though most of these were conducted on different soil types to those found in the Swan Coastal Plain (5, 7, 19, 20, 56, 72). More recently, Melo *et al.* investigated the characterisation of the sand fraction of subtropical soils using XRD and chemometrics, however, their method was deemed unsuitable for effective discrimination between these types of dry, sandy soils (54). A method for comparison of the quartz-recovered fine fraction was developed by Pitts and Clarke, which allowed for differentiation of minute and fractionated forensic samples using XRD (94). This method was shown to be suitable for soils such as those found in Perth, where the soil texture is dominated by quartz sand (94). However, it lacked any chemometric interpretation of results, relying solely on discrimination by database comparisons and visual examination of the XRD patterns.

While many studies have demonstrated methods for forensic soil analysis, fewer reports have explored the reproducibility of their technique (57, 73, 126, 127). Reproducibility is considered one of the most important aspects that should be assessed when validating a method for operational use, as outlined in the National Association of Testing Authorities – Technical Note 17 (128). Adhering to validated standards in a forensic setting improves the quality and reliability of the results obtained, and is therefore a crucial component to forensic evidence analysis (66, 129). Validation and verification of a technique is also necessary for resulting evidence to be considered admissible when being presented in court (129). In 2020, Pitts and Clarke demonstrated the reproducibility of the XRD data obtained through use of the quartz-recovered fine fraction by repeated analysis of a single soil sample and replicate analysis of eight sub-samples from the same bulk soil (94). However, their study only made use of a single XRD instrument, as is the case in most literature, so the effects of instrumental variation on method reproducibility were not able to be evaluated. There are many instrumental factors related to XRDs that can impact the results obtained, including instrument sensitivity and stability, optical alignment, and specific programme or software limitations (94). Forensic examiners would therefore encounter these issues when criminal investigations outlast the lifetime of the XRD used for analysis, such as in lengthy cold-cases, or when equipment is upgraded partway through an active investigation. It is also recognised that these diffractometers are expensive capital items, and some forensic laboratories may only have access to older instrumentation. In these instances, would it be viable to use the same method across multiple instruments that have different capabilities, and directly compare this data?

In this chapter, XRD combined with chemometrics was explored for the forensic analysis of the quartz-recovered fine fraction of soils from the Swan Coastal Plain in Perth, Western Australia. The methodology carried out is an advancement on the work of Pitts and Clarke by utilising their developed XRD method on a new set of soil samples, and with the addition of chemometric methods applied to the data. Chemometric analysis can not only allow for objective interpretation of the information collected, but can also increase the efficiency of the interpretation when

conducting analysis of many samples within a large dataset simultaneously. Data was collected using three different XRD instruments and compared to assess any variation attributed to instrument performance, hence providing a measure of the reproducibility of the method. PCA was performed on each XRD dataset to assess whether soils could be distinguished based on their location, or other attributes of the soil (such as dune system, visual appearance, or location type). LDA was then used on the combined datasets to classify and predict replicate samples using the independent test set method, and the accuracy of the model was assessed.

This chapter also expands on published literature even further by applying the above chemometric methods to the full XRD patterns. In published analyses of XRD datasets, generally peak picking or profile fitting occurs first, and the data generated from this is then utilised for multivariate analysis (19, 20, 72, 94). This study however, made use of the whole XRD pattern for chemometric analysis (minus the reflections associated with the low background plates) in the same way that microspectrophotometric (MSP) and ATR-FTIR spectra are used in their entirety. There are several advantages to doing this; use of the whole patterns can allow subtle variation that is undetectable by the human eye to be captured, it decreases the subjectivity surrounding the analysis by removing the need for human interpretation prior to statistical interpretation, and it is both faster and less complicated for the examiner when faced with large numbers of samples.

5.2 Experimental

Soil samples were collected as outlined in Sections 2.2 and 2.3.1. The quartz-recovered fine fractions were prepared from each of the soil samples, as described in Section 2.4. XRD patterns were collected from the quartz-recovered fine fractions as described in Section 2.5.4. PCA and LDA were then conducted on the resulting data as outlined in Section 2.6.

5.3 Results and discussion

5.3.1 Analysis of soils using XRD

XRD patterns were collected from the quartz-recovered fine fractions of 45 Swan Coastal Plain soil samples (outlined in Chapter 2: Table 2.1) using three different XRD instruments; at ChemCentre, half of the soil samples were analysed using a 1992 Philips Analytical PW1820 and the other half were analysed using a 2019 Malvern PANalytical EMPYREAN III with its settings adjusted to match those of the Phillips instrument as closely as possible, and at the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the entire soil collection was analysed again using a Malvern PANalytical EMPYREAN II operating at more modern capabilities (full details can be found in Section 2.5.4). Though the upgrade in ChemCentre equipment midway through analysis was unintentional, this presented a unique opportunity to assess the reproducibility of the method across older and newer instrumentation. The two modern diffractometers were of similar quality and capability, however, because the ChemCentre instrument analysis set-up was intended to replicate the older diffractometer to create a uniform ChemCentre dataset, it was expected that the CSIRO instrument would provide the best quality data. This study was therefore not intended to provide a direct comparison of instrumental capabilities between ChemCentre and CSIRO, but rather to demonstrate the degree of variation produced through analysis using different instrumentation. A selection of XRD patterns obtained from ChemCentre analysis has been illustrated in Figure 5.1, and from CSIRO analysis displayed in Figure 5.2, to demonstrate the variation in minerals observed across the collection of soil samples.

Assignments were based on reference patterns of minerals known to be found within Swan Coastal Plain soils (Table 5.1) (36, 94, 98). Due to the degree of overlap occurring in the reflections of some minerals, the most intense reflection for each mineral was used to identify whether it may be present within the soil samples, and 1 – 2 secondary reflections were then used to confirm its presence. Reflections that showed contribution from multiple different sources, e.g., quartz and microcline at approximately 24°, were generally not assigned to a specific mineral unless the

dominant sources could be confirmed, as in this instance. The presence of six dominant minerals were detected within these XRD patterns – quartz, kaolinite, microcline feldspar, goethite, gibbsite, and calcite. Most of the variation seen between samples was within the intensities or presence/absence of these reflections. For example, Sample 3a (from the Bassendean dunes) contained significantly higher concentrations of gibbsite in comparison to other samples, while Sample 7a (from the Pinjarra Plain) contained higher concentrations of goethite. Sample 15a was dominated by calcite, containing much lower relative concentrations of every other mineral, while Sample 11a was dominated by kaolinite and quartz, and Sample 19a by quartz and microcline feldspar.

Closer visual inspection of the patterns also indicated much smaller contributions likely from chloritised vermiculite (7.2° , 36.4°), mica (10.3°), anatase (29.5°), aragonite (30.5° – left shoulder of quartz reflection, 38.7° , 53.8°), and magnesian calcite (34.7° – right shoulder of calcite reflection) in some samples. Though due to the size of these in comparison to the level of noise and the general lack of secondary reflections present, these were difficult to accurately characterise, and so their presence was only inferred due to the likely occurrence of these minerals in Swan Coastal Plain soils (36, 94, 98), not confirmed. These smaller mineral reflections were easier visualised within the CSIRO patterns, primarily due to the improved signal to noise ratio. These were the only visible differences between ChemCentre and CSIRO patterns; all characteristic reflections of the dominant minerals and their intensities were comparable between the sources of analysis.

Interestingly, samples that were analysed using the older ChemCentre instrumentation (Samples 3a and 7a) exhibited less noise than those analysed using the newer ChemCentre XRD (Samples 11a, 15a, and 19a). This was likely due to the interpolation that occurred when replicating the previous set-up on the newer instrumentation, lowering the quality of the data obtained. This difference was noted for future PCA analysis to assess if any separation was displayed between ‘old’ and ‘new’ batches based on noise. Identifying all possible variations within the XRD patterns by hand would be a labour-intensive process involving careful human

examination and judgement. For this reason, chemometric interpretation is preferred and was applied to efficiently and objectively identify and maximise the differentiation between samples.

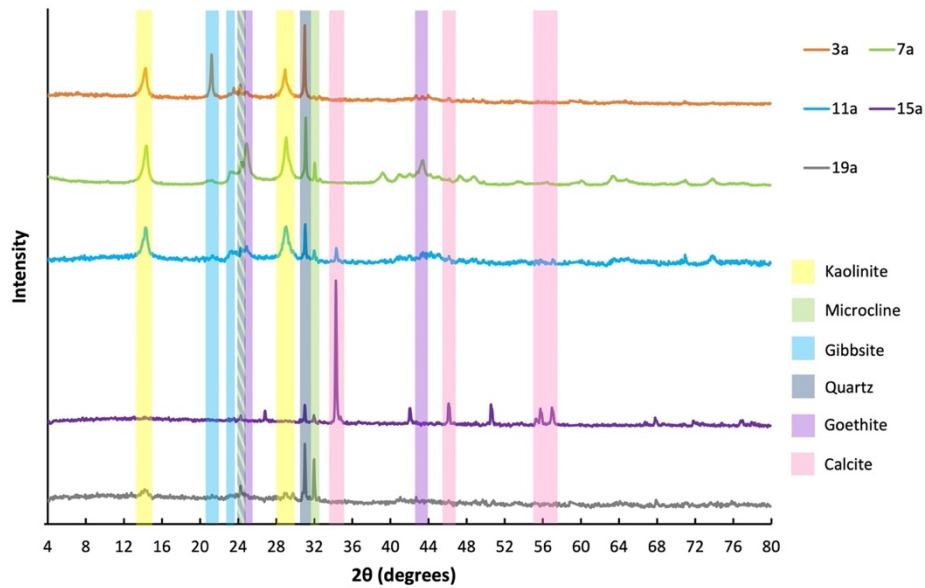


Figure 5.1 Baseline corrected and normalised XRD patterns obtained from ChemCentre, showing the variability in composition of a selection of soil samples collected from differing locations within the Swan Coastal Plain. Annotations were based off assignments outlined in Table 5.1.

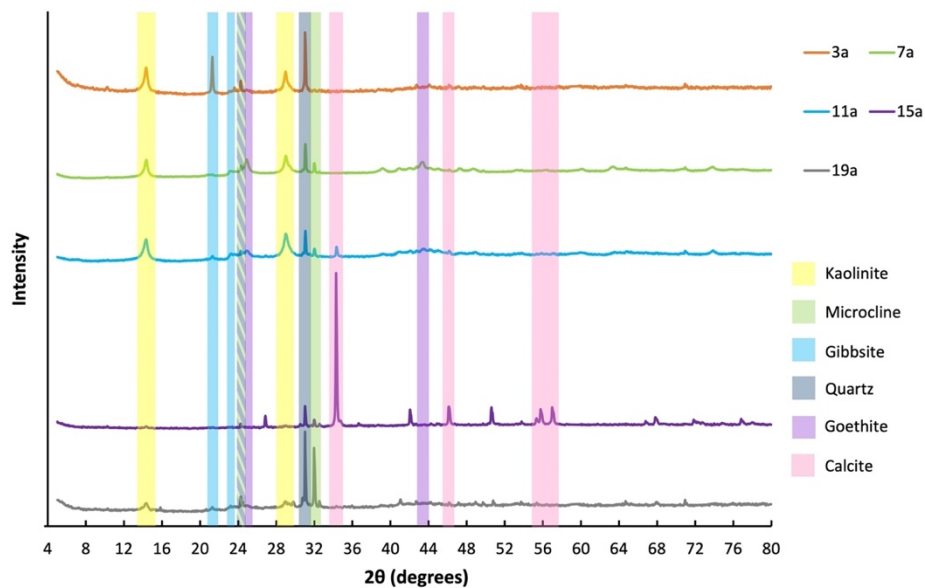


Figure 5.2 Baseline corrected and normalised XRD patterns obtained from CSIRO, showing the variability in composition of a selection of soil samples collected from differing locations within the Swan Coastal Plain. Annotations were based off assignments outlined in Table 5.1.

Table 5.1 Common minerals found in Swan Coastal Plain soils, and their most intense identifying XRD reflections¹ (rounded to one decimal place) that were used as a guide to identify minerals within the XRD patterns.

Co Radiation	2θ (°)	d-Spacing (Å)	ICDD Reference Code
Vermiculite	7.2	14.3	00-060-0341
Clinochlore	7.2	14.2	01-080-1119
Muscovite Mica	10.3	10.0	01-073-9865
Kaolinite	14.3	7.2	00-058-2005
Clinochlore	14.4	7.1	01-080-1119
Bohmite	16.8	6.1	00-021-1307
Gibbsite	21.3	4.8	00-033-0018
Clinochlore	21.7	4.7	01-080-1119
Muscovite Mica	23.2	4.4	01-073-9865
Gibbsite	23.6	4.4	00-033-0018
Quartz	24.3	4.2	00-046-1045
Microcline Feldspar	24.5	4.2	00-019-0932
Goethite	24.8	4.2	04-013-6663
Kaolinite	28.9	3.6	00-058-2005
Anatase	29.5	3.5	01-085-5943
Aragonite	30.5	3.4	00-041-1475
Quartz	31.0	3.3	00-046-1045
Microcline Feldspar	31.5	3.3	00-019-0932
Microcline Feldspar	32.0	3.2	00-019-0932
Bohmite	32.8	3.2	00-021-1307
Calcite	34.3	3.0	00-066-0867
Magnesian-Calcite	34.7	3.0	01-086-2336
Vermiculite	36.4	2.9	00-060-0341
Aragonite	38.7	2.7	00-041-1475
Muscovite Mica	41.0	2.6	01-073-9865
Goethite	43.2	2.4	04-013-6663
Calcite	46.1	2.3	00-066-0867
Aragonite	53.8	2.0	00-041-1475
Calcite	55.8	1.9	00-066-0867
Calcite	57.0	1.9	00-066-0867

¹ICDD (2022), PDF-4 Minerals 2022, version number: 4.2211.

5.3.2 Principal component analysis

In a similar approach to that carried out in Chapters 3 and 4, PCA was applied to the XRD data to assess whether soils could be differentiated based on their location, or any other attributes of the soil. In this chapter, it was also utilised as a tool to detect and explore any variation evident between patterns collected using different XRD instrumentation.

5.3.2.1 Discrimination using location-based ChemCentre data

PCA performed on the ChemCentre XRD patterns revealed that 85.1% of the total variance in the dataset could be described by the first five principal components (PCs), as illustrated in the scree plot (Figure 5.3). However, upon visualisation of the influence of PC-5 on the scores plots, it did not improve the discrimination of soils, and hence the first four PCs (accounting for 78.6% of the variance) were utilised for the PCA model.

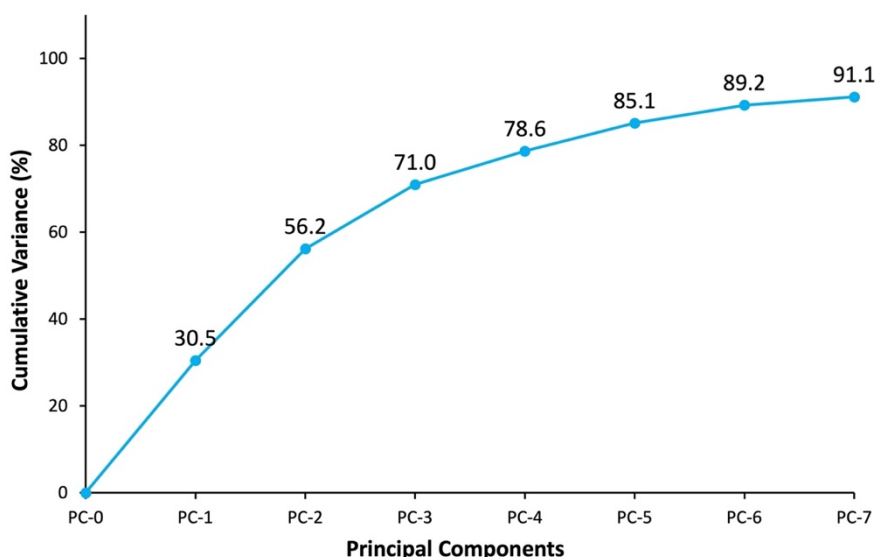


Figure 5.3 Scree plot depicting the cumulative variance in the ChemCentre XRD dataset retained by each PC.

Three-dimensional score plots generated using these PCs (Figure 5.4) resulted in most of the soils clustering based on the locations from which they were collected, however, samples from Locations 19 and 21 were shown to overlap with Locations 2 and 12 along PCs 1-3. Locations 1, 2, 6, 9, and 11 all showed significant intra-site

variability, with samples from the same location separating within the model. Three of these sites were subject to a high degree of human interference, being both easily accessible and highly managed e.g., a garden bed within a public park or roadside verge. It is therefore logical to draw the conclusion that human interference could have contributed to the variation shown across a number of these sites. The other two sites were dense bushland that produced grey sands, indicating higher inherent variability in these soils; previous chapters have demonstrated the same trends in grey bushland sands.

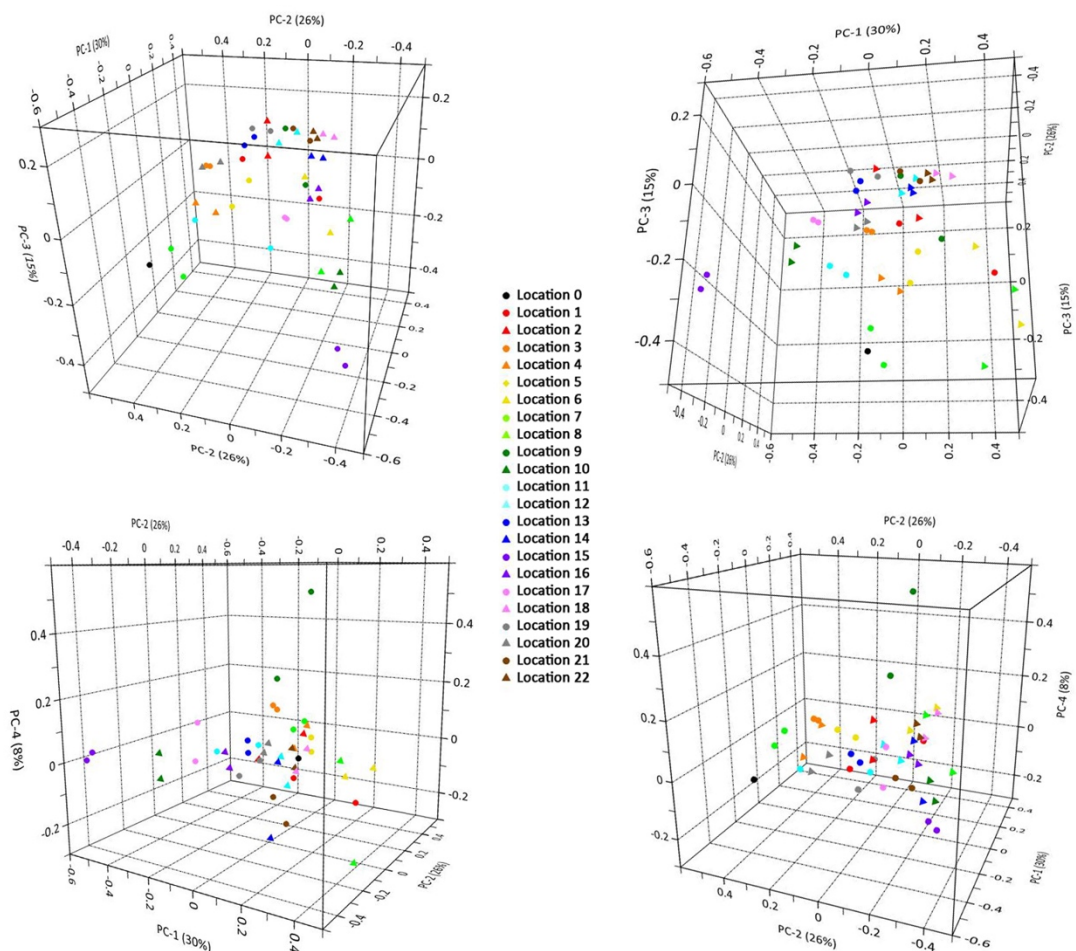


Figure 5.4 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations based on their corresponding ChemCentre XRD patterns.

In contrast to the ATR-FTIR model produced in Section 4.3.3, XRD achieved improved discrimination of the samples from Locations 4, 7, 13, 14, and 20. Modelling based on ATR-FTIR spectra had these clusters overlapping in initial trials, whereas using the

XRD data allowed for enhanced separation of nearly all these groups, with all intra-site samples still clustered closely. The XRD model also managed to reduce the intra-location separation seen in the ATR-FTIR model between samples originating from Locations 8 and 18. Chemometric analysis of the full XRD pattern shows strong promise for discriminating soils that cannot be entirely individualised through PCA of the ATR-FTIR data alone. The inclusion of PC-4 when visualising the scores plot allowed for the discrimination of Locations 6, 9, and 21 (Figure 5.4), leaving only Locations 1, 2, and 19 unable to be individualised within the XRD model. As Locations 1 and 19 were difficult to differentiate across both ATR-FTIR and XRD models, the variability within these bushland soils was emphasised. Despite these challenges, all three of these locations were able to be differentiated by PCA using ATR-FTIR data, highlighting the importance of using several complementary techniques within a sequence to achieve maximum differentiation between samples.

The factor loadings for the first four PCs were studied to determine the reflections in the XRD patterns that were associated with the discrimination of samples along each component (Figure 5.5). There was some uncertainty stemming from identification based on the XRD pattern alone as some reflections were difficult to resolve, for example kaolinite and halloysite, however, the presence of halloysite is questionable in sandy soils considering it is typically derived from igneous rock (130-132). Hence, using the full pattern with PCA to capture all the information present was valuable, so that complete identification was not essential. It was also noted that mica, aragonite, and magnesian calcite were detected within the loadings despite being difficult to visualise within the XRD patterns, illustrating the advantage that chemometric interpretation has over human interpretation, as it can easily identify subtle variation in minerals that visual examination cannot. In this instance, these minerals did not significantly contribute to the separation between samples, but may be of assistance as the dataset grows to include more Quindalup soils.

The loadings plot for PC-1 exhibited an overall positive correlation with microcline feldspar and quartz, and a negative correlation with calcite. Samples from location 15 were best separated along PC-1 in a negative direction, and therefore were assumed to have higher concentrations of calcite and lower concentrations of

microcline feldspar and quartz in their soil compared to other samples. This was confirmed by examining the original, pre-processed XRD patterns. Location 15 soils originated from the Quindalup dune system, nearest to the coast (as illustrated in Chapter 1: Figure 1.2), which is known to contain higher concentrations of calcite and lower relative quartz levels, due to the presence of marine-life and limestone (36, 94).

The loadings plot for PC-2 appeared to be positively influenced by kaolinite, with possible smaller contribution from goethite and gibbsite, and negatively influenced by calcite, microcline feldspar, and quartz. Location 15 soils were once again the most negative along PC-2, primarily due to their significantly higher concentrations of calcite relative to other samples. Samples from Locations 0 and 7 were positioned at the positive end of PC-2, as they contained higher amounts of kaolinite, goethite, and/or gibbsite. These soils all originated from the Pinjarra Plain which represents an accumulation of fluvial deposits, and hence these soils can contain relatively high concentrations of these minerals (43, 94).

The loadings plot for PC-3 was positively correlated with quartz and microcline feldspar, and negatively correlated with calcite, microcline feldspar, quartz, and kaolinite. Quartz and microcline feldspar were correlated with both positive and negative loadings across PC-3, so the original XRD patterns were assessed to see whether z-offset corrections were needed between the two different instruments used; this was not required, so the split reflections were presumed to be due to a potential normalisation issue, and as such, their influence on the loadings data was negated. Due to this, the separation achieved between samples along PC-3 was not significant, with samples from locations 8, 7, and 15 positioned most negatively presumably due to their higher relative concentrations of calcite and/or kaolinite.

The loadings plot for PC-4 exhibited positive association with quartz, and gibbsite to a lesser degree, and negative association with microcline feldspar and kaolinite. Location 9 was best separated along PC-4 in a positive direction, followed by Location 3, signifying higher relative amounts of quartz and/or gibbsite within these soils. These samples were the only ones that originated from the Bassendean dune system,

which has previously been shown to be dominated by quartz and gibbsite in comparison to the more westerly systems (36, 94).

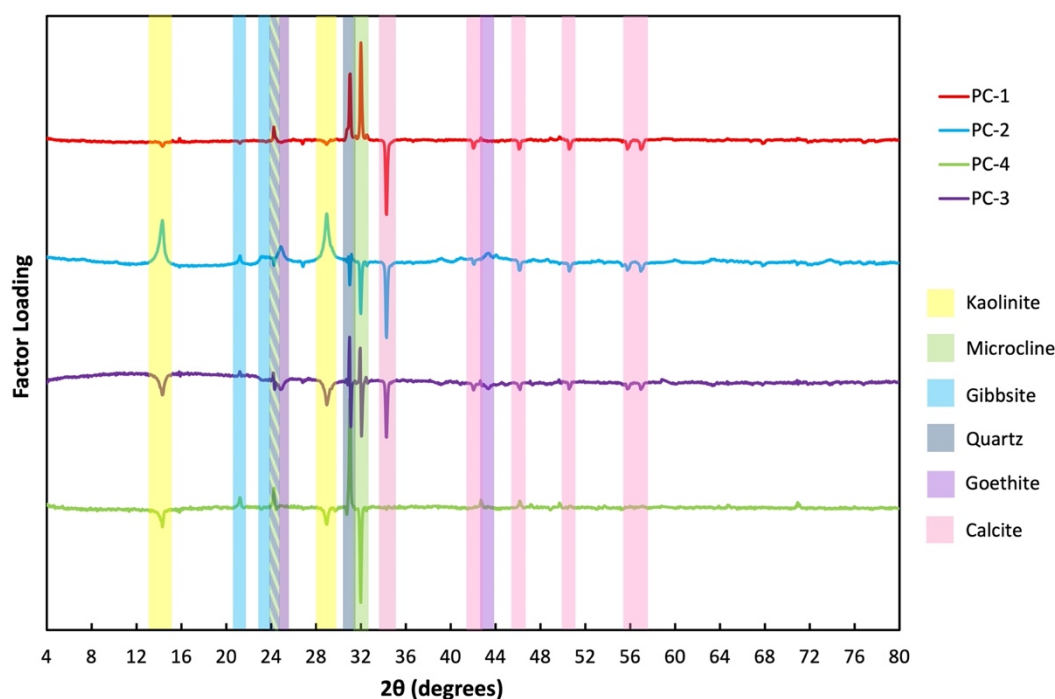


Figure 5.5 Factor loadings plot of PCs 1-4 for PCA of the ChemCentre XRD dataset, with the main reflections of interest highlighted and annotated with their contributing compounds.

The inclusion of samples from additional non-metropolitan sites could help to increase separation. For example, aluminium substitution for iron within goethite will shift the positioning of goethite reflections, potentially allowing for greater differentiation between soils containing variably substituted goethite (50, 94, 133). This could also assist with determining the sample's original location, as Swan Coastal Plain soils generally experience more substitution by aluminium compared with the iron-rich soils located further inland (98, 134). The same phenomenon could also be applied to other minerals such as kaolinite and vermiculite (chloritized or not), where numerous substitutions are possible. Hence using the full XRD pattern, which captures any shifts away from 'standard' reflection positions, is recommended.

5.3.2.2 Discrimination using feature-based ChemCentre data

As XRD specifically identifies the mineralogy of the soil samples, the PCA model was re-visualised (Figure 5.6) to present the samples grouped according to the dune

system within the Swan Coastal Plain from which they originated (outlined in Chapter 2: Table 2.1). Visual discrimination of the four dune systems was achieved, despite the spread of the samples within each cluster, confirming that the variation in mineral composition detectable by XRD was associated with the soils' original dune system. Soils from the Pinjarra Plain, located further inland than the other systems, were shown to contain higher relative levels of kaolinite in comparison to other samples. Bassendean soils were indicated to contain higher relative concentrations of quartz, as demonstrated in previous studies (94). Samples collected from the Spearwood dunes were also consistent with this previous study, containing higher levels of microcline feldspar and mica than their easterly counterparts (94). Quindalup soils contained higher levels of calcite, due to their westerly coastal positioning (36, 94). These trends allowed the dune systems to cluster and be discriminated from one another within the scores plot.

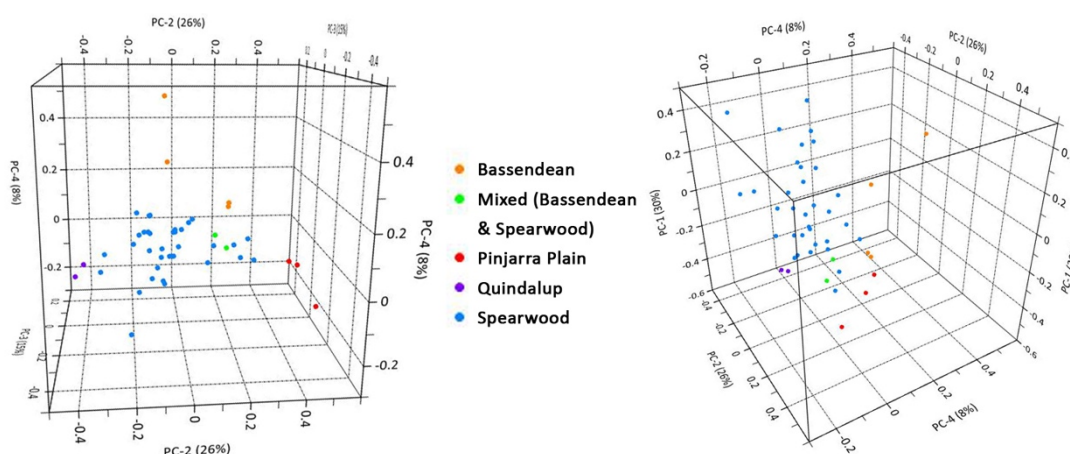


Figure 5.6 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing dune systems based on their corresponding ChemCentre XRD patterns.

When samples were grouped according to the ChemCentre instrumentation that was used for their analysis, the model did exhibit broad clustering based on these two 'old' and 'new' batches (Figure 5.7). This may be coincidental, as most of the variation across the loadings was clearly due to differences in mineral content, however, it is also possible that the variation in signal-to-noise ratio was unintentionally encapsulated within the loadings, and was therefore influencing the model.

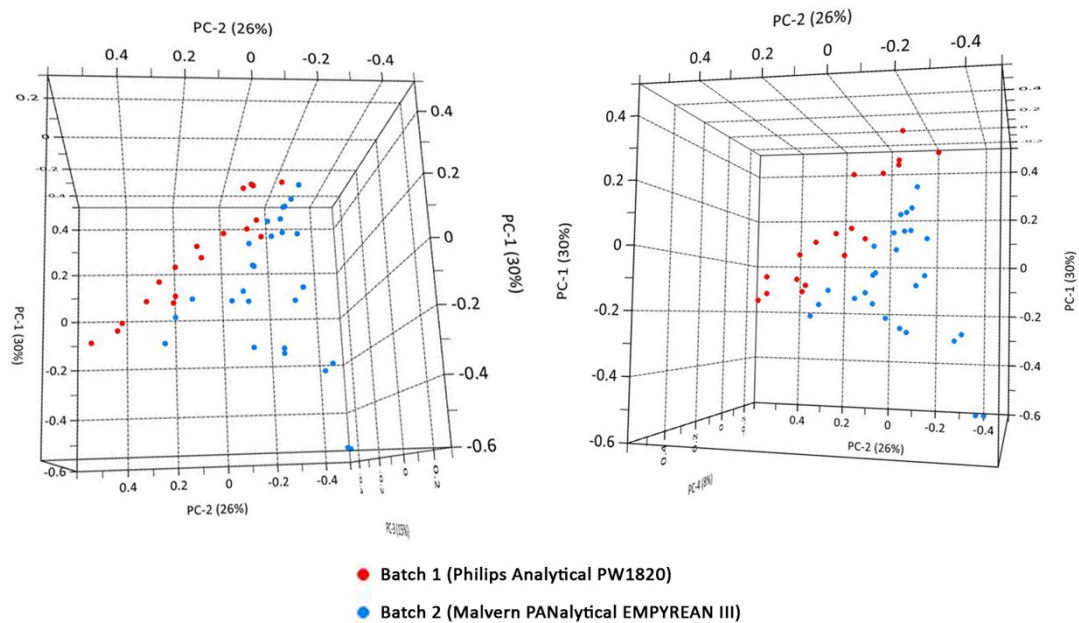


Figure 5.7 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples analysed with different XRD instrumentation, based on their corresponding ChemCentre XRD patterns.

Improved separation was again investigated by grouping the samples based on other attributes of the soil, such as their visual appearance (outlined in Chapter 2: Table 2.1). This resulted in poorly formed clusters that significantly overlapped with other classes (Figure 5.8). The only group that was able to be discriminated from the rest of the population was the red-coloured soils, however, these were still situated in close proximity to other samples. The presence of some distinctively coloured minerals such as goethite and kaolinite (often stained red brown from iron oxides (125)) are evidently shaping the model, however, there are other compounds influencing the PCs that are likely not associated with the overall colour of the soils. The use of visual colour to group samples within the model did not give increased discrimination over grouping based on locations. It is likely that while the appearance of the soil is linked to the mineralogy, the minerals responsible for colour are present at low levels and hence not significant in the XRD patterns seen. This is especially true for the iron oxides, with a hint of iron making soils coloured. It is expected that there would also be small amounts of organic material present within samples which, even if small, would contribute to the colour of the soils without being identified by

XRD. No additional structure in the dataset was visualised when grouping the soils according to the type of location that they were collected from (Figure 5.9).

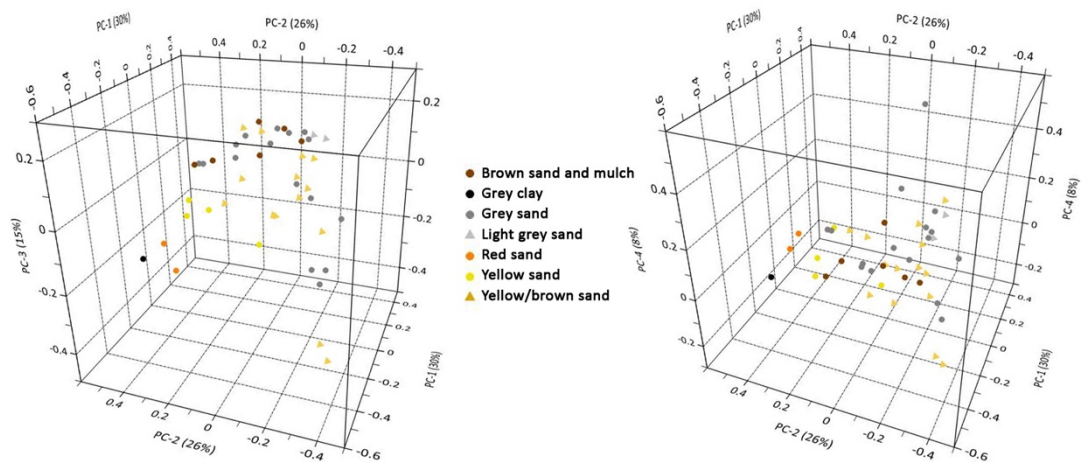


Figure 5.8 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability in the visual appearance of soil samples based on their corresponding ChemCentre XRD patterns.

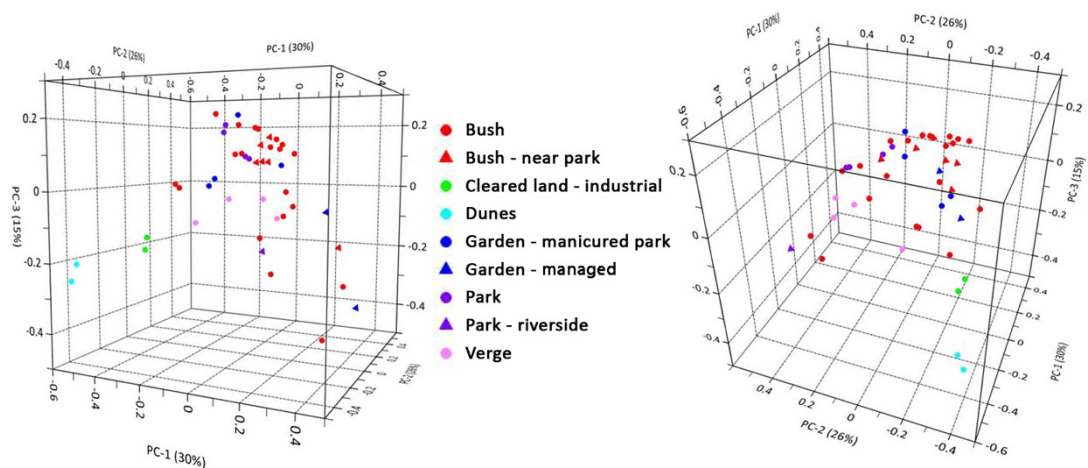


Figure 5.9 3-dimensional PCA scores plot (shown from two perspectives) displaying the variability of soil samples from differing types of locations based on their corresponding ChemCentre XRD patterns.

5.3.2.3 Discrimination using CSIRO data

PCA performed on the CSIRO XRD patterns revealed that 85.2% of the total variance in the dataset could be described by the first four PCs, as illustrated in the scree plot (Figure 5.10). Three-dimensional score plots generated using the first three PCs (Figure 5.11) resulted in a similar degree of clustering compared to the ChemCentre

model. Samples from Locations 19 and 21 showed improved discrimination, no longer overlapping with other clusters across PCs 1-3. Inspection of both the ChemCentre and CSIRO XRD patterns for Location 19 (Figure 5.12) revealed the key difference between them to be the signal-to-noise ratio, which was enhanced in the modern CSIRO patterns and hence allowed for improved detection of some additional subtle mineral reflections, such as gibbsite at 21.3° .

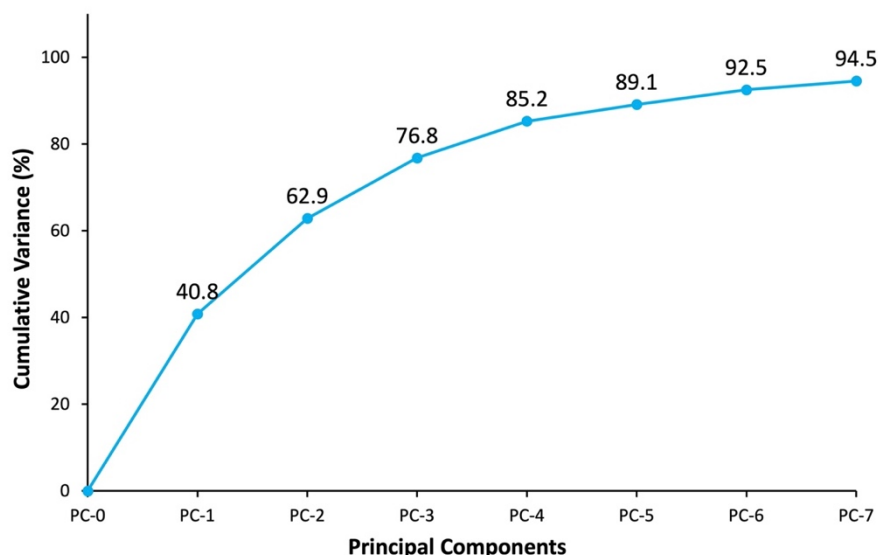


Figure 5.10 Scree plot depicting the cumulative variance in the CSIRO XRD dataset retained by each PC.

However, in addition to Locations 1, 2, 6, 9, and 11, Location 4 also showed significant intra-site variability between its samples. Further examination of these XRD patterns (Figure 5.12) revealed that the CSIRO pattern for Sample 4b was missing the basal kaolinite reflection at 14.3° . While this reflection has been shown to be affected by the thermal collapse of kaolinite, this generally occurs after heating, and these samples were stored and analysed at room temperature (135). The absent kaolinite reflection also appears to be associated with lower reflection intensities across the entire XRD pattern. It is therefore more likely to be due to an alteration effect of the sample, such as drying or ageing, or the other sample deposits being too thin and increasing preferred orientation of the secondary clay minerals, causing enhancement of basal peak intensities within the 4a and ChemCentre 4b samples

(94, 136). However, this could not be confirmed without further repeated analysis of the sample.

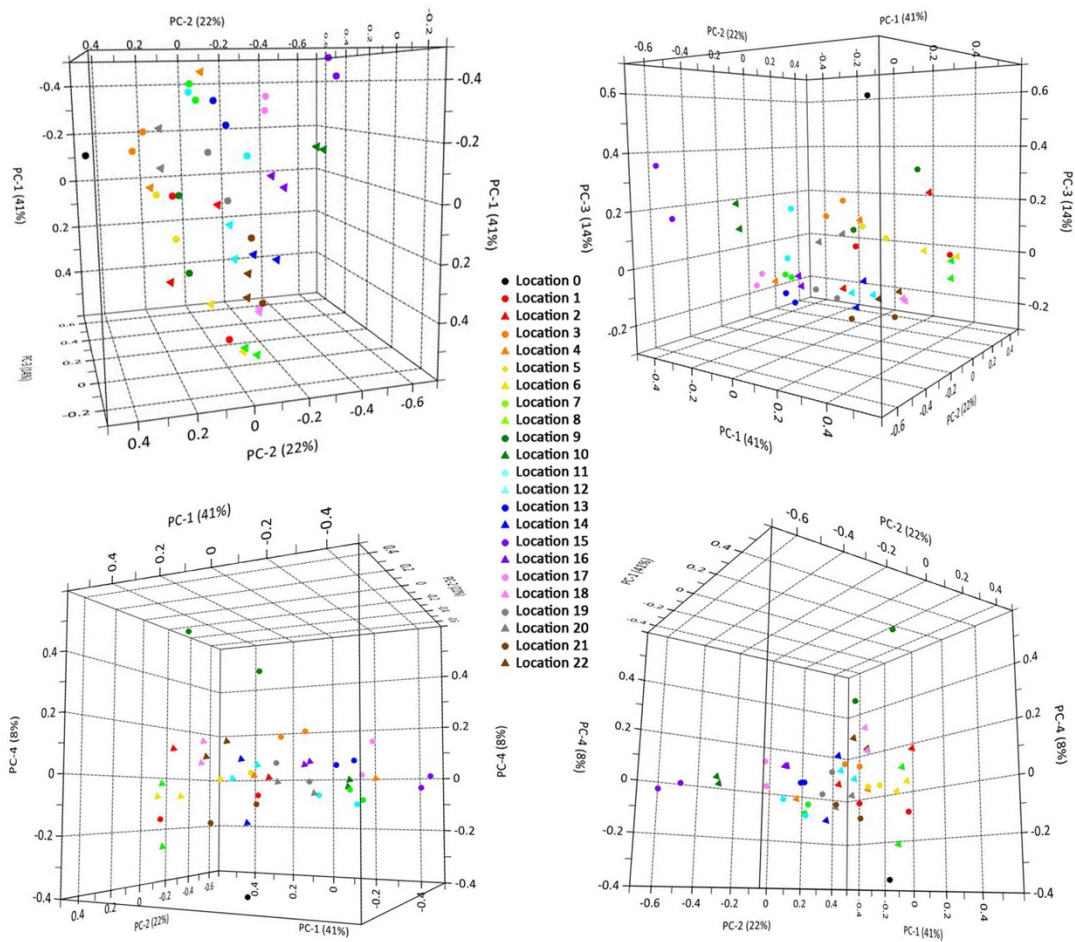


Figure 5.11 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations based on their corresponding CSIRO XRD patterns.

The inclusion of PC-4 when visualising the scores plot allowed for the further discrimination of Locations 6 and 9 (Figure 5.11), this time leaving Locations 1, 2, 4, and 11 unable to be individualised within the model. So while the ChemCentre model was able to discriminate Location 4 and 11 soils and the CSIRO model was not, the CSIRO model was able to discriminate Location 19 soils unlike the ChemCentre model. The variation seen across samples from Locations 4 and 19 was tentatively attributed to alteration / orientation effects and signal-to-noise ratio, as outlined above, however, the differences in Location 11's patterns (Figure 5.12) appeared to be due to sampling variation between duplicate samples, or inherent variation within

the soil from this site. Overall, the only soils that were unable to be discriminated within any XRD models were those from Locations 1 and 2. As previously stated, these locations were both discriminated within the ATR-FTIR models presented in Chapter 4, reinforcing the enhanced capabilities of the method when performed within an analysis sequence, rather than in isolation.

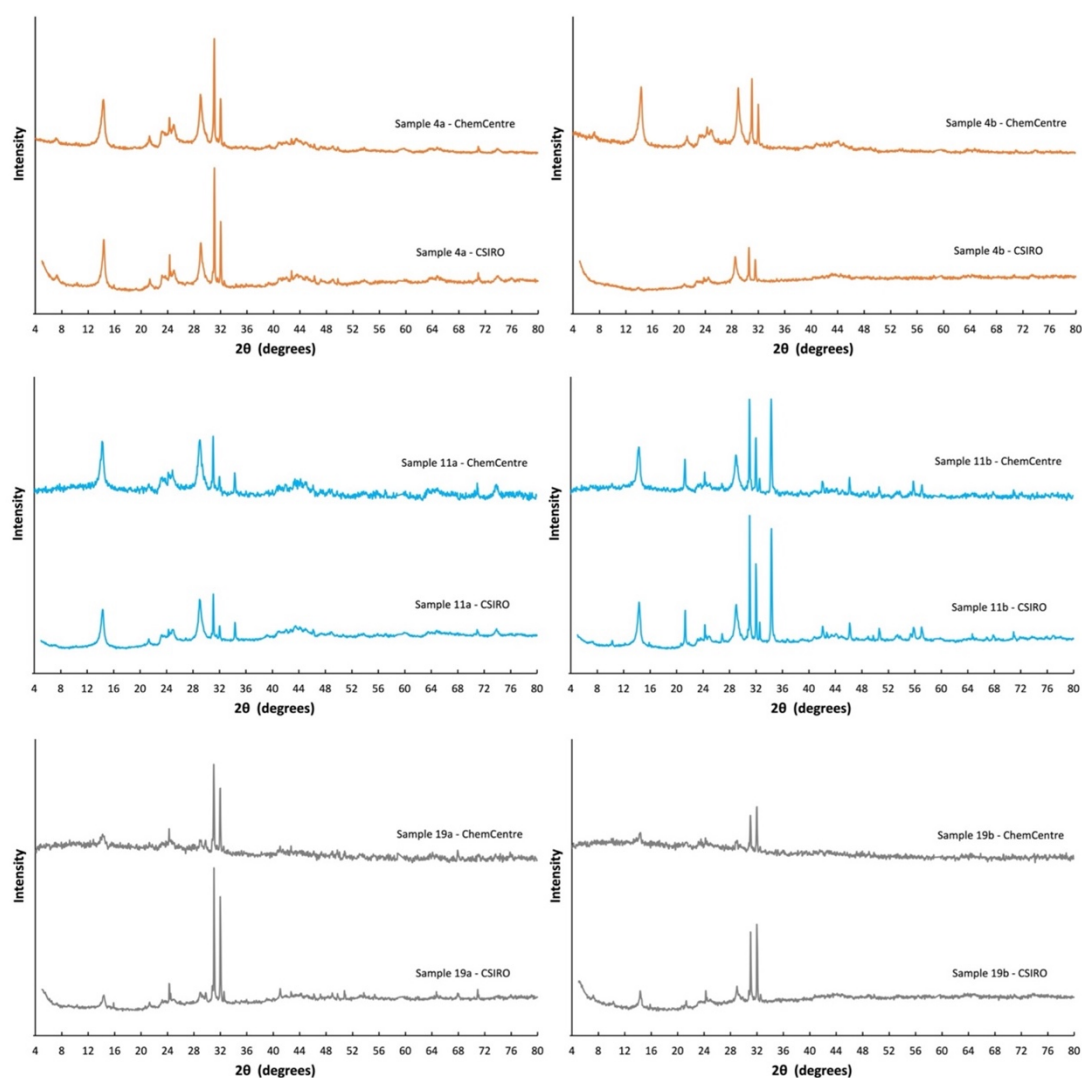


Figure 5.12 XRD patterns obtained from soil samples from Locations 4, 11, and 19, illustrating both the variation in patterns between different instrumentation, and between duplicate samples.

The loadings plot for the CSIRO model was largely the same as that from the ChemCentre model, with a few variations in reflection intensities (Figure 5.13); PC-1 showed a slight reduction in the degree of negative influence attributed to calcite, PC-2 showed less contribution from goethite and a reduction in influence attributed

to kaolinite, quartz, and microcline, and PC-4 showed an increase in influence attributed to kaolinite with a simultaneous decrease in influence from microcline. PC-3 was flipped in comparison to the ChemCentre loadings, and quartz and microcline no longer displayed any normalisation errors, leaving positive contributions from calcite, kaolinite, and quartz, and negative contributions from microcline feldspar.

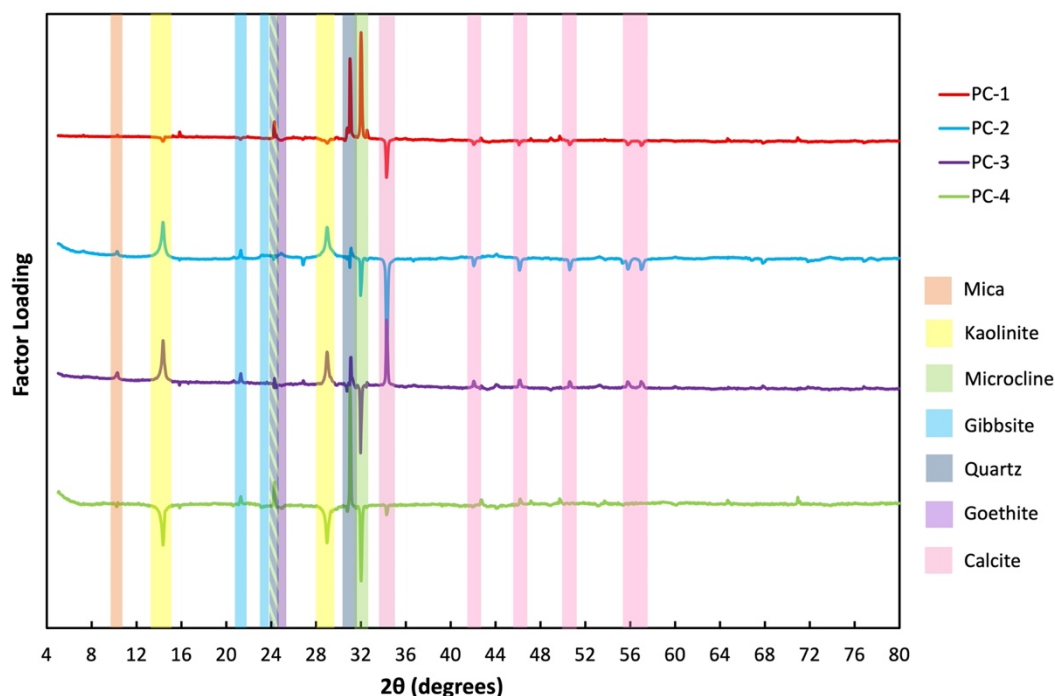


Figure 5.13 Factor loadings plot of PCs 1-4 for PCA of the CSIRO XRD dataset, with the main reflections of interest highlighted and annotated with their contributing compounds.

The CSIRO loadings also detected variation in some of the minerals seen in lower concentrations within the soil samples due to its increased sensitivity, e.g., mica, which were not encapsulated by the ChemCentre loadings. However, the influence attributed to these minerals was again not significant. The variations seen between the loadings plots were substantial enough to alter the discrimination of samples along PC-3, but most samples were situated in comparable positions on PCs 1, 2, and 4 relative to the ChemCentre model. For example, Location 15 was still the most separated along PCs 1 and 2, with Location 0 located on the extreme opposite side of PC-2, and Location 9 was again the most separated along PC-4. This highlights the

reproducibility of the method, despite utilising multiple XRD instruments with differing capabilities.

5.3.3 Linear discriminant analysis

LDA was used as a classification technique to predict the original locations of soil samples based on their chemical data. The XRD data from both ChemCentre and CSIRO were combined to produce one dataset for use with LDA. This was done to ensure an appropriate number of samples within each location class, and hence increase the reliability of the model. To do this, CSIRO XRD patterns were adjusted for their step size post-analysis to match the 0.05° step size of the ChemCentre XRD patterns, and ChemCentre XRD patterns were truncated from $5\text{--}80^\circ$ 2θ to match the range of the CSIRO patterns (details of how this was done can be found in Section 2.6.2).

The full dataset was then split into two for PCA-LDA; three out of the four replicates per location were used to build the calibration model, and the remaining one replicate was used as a validation set and predicted onto the model to assess its performance. The validation set consisted of replicates that were not utilised within the calibration set, to ensure no overlap between models and a more realistic estimation of the model's performance. LDA was conducted twice; firstly, using all of the ChemCentre patterns and one replicate from each of CSIRO's samples as the calibration set, and the remaining CSIRO patterns as the validation set, and then swapping the data so that all of the CSIRO patterns and one replicate from each of ChemCentre's samples were used as the calibration set, and the remaining ChemCentre patterns were the validation set. Location 0 was excluded from LDA, due to its limited sample size (only two replicates total). While up to four PCs allowed maximum separation between locations through PCA, it was only possible to utilise the first two PCs for LDA, due to the available number of samples per location.

5.3.3.1 Predicting CSIRO's validation samples onto the XRD LDA model

LDA was performed on the combined XRD dataset post-PCA, with each location treated as an individual class. This discriminant model returned a calibration accuracy

of 65.1% (Table 5.2). The majority of the locations misclassified were those that either were unable to be separated within the PCA scores plots, or only discriminated via PCs 3 or above. The combination of the ChemCentre and CSIRO datasets meant that the LDA model had increased uncertainty compared to the PCA models, as it not only encapsulated variation based on sample mineralogy and sample preparation, but also variation due to instrumental analysis.

Table 5.2 Number of correct vs incorrect location classifications for samples in the first combined XRD calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Classified	% Correct
1	0	3	2, 5, 6	0
2	1	2	1 & 19	33
3	2	1	20	67
4	3	0	-	100
5	3	0	-	100
6	3	0	-	100
7	2	1	20	67
8	2	1	18	67
9	1	2	2 & 12	33
10	3	0	-	100
11	1	2	17 & 20	33
12	1	2	1 & 2	33
13	2	1	11	67
14	3	0	-	100
15	3	0	-	100
16	3	0	-	100
17	3	0	-	100
18	2	1	22	67
19	1	2	2 & 13	33
20	0	3	3 & 7	0
21	2	1	9	67
22	2	1	18	67
% Total Correct				65

This LDA model was then used to predict the locations of 22 CSIRO replicates from the validation dataset. Eight samples, or 36.4%, had their location correctly predicted (Table 5.3). Despite Locations 8, 13, and 22 being misclassified within the calibration set, they all achieved 100% validation accuracy, though the confidence surrounding several of these predictions was low; the associated discriminant values (Table 5.4), indicating the distance between the sample and the centroid of each class, showed that the majority of samples were close to being predicted as originating from at least one other location. Most of the samples that exhibited 100% calibration accuracy also achieved the same validation accuracy, except for Locations 4, 5, and 6, due in part to the significant intra-site variation. As expected, all of the samples that were unable to be discriminated within the scores plots of either XRD model (Locations 1, 2, 4, 11, and 19) had their locations predicted incorrectly. The level of separation between classes, and hence the resulting discriminant values, could have potentially been improved by incorporating further PCs into the analysis, or utilising a larger number of replicates within each class so that separate LDA models could be created for ChemCentre versus CSIRO patterns.

Table 5.3 Number of correct vs incorrect location predictions for samples in the CSIRO validation set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Predicted	% Correct
1	0	1	6	0
2	0	1	1	0
3	0	1	7	0
4	0	1	11	0
5	0	1	4	0
6	0	1	8	0
7	0	1	20	0
8	1	0	-	100
9	0	1	22	0
10	1	0	-	100
11	0	1	17	0
12	0	1	9	0
13	1	0	-	100
14	1	0	-	100
15	1	0	-	100
16	1	0	-	100
17	1	0	-	100
18	0	1	8	0
19	0	1	13	0
20	0	1	4	0
21	0	1	18	0
22	1	0	-	100
% Total Correct				36

Table 5.4 Discriminant values of replicates from the CSIRO XRD validation dataset (rounded to three decimal places), with correct predictions shaded green and incorrect predictions shaded red. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

Sample Replicate	Discriminant Values																						Predicted Location	Closest Value Within (%)
	Location 1	Location 2	Location 3	Location 4	Location 5	Location 6	Location 7	Location 8	Location 9	Location 10	Location 11	Location 12	Location 13	Location 14	Location 15	Location 16	Location 17	Location 18	Location 19	Location 20	Location 21	Location 22		
1b	-12.525	-19.617	-38.846	-29.549	-17.495	-3.934	-44.298	-4.452	-17.779	-128.864	-51.071	-14.050	-40.315	-17.352	-249.575	-44.528	-71.732	-7.614	-28.453	-39.857	-14.620	-9.296	6	13
2b	-4.618	-8.853	-19.126	-12.242	-6.008	-4.783	-21.875	-10.072	-8.704	-122.304	-31.740	-6.712	-22.873	-14.023	-239.365	-36.802	-57.553	-6.768	-15.048	-19.877	-8.118	-6.232	1	4
3b	-12.485	-8.800	-3.667	-5.211	-9.163	-30.455	-3.567	-40.876	-10.850	-100.888	-9.613	-13.059	-6.676	-22.449	-197.573	-29.611	-35.057	-24.548	-7.556	-3.600	-14.479	-20.167	7	1
4b	-26.032	-16.935	-10.275	-19.192	-24.397	-49.924	-12.896	-56.183	-18.157	-63.327	-4.510	-22.582	-6.215	-24.866	-136.517	-18.267	-15.827	-35.651	-10.591	-9.606	-21.890	-30.806	11	38
5b	-6.526	-6.400	-6.353	-3.820	-4.386	-17.828	-6.349	-27.452	-7.902	-114.046	-16.524	-8.418	-10.870	-18.523	-221.637	-33.159	-44.865	-16.007	-8.416	-6.613	-10.194	-12.947	4	15
6b	-16.675	-24.227	-46.383	-36.964	-23.032	-5.645	-53.037	-4.128	-21.732	-127.660	-57.361	-17.616	-46.400	-18.865	-247.376	-46.236	-74.902	-9.106	-33.484	-47.429	-17.701	-11.503	8	37
7b	-21.720	-14.740	-6.478	-12.159	-18.297	-44.886	-6.643	-54.586	-16.816	-86.181	-6.790	-20.601	-6.609	-27.649	-170.678	-26.998	-26.913	-34.538	-10.300	-6.001	-21.177	-29.384	20	10
8b	-16.657	-23.702	-46.059	-37.382	-23.343	-5.928	-53.200	-3.742	-21.050	-121.765	-55.723	-17.101	-45.203	-17.381	-238.967	-43.388	-71.436	-8.509	-32.502	-47.058	-16.924	-10.937	8	58
9b	-4.795	-7.103	-19.459	-15.614	-8.124	-5.161	-24.866	-6.812	-6.032	-95.778	-26.149	-4.551	-19.010	-6.974	-201.980	-24.205	-43.031	-3.398	-11.603	-20.055	-4.686	-3.279	22	4
10b	-111.171	-93.816	-104.814	-128.304	-122.575	-129.838	-123.472	-113.089	-88.837	-4.005	-66.823	-94.788	-76.475	-67.891	-21.448	-34.590	-26.438	-96.644	-80.228	-103.460	-86.622	-95.670	10	436
11b	-40.140	-28.889	-32.589	-46.247	-44.891	-58.538	-42.907	-54.025	-27.088	-20.787	-14.382	-31.351	-18.380	-20.784	-73.797	-5.651	-3.580	-37.588	-20.848	-31.829	-27.516	-35.056	17	58
12b	-4.648	-4.578	-10.827	-9.376	-5.965	-11.578	-13.884	-15.848	-4.432	-90.678	-15.948	-4.516	-10.710	-8.672	-191.669	-21.158	-34.668	-7.713	-6.566	-11.027	-4.778	-6.195	9	2
13b	-15.349	-9.382	-8.016	-13.513	-15.229	-32.835	-11.052	-37.668	-9.893	-66.440	-5.213	-12.796	-4.612	-14.890	-148.254	-14.505	-17.763	-21.852	-5.864	-7.610	-12.192	-18.381	13	13
14b	-7.549	-6.908	-18.067	-18.227	-11.685	-11.474	-24.465	-11.092	-5.431	-71.413	-18.356	-5.265	-13.814	-4.039	-164.922	-13.663	-27.963	-5.181	-8.512	-18.271	-4.241	-4.768	14	5
15b	-179.852	-154.624	-158.285	-191.524	-190.822	-210.594	-178.595	-192.235	-149.821	-16.608	-107.674	-159.275	-123.642	-127.025	-5.626	-75.726	-54.780	-167.023	-133.759	-156.155	-149.349	-164.092	15	195
16b	-27.659	-19.575	-26.133	-35.926	-32.653	-41.373	-35.330	-37.483	-17.662	-30.468	-12.658	-20.633	-14.233	-12.141	-94.637	-3.964	-6.845	-24.503	-14.303	-25.598	-17.400	-22.799	16	73
17b	-38.432	-26.917	-26.202	-39.560	-41.094	-60.309	-33.718	-59.120	-26.090	-30.401	-10.180	-30.960	-14.323	-23.884	-86.262	-9.046	-5.072	-40.179	-18.373	-25.286	-27.834	-36.685	17	78
18b	-9.099	-13.228	-30.132	-24.634	-14.231	-4.829	-36.124	-4.293	-11.295	-102.303	-36.951	-8.934	-28.692	-9.687	-211.577	-29.706	-52.096	-4.368	-19.208	-30.783	-8.645	-5.465	8	2
19b	-10.720	-6.424	-5.968	-9.253	-10.295	-26.051	-8.475	-31.814	-7.089	-76.211	-6.515	-9.254	-4.485	-13.158	-164.885	-16.933	-22.786	-17.530	-4.511	-5.764	-9.224	-14.296	13	1
20b	-9.613	-7.485	-4.321	-4.045	-6.684	-24.858	-4.076	-35.135	-9.353	-107.166	-12.263	-10.854	-8.136	-20.777	-208.844	-31.254	-39.343	-20.807	-7.646	-4.386	-12.473	-16.950	4	1
21b	-8.034	-11.612	-27.458	-22.545	-12.837	-4.984	-33.289	-4.800	-9.832	-99.196	-33.830	-7.754	-25.976	-8.649	-207.062	-27.598	-48.938	-3.978	-17.084	-28.059	-7.471	-4.807	18	21
22b	-5.398	-7.367	-19.315	-15.987	-8.842	-6.514	-24.114	-8.013	-6.224	-93.168	-24.887	-5.071	-18.203	-7.094	-197.661	-23.090	-41.079	-4.141	-11.309	-19.741	-4.955	-4.013	22	3

5.3.3.2 Predicting ChemCentre's validation samples onto the XRD LDA model

LDA was again performed on the combined XRD dataset post-PCA, with each location treated as an individual class. This discriminant model returned a calibration accuracy of 59.1% (Table 5.5). While Locations 8, 13, and 19 all achieved 100% classification accuracy, they did not exhibit the same classification accuracy in the previous LDA model, indicating that these samples were better discriminated from the rest of the population by the CSIRO data; the previous model utilised more ChemCentre data within the calibration set, whereas this model utilised more CSIRO data. Locations 8 and 13 were both also predicted correctly within the previous CSIRO validation set, reinforcing this. Comparing their XRD patterns in Figure 5.14, no observable differences could be identified from the use of the two instruments other than the improved sensitivity and signal-to-noise ratio obtained by the modern CSIRO XRD. In contrast, Locations 4 and 6 both achieved 100% classification accuracy in the previous model, but did not exhibit the same results within this model, indicating that these locations were better discriminated by the ChemCentre data. For Location 4, this was due to the absent basal kaolinite reflection in the CSIRO pattern, as outlined in Section 5.3.2.3. However, Location 6 showed greater disparity between reflection intensities of quartz and microcline feldspar in the CSIRO patterns compared to the ChemCentre patterns (Figure 5.14). Once again, the majority of the locations misclassified were those that either were unable to be separated within the PCA scores plots, or only discriminated by PCs 3 or above.

Table 5.5 Number of correct vs incorrect location classifications for samples in the second combined XRD calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Classified	% Correct
1	0	3	2, 5, 6	0
2	0	3	1, 12, 19	0
3	2	1	7	67
4	0	3	5 & 13	0
5	3	0	-	100
6	2	1	8	67
7	2	1	3	67
8	3	0	-	100
9	1	2	19 & 22	33
10	3	0	-	100
11	0	3	7, 13, 17	0
12	2	1	9	67
13	3	0	-	100
14	3	0	-	100
15	3	0	-	100
16	3	0	-	100
17	3	0	-	100
18	2	1	22	67
19	3	0	-	100
20	0	3	3, 4, 7	0
21	0	3	9, 12, 18	0
22	1	2	18 & 21	33
% Total Correct				59

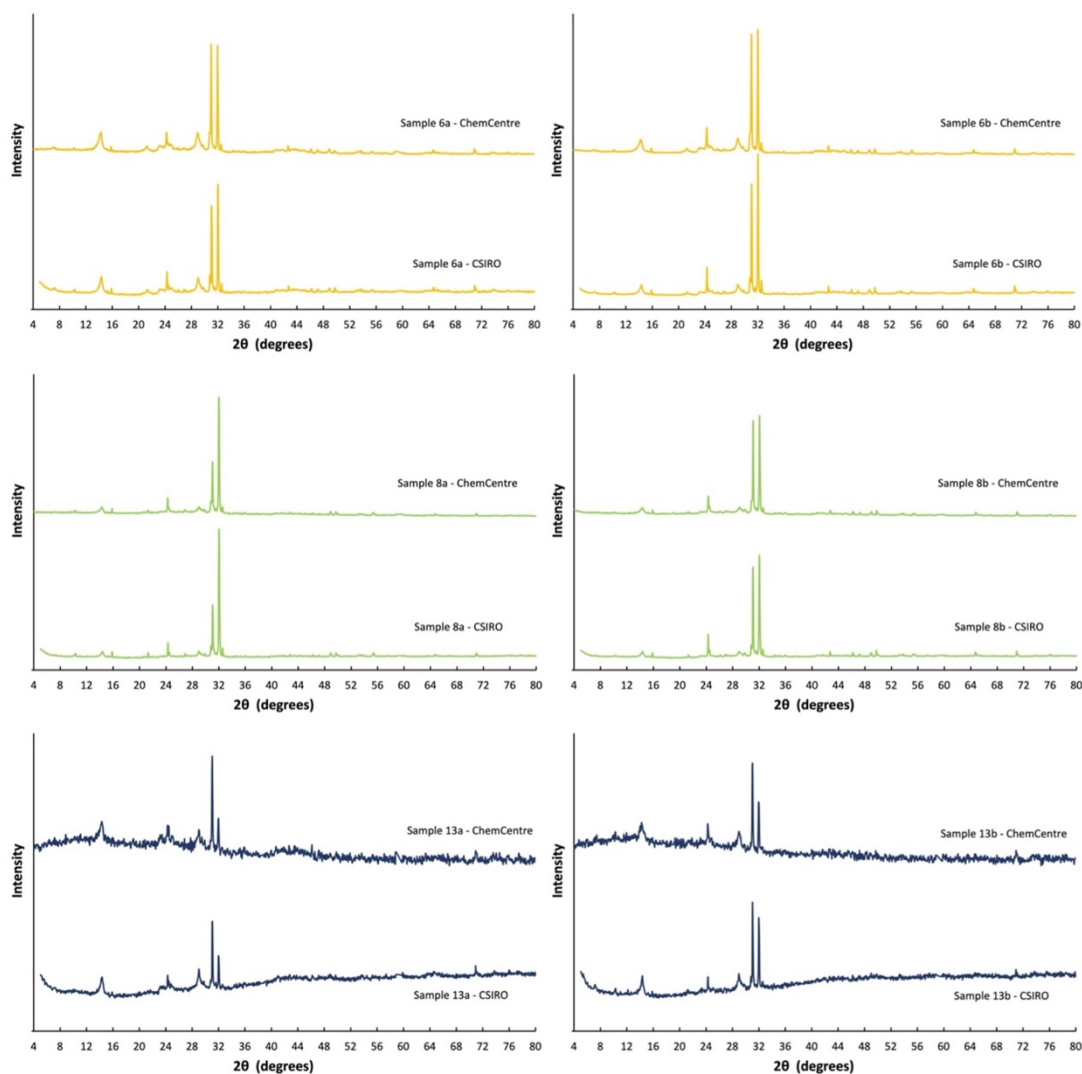


Figure 5.14 XRD patterns obtained from soil samples from Locations 6, 8, and 13, illustrating both the variation in patterns between different instrumentation, and between duplicate samples.

This LDA model was then used to predict the locations of 22 ChemCentre samples from the validation dataset. Nine samples, or 40.9%, had their location correctly predicted (Table 5.6). Despite Locations 2, 3, and 6 showing misclassifications within the calibration set, they all achieved 100% validation accuracy, though the confidence surrounding these predictions was very low; the associated discriminant values (Table 5.7) showed that they were close to being predicted as originating from at least one other location. Most of the samples that exhibited 100% calibration accuracy also achieved the same validation accuracy, except for Locations 8, 13, and 19, suggesting noteworthy variation between the different XRD patterns associated with these samples. Despite unsuccessful discrimination within both scores plots and

all calibration samples being misclassified, the validation sample for Location 2 had its origin predicted correctly; this was presumably by chance, as the discriminant values indicated that the centroids of several other locations were also situated very closely (within 4%). All of the remaining samples that were unable to be discriminated within the scores plots of either XRD model again had their locations predicted incorrectly.

Table 5.6 Number of correct vs incorrect location predictions for samples in the ChemCentre validation set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Predicted	% Correct
1	0	1	6	0
2	1	0	-	100
3	1	0	-	100
4	0	1	3	0
5	1	0	-	100
6	1	0	-	100
7	0	1	3	0
8	0	1	18	0
9	0	1	21	0
10	1	0	-	100
11	0	1	17	0
12	0	1	9	0
13	0	1	19	0
14	1	0	-	100
15	1	0	-	100
16	1	0	-	100
17	1	0	-	100
18	0	1	21	0
19	0	1	13	0
20	0	1	3	0
21	0	1	9	0
22	0	1	9	0
% Total Correct				41

Table 5.7 Discriminant values of replicates from the ChemCentre XRD validation dataset (rounded to three decimal places), with correct predictions shaded green and incorrect predictions shaded red. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

Sample Replicate	Discriminant Values																						Predicted Location	Closest Value Within (%)
	Location 1	Location 2	Location 3	Location 4	Location 5	Location 6	Location 7	Location 8	Location 9	Location 10	Location 11	Location 12	Location 13	Location 14	Location 15	Location 16	Location 17	Location 18	Location 19	Location 20	Location 21	Location 22		
1b	-6.147	-8.042	-20.786	-18.202	-10.078	-3.254	-24.660	-4.513	-7.420	-81.038	-25.565	-7.026	-21.471	-8.364	-148.488	-22.773	-36.133	-3.940	-13.680	-20.285	-6.038	-4.719	6	21
2b	-3.772	-3.532	-7.800	-6.801	-4.178	-8.157	-10.365	-13.384	-3.665	-75.360	-12.881	-4.163	-9.326	-7.926	-137.397	-17.601	-24.913	-7.272	-5.437	-7.617	-4.747	-5.323	2	4
3b	-7.997	-6.109	-3.995	-4.134	-5.792	-17.872	-5.415	-25.990	-6.781	-75.939	-8.958	-7.894	-5.896	-13.310	-133.491	-19.139	-21.791	-15.660	-5.053	-4.033	-9.634	-11.667	3	1
4b	-7.570	-6.139	-4.099	-4.311	-4.979	-17.188	-6.059	-25.986	-7.121	-85.749	-11.824	-8.098	-7.388	-15.130	-147.033	-23.311	-26.869	-15.806	-5.796	-4.170	-10.095	-11.819	3	2
5b	-4.659	-4.216	-5.860	-5.397	-3.542	-11.086	-8.468	-18.260	-4.914	-85.697	-13.842	-5.488	-8.983	-11.695	-149.848	-22.256	-28.563	-10.504	-5.594	-5.799	-6.840	-7.699	5	19
6b	-7.840	-10.174	-24.749	-21.780	-12.554	-3.378	-28.793	-3.720	-9.373	-82.785	-29.260	-8.738	-25.083	-9.458	-151.175	-24.725	-39.328	-4.308	-16.515	-24.165	-7.396	-5.685	6	10
7b	-10.527	-8.371	-3.413	-3.975	-6.726	-22.612	-4.481	-32.667	-9.617	-86.975	-10.955	-10.618	-6.497	-18.445	-146.223	-24.783	-26.168	-20.425	-6.400	-3.531	-13.155	-15.536	3	3
8b	-5.691	-6.702	-18.168	-15.579	-9.632	-4.122	-20.862	-4.980	-5.788	-66.465	-19.219	-5.504	-16.838	-5.060	-128.210	-15.620	-26.898	-3.334	-10.664	-17.626	-4.401	-3.751	18	13
9b	-4.165	-4.146	-11.375	-9.645	-6.306	-6.228	-13.595	-9.168	-3.636	-64.390	-12.913	-3.873	-10.741	-4.800	-123.813	-13.267	-21.627	-4.809	-6.407	-11.014	-3.580	-3.850	21	2
10b	-64.988	-59.595	-65.895	-63.125	-71.640	-70.096	-61.317	-64.965	-55.459	-4.418	-34.721	-57.022	-48.464	-38.944	-19.810	-19.559	-17.127	-58.638	-53.110	-65.086	-53.715	-57.641	10	288
11b	-26.327	-22.316	-24.395	-22.946	-28.897	-33.983	-22.225	-34.845	-20.320	-21.213	-8.867	-21.677	-15.294	-13.957	-54.513	-5.155	-4.504	-26.370	-17.854	-23.960	-20.667	-23.899	17	14
12b	-4.543	-3.715	-7.041	-6.122	-4.893	-9.888	-8.976	-14.967	-3.679	-67.485	-9.896	-4.341	-7.472	-7.052	-125.937	-14.317	-20.258	-8.136	-4.612	-6.857	-4.881	-5.904	9	1
13b	-8.564	-6.264	-5.086	-4.860	-7.367	-17.912	-5.727	-24.497	-6.385	-61.551	-5.827	-7.502	-4.667	-10.209	-114.183	-13.069	-15.091	-14.577	-4.502	-5.017	-8.666	-10.926	19	4
14b	-6.132	-5.457	-12.297	-10.549	-8.640	-8.781	-13.830	-11.006	-4.580	-53.303	-10.367	-5.050	-9.979	-4.013	-107.815	-9.350	-16.356	-6.278	-6.717	-11.912	-4.485	-5.214	14	12
15b	-125.303	-116.783	-119.596	-116.845	-132.653	-134.371	-111.332	-127.088	-111.334	-10.269	-74.041	-113.657	-94.850	-88.107	-4.300	-54.680	-45.272	-117.804	-104.929	-118.673	-109.599	-115.819	15	139
16b	-18.944	-16.048	-20.569	-18.802	-22.052	-24.288	-19.470	-24.846	-14.194	-26.564	-8.170	-15.205	-12.921	-8.665	-65.670	-3.910	-5.997	-18.116	-13.509	-20.095	-14.052	-16.402	16	53
17b	-26.238	-21.898	-22.105	-20.970	-27.899	-35.427	-19.668	-37.355	-20.170	-23.423	-7.423	-21.591	-13.469	-14.909	-56.635	-5.721	-3.976	-27.638	-16.714	-21.731	-21.013	-24.572	17	44
18b	-4.236	-4.479	-12.542	-10.710	-6.672	-5.568	-15.053	-8.228	-3.933	-66.174	-14.454	-4.128	-12.106	-5.028	-126.623	-14.292	-23.373	-4.462	-7.267	-12.169	-3.714	-3.760	21	1
19b	-9.994	-7.344	-5.469	-5.320	-8.675	-20.014	-5.749	-26.629	-7.401	-58.566	-5.101	-8.616	-4.514	-10.862	-109.261	-12.388	-13.563	-16.198	-4.970	-5.406	-9.800	-12.334	13	10
20b	-8.146	-6.125	-3.820	-3.917	-5.946	-18.212	-4.968	-26.242	-6.752	-73.789	-8.149	-7.824	-5.307	-12.954	-130.428	-18.149	-20.464	-15.719	-4.742	-3.839	-9.544	-11.680	3	0
21b	-4.299	-4.083	-10.673	-9.087	-6.206	-6.932	-12.825	-10.132	-3.605	-63.578	-12.066	-3.964	-10.074	-4.954	-122.335	-12.919	-20.772	-5.373	-6.065	-10.344	-3.769	-4.204	9	5
22b	-4.282	-4.238	-11.339	-9.702	-6.358	-6.463	-13.705	-9.518	-3.747	-64.925	-13.052	-4.090	-10.923	-5.111	-124.422	-13.639	-21.961	-5.132	-6.578	-11.006	-3.829	-4.118	9	2

5.3.4 Reproducibility of data

Not only does this study highlight the value gained from interpretation of soil evidence and the applicability of the method, but it also showcases how instrument capabilities can affect the results obtained. Despite the modern capabilities of the XRD instrument at CSIRO, both ChemCentre and CSIRO models showed similar results when the data was analysed using PCA; the loadings were primarily influenced by the same reflections, and the resulting scores plots showed most samples obtained an equivalent degree of discrimination between the two models. CSIRO XRD patterns were less affected by noise and allowed for the detection of some additional minerals that could not be visualised within the ChemCentre patterns, due to the increased sensitivity associated with more modern instrumentation. However, these were present in such low quantities within the soil that their presence could not be confirmed, and their detection did not considerably alter the discrimination between samples. This demonstrates that the PCA methods utilised throughout this chapter are reproducible and robust, making them suitable for forensic casework.

For LDA, combining the ChemCentre and CSIRO data into a singular dataset produced sub-optimal results, due to the encapsulated variation attributed to sensitivity and noise. If a situation arose in forensic casework in which samples were analysed over multiple different instruments, it would be impractical to subject this collective data to predictive chemometric methods such as LDA without first gaining a proper understanding of the instrumental variation. Different diffractometers will show inevitable variations in their diffraction intensity data due to instrument stability, optical alignment, and the in-built procedures for data correction (94). For best results, it is therefore important that proficiency testing, validation, and casework are all carried out using the same instrumentation. Sometimes situations will occur in which equipment unavoidably changes over the course of a single investigation; this study has shown that this data is still valuable for questioned versus known comparisons if the examiner has a strong understanding of what differences will occur, so that they can be accounted for appropriately. By comprehending how to best interpret these differences and optimise the equipment available, these

methods can be suitable for use in forensic laboratories that may be faced with outdated instrumentation.

5.4 Conclusions

This chapter explored the use of XRD combined with chemometrics for analysing the quartz-recovered fine fraction of Swan Coastal Plain soils, which has never before been demonstrated in the open literature. The use of the quartz-recovered fine fraction has been developed to allow for detection of the subtle variations in minerals that could not otherwise be identified, due to the minimal amounts of clay and fine silt in quartz-dominated soils. PCA and LDA conducted on the entire XRD patterns allowed for discrimination of most samples, including those that produced visually similar patterns. The use of XRD on soils post-ATR-FTIR analysis provided additional discrimination of some samples, confirmed the associations made, and aided in improving the confidence of the results.

The use of the collective ChemCentre XRD patterns for PCA, which were produced using 'older' instrumentation settings, resulted in the discrimination of most of the soils based on the locations from which they originated. Some sites however, showed significant intra-site variability, though this was potentially due to anthropogenic influence. Factor loadings showed that the variation observed within the mineral content of the soil appeared to correlate with the soil's original dune system, and hence visual discrimination of the four dune systems was achieved. There also appeared to be some grouping within the ChemCentre model based on batches of samples that were analysed with 'old' versus 'new' instrumentation, highlighting the degree of instrumental variation due to the interpolation of data. However, the effect of this on the overall discrimination of samples was minimal.

The use of the 'newer' CSIRO XRD patterns for PCA resulted in a similar degree of clustering compared to the ChemCentre model. Due to the considerable similarities between loadings, the discrimination between soils was comparable between both models, with most samples situated in equivalent positions within the scores plots. The CSIRO instrumentation appeared to be more sensitive when operated at modern

capabilities, with improved signal-to-noise ratios allowing for the detection of some minerals seen in lower concentrations within the soil samples, e.g., mica. However, this did not improve the overall differentiation of soils, highlighting that the degree of discrimination achieved was still attainable using outdated instrumentation. It was therefore demonstrated that this method is both reproducible across different instrumentation, and robust in its results achieved, making it well-suited for use within forensic casework.

Using the combined XRD dataset for LDA post-PCA caused overestimation of the calibration accuracies (65.1%, 59.1%), with the validation accuracies lower than expected (36.4%, 40.9%). Most of the samples that were misclassified were unable to be discriminated within the scores plots of either XRD model. In addition, the majority of the correctly predicted samples exhibited low discriminant values across several classes, indicating that the confidence surrounding their separation was low. The LDA results could have potentially been improved by utilising a larger number of replicates within each class. The combined ChemCentre and CSIRO LDA models had increased uncertainty compared to the individual PCA models, as they encapsulated a substantial amount of variation due to instrumental analysis. This is an issue that could be encountered in forensic casework; to avoid this, analysis of both questioned and known samples should ideally be conducted using the same instrumentation when using this data for predictive interpretation.

Overall, utilising entire XRD patterns for chemometric analysis was successful at discriminating between most soils from different locations. Despite issues predicting the locations of these samples due to instrumental variation, this did not pose a major problem because the differentiation of these locations was already demonstrated through the use of other techniques in previous chapters. While XRD can be effective as a stand-alone method for analysis of the quartz-recovered fine fractions, it can only discriminate soils based on their mineral or crystalline components, unlike ATR-FTIR, and hence it is recommended to be used in sequence with other techniques to maximise the information obtained. XRD also achieved improved discrimination of some samples that were shown to overlap within the

ATR-FTIR PCA model produced in Chapter 4, and reduced the intra-site separation seen between others, demonstrating how these methods can be used in sequence with each other to provide complementary information. Therefore, chemometric analysis of the full XRD pattern shows strong promise for sequentially discriminating soils that cannot be entirely individualised through PCA of the ATR-FTIR data alone, with the conclusions reinforced by a scientific measure of similarity that can be presented as evidence in court. This technique would ideally be applied by forensic practitioners at the end of the sequence, to gain additional discrimination on soil samples that could not be differentiated using MSP or ATR-FTIR and to improve the confidence surrounding the previously obtained results. The XRD patterns produced through analysis also showed correlation with the dune system that the soil originated from, indicating that this technique could be further developed to provide intelligence to forensic investigators that can direct them to regions of interest during active criminal investigations.

Additionally, this thesis offers preliminary data towards validation of the XRD method, by demonstrating its ability to provide consistent information within a forensic investigation despite the use of multiple instruments. Within each XRD model, the majority of soils achieved comparable levels of discrimination due to very similar loadings, regardless of the capabilities of the instrumentation used. While this chapter also demonstrated that it is possible to utilise different diffractometers for chemometric analysis, the results are contingent on a strong understanding by the examiner of what variation will occur, so it can be accounted for and interpreted appropriately. For example, standard questioned versus known comparisons using PCA appeared to be mostly unaffected by the instrumental variation, with the different sensitivities and levels of noise having minimal effects on the overall discrimination of data. However, this was not the case for LDA, which required further examination of the discriminant values to assess the confidence in the model, and hence would need human interpretation of the data by the forensic examiner post analysis. Further research is required to explore other factors that may influence the results, such as sample collection or storage, in order to advance towards full validation of the method.

Chapter 6. Application of Developed Analytical Sequence Incorporating Chemometrics to a Forensic Case Simulation

Portions of this chapter have been published in the following article:

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics: Application to simulated forensic soil casework." *Forensic Chemistry*, 2023. 33: 100481.

6.1 Introduction

In Chapters 3 to 5, a sequence was developed and proposed for the forensic analysis and interpretation of sandy soil evidence. This sequence utilised several spectroscopic techniques in combination with chemometric methods to allow for the discrimination of soils from the Swan Coastal Plain based on their individual locations. Sequenced analysis can be especially useful for soils such as these, where the bulk samples are relatively similar and analysis techniques are limited due to their lack of clay and organic matter (15, 17, 73). It was shown that each technique allowed for the enhanced discrimination of different soils through the analysis of complementary characteristics within the sample, and hence outcomes were improved by performing them sequentially and interpreting the resulting data as a whole. Additionally, the use of a sequence allowed for more accurate identification of the chemical components responsible for the variance between soils, and improved the overall confidence surrounding the results by confirming them numerous times. However, in order to best assess the suitability of the sequence for use within forensic casework, it should be further tested through application to a case simulation.

Forensic casework generally involves the comparison of a questioned soil sample, recovered from an item or suspect, to known soil samples, collected from locations such as the crime scene or alibi sites (20, 29, 65, 76). These soils are examined physically and chemically to collect information to allow for an association or an exclusion to be made. With naturally occurring forms of trace evidence like soil, comparisons are complicated by issues such as non-representative sampling (the soil has already been unintentionally sampled by the item or suspect, likely retaining only portions of the bulk material), sample preparation and storage, limitations on sample size, and the ability of available analytical methods to detect key variation within the sample (10, 38). Therefore, it is important to ensure that appropriate methods are selected, and that these have been well-tested and verified before application to forensic casework.

Case simulations are especially useful for evaluating the capabilities of developed methods and showcasing how they can be integrated within forensic casework. Limited case studies for verification of proposed methods for forensic soil analysis have been documented within the literature, and these arguably should be more prevalent. In 2011, Reidy *et al.* ran a successful mock crime scene scenario in Mississippi, USA, in which students were asked to blindly conduct comparative principal component analysis (PCA) and discriminant analysis of soils based on elemental distribution fingerprints determined from inductively coupled plasma-mass spectrometry (ICP-MS) (10). All students were able to correctly classify their unknown suspect sample to its source, hence verifying the capabilities of the method. Additionally, the case simulation allowed for identification of the likely complications encountered within forensic casework; relatively wide areas of soil may have comparable elemental profiles if the soil type and source material are similar, and PCA alone may not be the best tool for data discrimination. Therefore, the results could be further improved by incorporating additional chemometric methods, as well as data obtained from analysis techniques that characterise other aspects of the soil.

In South Australia, Young *et al.* demonstrated the utility of chemometrics combined with high throughput eukaryote DNA sequencing (HTS) and mid-infrared (MIR) spectroscopy to discriminate between different soil sites in a crime scene setting, and were able to link an unknown soil to a particular reference location (66). This case simulation showed that HTS was both effective and robust to environmental variation, transfer and storage effects, and spatial variation. The results obtained through HTS also complemented the MIR spectroscopy soil profiling, once again suggesting the value of utilising several techniques within a sequence. Finally, Woods *et al.* reported a blind trial that demonstrated the use of microspectrophotometry (MSP), attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR), X-ray fluorescence spectroscopy (XRF), scanning electron microscopy coupled with energy dispersive X-ray spectroscopy (SEM-EDX), and laser induced breakdown spectroscopy (LIBS) (38). This analysis sequence was combined with PCA and linear discriminant analysis (LDA) for the effective discrimination and classification of

unknown Australian soil samples against a reference database. Individually, each of the instrumental techniques demonstrated relatively high discrimination of the soils analysed, however, when combined they resulted in full discrimination of the soil set.

Case simulations have also been utilised across other areas of forensic research, such as document examination, showcasing the value that these methods hold for analysis of forensic evidence. In 2019, Sauzier *et al.* demonstrated the power of video spectral comparator (VSC) spectroscopy with chemometrics for comparing visually similar ballpoint inks on paper documents (137). This blinded case simulation verified the suggested approach through comparison against traditional visual examination, and provided quantitative data to support expert opinions. Additionally, it demonstrated the challenges that may be faced during application of this method to forensic casework, by revealing that some inks with very similar optical properties may not be distinguishable through colour alone. Recommendations were also made to reduce the effects of background substrate and ageing on the results.

Not only are case simulations extremely important for research, they are also a part of the accreditation and validation processes dedicated to quality assurance (128, 138). In Australia, forensic laboratories are accredited through the National Association of Testing Authorities (NATA). Every year, a proficiency test is conducted in the form of a blinded case simulation, as part of the requirements for accreditation (138); this 'case' is analysed and reported as a 'real' case would be by each laboratory, and then the conclusions are assessed by NATA to evaluate their competency and detect any issues associated with methodology or evidence interpretation. These proficiency tests are crucial for evaluating whether current forensic methodologies are appropriate, reliable, and reproducible, and that the scientists can perform them accurately.

In this chapter, the sequence developed in Chapters 3 to 5 was verified through application to a blinded case simulation. MSP, ATR-FTIR spectroscopy, and X-ray diffraction (XRD) were conducted on the quartz-recovered fine fraction of soils collected from locations within the Swan Coastal Plain in Perth, Western Australia, as per the previous chapters. PCA was performed on each dataset to assess any

similarities or differences between a questioned soil and four known soils. LDA was then performed to predict the most likely source of the unknown soil sample. These results were then interpreted as a whole, to determine if an association could be made between the questioned soil and any of the known soils.

6.2 Experimental

6.2.1 Sample selection

A case simulation was generated by a third-party, who is an experienced forensic soil examiner with many years of forensic casework expertise, to allow for the blind analysis of five soil samples using the methods developed in previous chapters. The five soils utilised for this study were collected by the third-party as outlined in Sections 2.2 and 2.3.2. The soils selected originated from areas that shared many similarities, such as soil colour, dune system, and type of locations, in order to make the simulation challenging enough to test the full capabilities of the developed method. The brief provided along with the soil samples stated:

“A suspect has come to the attention of police, and a sample of soil has been recovered. The investigators wish to know if the soil corresponds to one of four potential locations, which you have been provided samples from; the crime scene, a potential secondary site, and two possible alibi sites. The investigators want the soil recovered from the suspect compared with the four collected control samples, using the methods you have developed, to see what can be determined.”

Information regarding the origin of the four known samples was communicated prior to analysis, however, all details surrounding the *suspect recovered* soil were not disclosed until after the spectroscopic and chemometric analyses had been completed and results were presented.

6.2.2 Sample analysis

Duplicate samples were obtained from each soil. The quartz-recovered fine fractions were prepared from each of the soil duplicates, as described in Section 2.4. Spectroscopic analysis was conducted on the extracted quartz fine fractions of each

of the sample duplicates using MSP, ATR-FTIR spectroscopy, and XRD, as described in Section 2.5. All ChemCentre XRD patterns were obtained using the newer instrument – Malvern PANalytical EMPYREAN III Diffractometer. Chemometric analysis was then conducted on the resulting data as outlined in Section 2.6.

While the different methods of analysis were conducted on the soils across multiple days, all analysis using each technique was conducted within one session; for example, all of the MSP analysis was conducted consecutively, with the instrument calibrated once at the beginning of the session. This was done to minimise the effects of daily fluctuations in the instrument performance, and to replicate the most accurate representation of a casework scenario (comparing questioned versus known samples).

6.3 Results and discussion

On receipt of the case simulation soils, forensic analysis was carried out according to the developed sequence, as outlined in Figure 6.1 and previously described in Sections 6.1 and 6.2 above. First, an overall visual examination of the soils was carried out, before the quartz-recovered fine fractions were isolated from each sample and analysed using the three spectroscopic techniques explored in previous chapters; MSP, ATR-FTIR spectroscopy, and XRD. PCA was conducted on the resulting data to identify any variation between samples and maximise their differentiation, and then LDA was conducted to predict the source of the *suspect recovered* soil. The following sections discuss the results generated at each stage of the sequence.

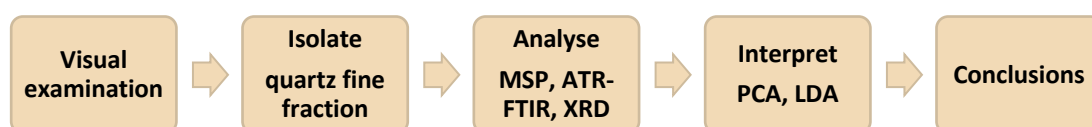


Figure 6.1 Flow diagram outlining the overall analysis sequence, previously developed throughout Chapters 3 to 5, that was applied to soil samples within the blinded case simulation.

The source of the *suspect recovered soil* was subsequently revealed to have originated from the *crime scene* site (Table 6.1). This information was unknown throughout the analysis and interpretation of the samples, however, is presented

upfront in this chapter to allow for comprehensive comparison and interpretation of the results.






Table 6.1 The soils utilised for the blinded case simulation, with their associated information. Details regarding the suspect recovered soil were kept confidential until all analyses were completed, and results were presented.

Location	Visual Appearance	Suburb	Dune System	Location Type
Alibi Site 1	Grey sand (heavy organic matter)	North Perth	Spearwood	Manicured park
Suspect Recovered (Crime Scene)	Grey sand	Leederville	Spearwood	Verge
Crime Scene	Grey sand	Leederville	Spearwood	Verge
Alibi Site 2	Grey sand	Wembley Downs	Spearwood	Verge / median strip
Potential Site of Interest	Grey sand	Hazelmere	Pinjarra Plain	Bushland (near road construction)

6.3.1 Visual examination of soil samples

The provided soil samples were first examined visually to note differences in appearance. Visual descriptions of each soil provided are outlined in Table 6.2. The *suspect recovered* soil was visually the most similar to the *crime scene* sample; both contained similar ratios of sand/quartz to organic matter, with rocks and seed pods found within both samples. The *crime scene* sample also contained a road-marking bead, indicative of its roadside location (verge), while the *suspect recovered* sample contained a glass fragment, possibly also due to originating from a roadside location (vehicular accident) or damage to property (i.e., at the crime scene). In a real-life investigation, this glass shard would likely have been compared to other glass samples collected from the crime scene, however, as this was a case simulation with a focus on soil, no further analysis on the glass was conducted.

Table 6.2 Visual descriptions and photographs of the soil samples provided for analysis within the blinded case simulation.

Sample	Photograph	Description
Alibi Site 1		Grey sand dispersed through abundant dark brown organic matter, lots of brown mulch/bark, sticks and leaves present, minimal quartz throughout soil
Crime Scene		Grey sand with moderate amount of brown mulch, small white and grey rocks present (tentatively identified as limestone) as well as some small brown-orange leaves/seed pods and a few white roots/strands, spherical road-marking bead
Alibi Site 2		Grey sand with minimal brown mulch, very low on organics, dominated by sand and quartz
Potential Site of Interest		Grey sand with small sticks and dried organic matter, light brown leaves present as well as several long strands of grass or plant stems/roots
Suspect Recovered (Unknown)		Grey sand with moderate amount of brown mulch, small white and grey rocks present (tentatively identified as limestone) as well as some small brown-orange leaves/seed pods and a few white roots/strands, shard of glass

While the appearance of the questioned sample was most comparable to the *crime scene* sample, it still shared similarities with most of the other soils, and therefore could not confidently be attributed to an individual source above the three, nor conclusively above others outside of the four presented. The only soil that was visually different to the *suspect recovered* sample was the soil from *alibi site 1*, due to the increased level of organic matter present and sparse amount of quartz grains; it was considered unlikely that these two soils originated from the same location. While the variation could be the product of non-representative sampling, this was noted as a potential difference.

6.3.2 Microspectrophotometric analysis of soil samples

MSP spectra were collected from the quartz-recovered fine fractions of the supplied soils (Figure 6.2), as outlined in Section 2.5. As expected, the soils had minimal distinctive features that separated them from the other sites. Both *alibi site 1* and *alibi site 2* soils had more distinctive curves in comparison to the relatively straight spectra obtained from the other locations, with exponential slopes that indicated higher proportions of orange-red components, however, this difference was undetectable by visual examination. The visually linear spectra were indicative of the similar shades of grey exhibited by all the soils. While *alibi site 1* was noticeably different in appearance, due to its higher concentration of dark brown organic matter, most of this was removed during preparation, and hence the quartz-recovered fine fraction was not representative of this. Colour determination methods therefore may be more accurate when applied to the bulk soil sample, as previously demonstrated in Chapter 3, however, this is generally not possible when working with partially-representative, recovered forensic samples. The spectrum of the *suspect recovered* soil was most aligned with that of the *crime scene* soil, but still showed a strong likeness with the other samples. Chemometric methods were subsequently employed to identify and enhance any subtle differences within the spectra as described below.

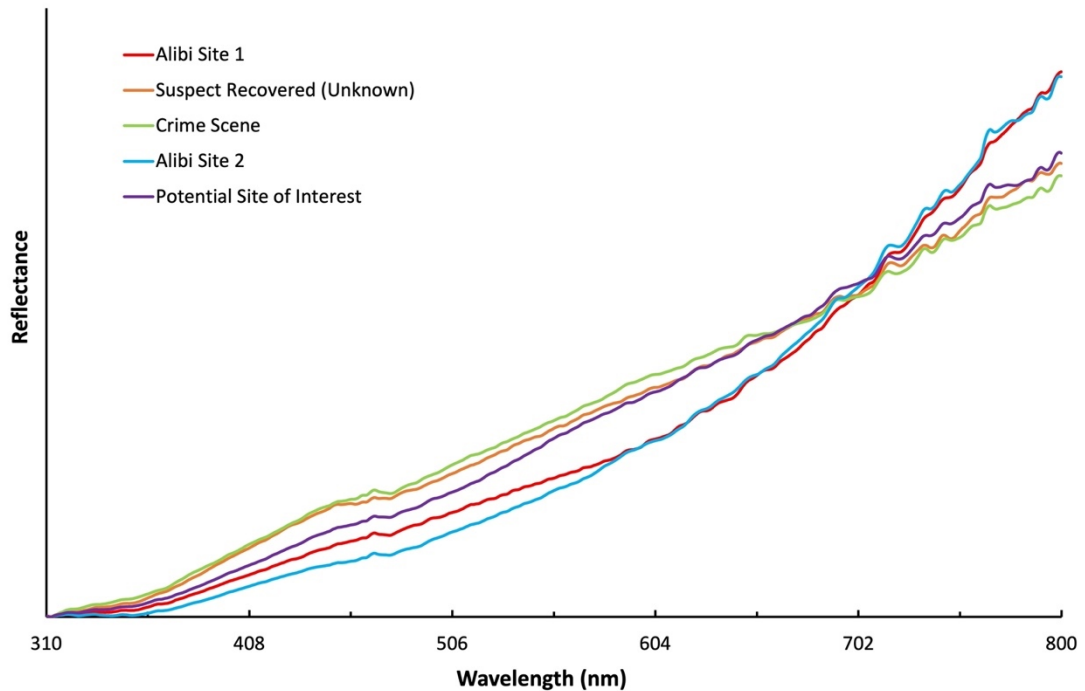


Figure 6.2 Baseline corrected and normalised MSP reflectance spectra showing the variability in composition of the soil samples collected from differing locations representative of a blinded case simulation.

6.3.2.1 Principal component analysis utilising entire MSP spectra

As per the method applied in Section 3.3.1, PCA was conducted on the entire MSP replicate spectra to maximise any variation between samples; 99.1% of the total variance in the dataset could be described by the first four principal components (PCs) (Appendix 6.1). Three-dimensional score plots generated using these PCs (Figure 6.3) showed wide-spread, loosely defined clusters based on the source of the soil. Samples from the *potential site of interest* were the best separated from the rest of the population, however, they still exhibited minimal overlap with other clusters. Samples from the *suspect recovered* soil were spread out considerably, overlapping with all the other soils. Whilst different colour regions within the spectra were associated with positive and negative loadings across the principal components (Figure 6.4), the comparable grey colourings of all the soils meant that the variation detected within these regions was not significant enough to allow for their discrimination. There was also a substantial amount of noise encapsulated within the loadings, which contributed to the separation seen and interfered with the results.

Based on the level of overlap observed within the scores plot, the *suspect recovered* soil was unable to be differentiated from or associated with any of the known sites.

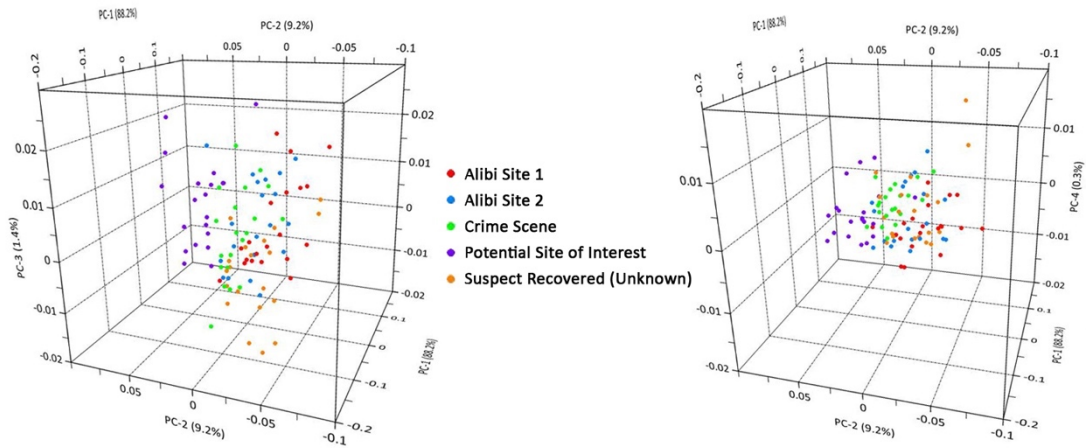


Figure 6.3 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding MSP spectra.

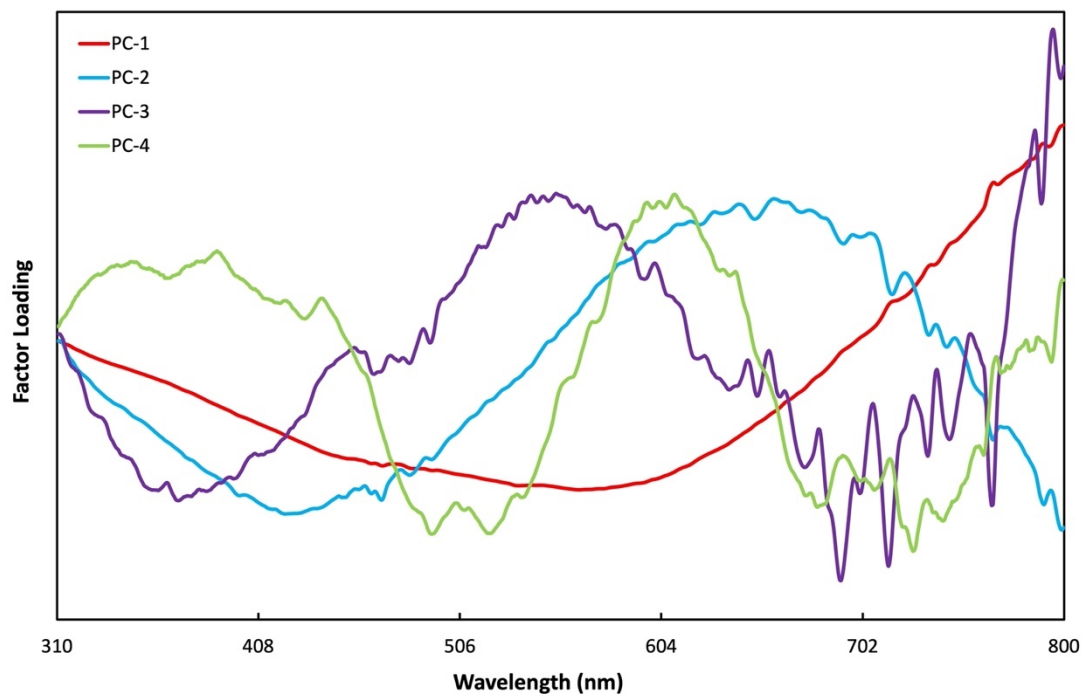


Figure 6.4 Factor loadings plot of PCs 1-4 for PCA of the blinded case simulation MSP reflectance dataset.

6.3.2.2 Principal component analysis utilising average MSP spectra

As per Section 3.3.2, the MSP spectra collected from the quartz-recovered fine fractions of the five supplied soils were averaged to obtain two average spectra per soil. PCA performed on these average spectra revealed that 99.7% of the total variance in the dataset could be described by the first four PCs (Appendix 6.2). Three-dimensional score plots using these PCs (Figure 6.5) showed improved separation between soils compared to the previous full MSP model. Samples from the *potential site of interest* were tightly clustered and well-separated from all other soils, indicating significant differences between this soil and the *suspect recovered*, and therefore they were unlikely to have originated from the same location. Samples from *alibi site 1*, although displaying significant separation between intra-site samples along PC-3, did not exhibit any overlap with other sites. However, they could not be confidently discriminated from the rest of the population due to their close proximity to several other soils. The samples from the *crime scene* and *alibi site 2* also experienced significant intra-site separation, causing these groups to overlap with one another.

The loadings for PCs 1-3 were comparable to those for the previous full MSP model and hence were unable to provide any further association between samples (Figure 6.6); the only difference was that the influence of PC-2 was reversed i.e., shorter wavelengths of light were associated with negative influence (previously positive) and longer wavelengths were associated with positive influence (previously negative). The detection of noise was again a significant issue, substantially contributing to the loadings across PCs 2-4 more so than in the previous model. The *suspect recovered* samples were not clustered together along PCs 1 or 3, nor were they situated nearby any of the other soils, so were not able to be associated with a known sample. The inclusion of PC-4 allowed for improvement in the clustering and discrimination of all of the known soil groups, however, the *suspect recovered* samples still showed significant intra-site separation along PC-1.

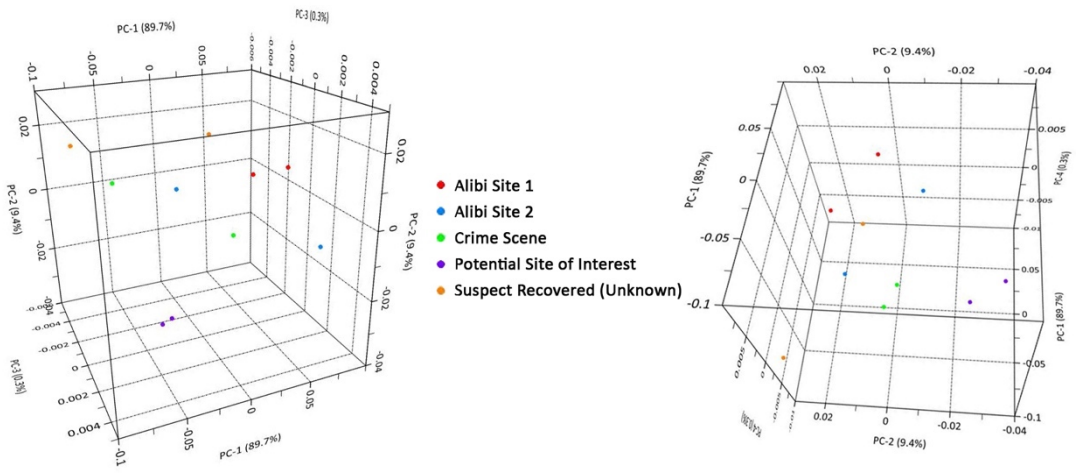


Figure 6.5 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding average MSP spectra.

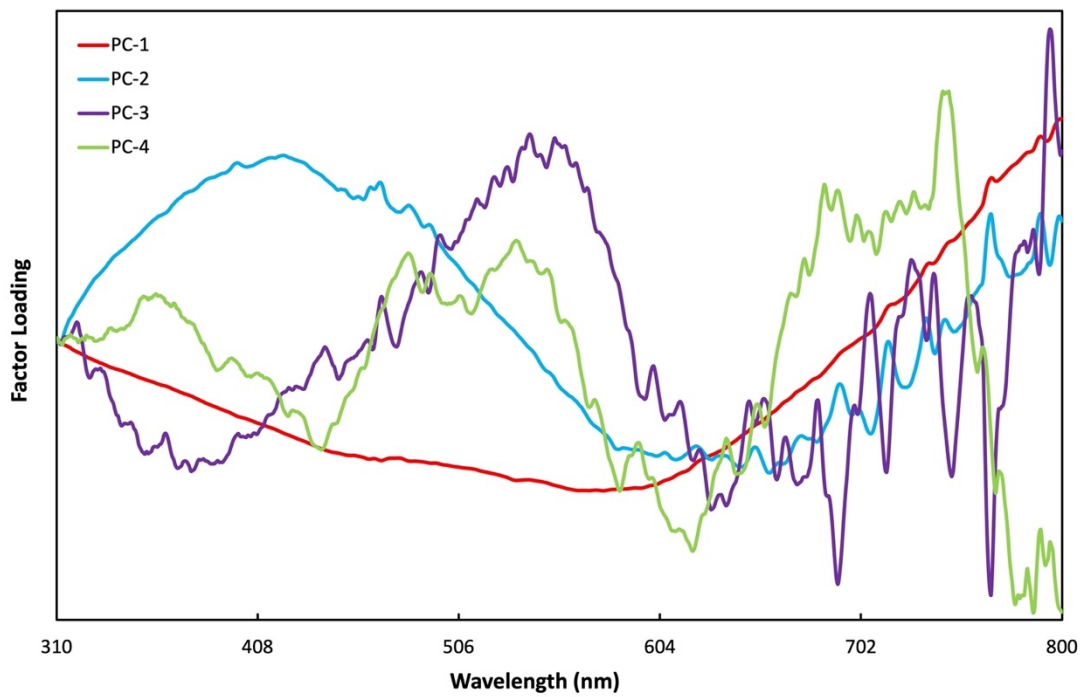


Figure 6.6 Factor loadings plot of PCs 1-4 for PCA of the blinded case simulation average MSP reflectance dataset.

6.3.2.3 Principal component analysis utilising L*a*b* colourimetric values

As per Section 3.3.3, the average MSP spectra collected above were converted into L*a*b* chromaticity values (Appendix 6.3), where L* is a measure of the lightness, a* is a measure of the green-red components, and b* is a measure of the blue-yellow components. PCA performed on these L*a*b* values (without baseline correction or normalisation) revealed that 100% of the total variance in the dataset could be described by the first three PCs (Appendix 6.4). Three-dimensional score plots using these PCs (Figure 6.7) revealed clustering and discrimination of the samples from *alibi site 2* and the *potential site of interest* respectively, indicating that these sites were not associated with the *suspect recovered* soil.

Samples from the remaining known sites however, showed significant intra-site separation within the model on at least one PC; for example, one *crime scene* sample was situated at the extreme positive end of PC-1 whereas the other was at the extreme negative. The loadings across PC-1 were largely influenced by variation in the lightness of the soil (Figure 6.8), and hence the separation between samples was due to the large difference observed in the L* values obtained. The *suspect recovered* samples were also separated from each other across PCs 3 and 2, due to variation in their red-green and blue-yellow components; sample 'a' had higher a* and b* values causing it to be positioned at a more positive position along both axes than the 'b' sample. The substantially different values highlight the high level of variation within these soils, even from samples that we know have originated from the same location. This can lead to challenging questioned versus known comparisons; are the differences seen caused by natural variation within the site or do they indicate meaningful variation caused by their origin? Due to the excessive separation between the two *suspect recovered* samples, this soil again did not show clustering with any of the known samples, preventing it from being associated with a single source.

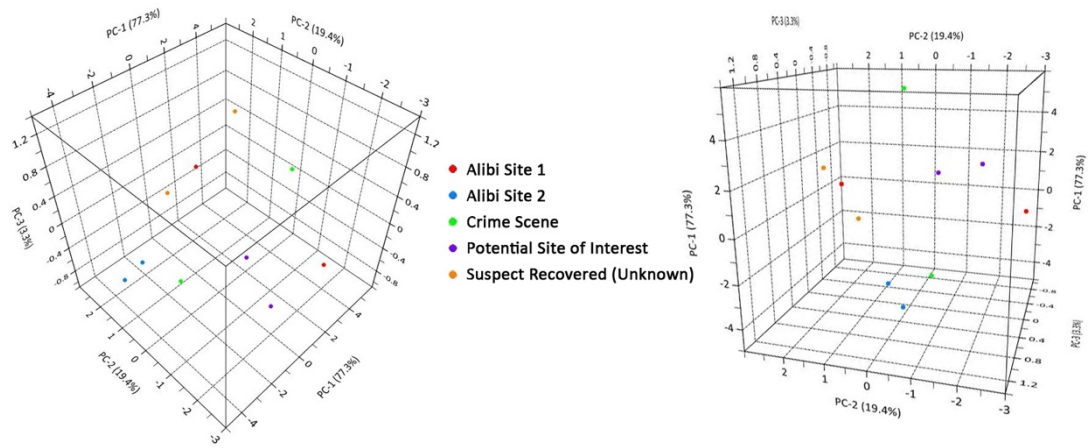


Figure 6.7 3-dimensional PCA scores plots generated using the first three PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding $L^*a^*b^*$ chromaticity values.

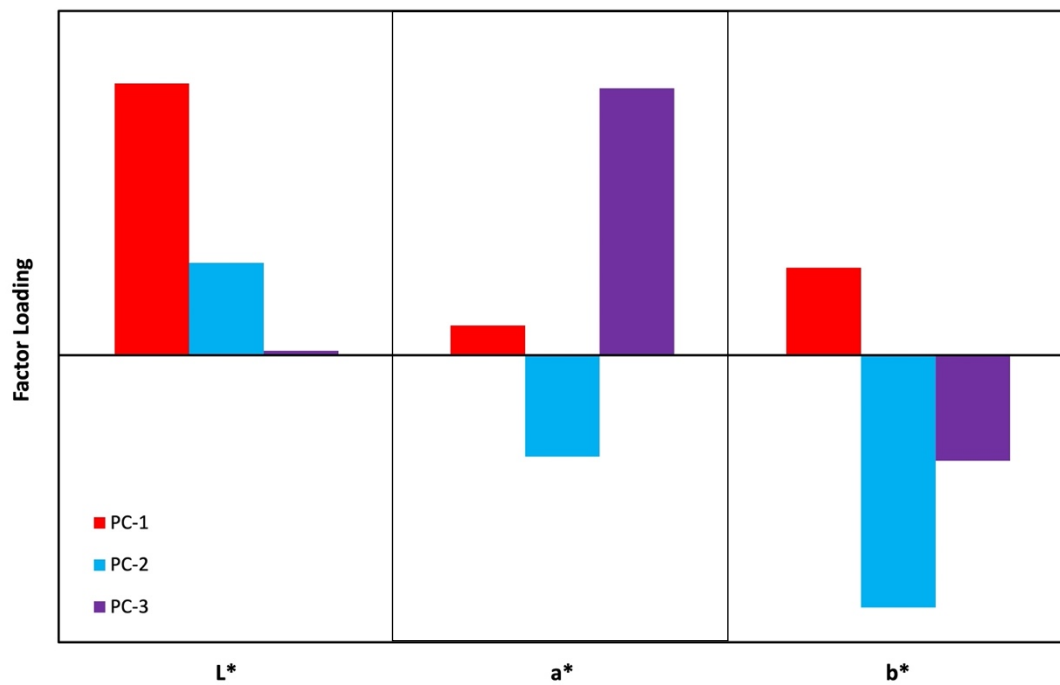


Figure 6.8 Factor loadings plot of PCs 1-3 for PCA of the blinded case simulation $L^*a^*b^*$ chromaticity values dataset.

6.3.2.4 Linear discriminant analysis

LDA was performed on the entire MSP dataset with each location treated as an individual class, returning a calibration accuracy of 76.2% (Appendix 6.5). Samples from the *crime scene* and *potential site of interest* had the highest percentage of correct classifications, most likely due to the tighter clustering (and hence less variation) seen within these groups. Because of the high degree of overlap between groups in the scores plot, none of the soils achieved a 100% classification accuracy.

This LDA model was then used to predict the source of the *suspect recovered* soil samples. 65% of sample spectra were predicted as originating from the *crime scene*, and the other 35% were predicted as *alibi site 1* (Table 6.3). The discriminant values (Appendix 6.6) for a significant number of these predictions indicated that they were close to being associated with several other locations, so the confidence surrounding these results was low. The increased classification accuracy of the *crime scene* samples within the LDA model meant that the predictions attributing the *suspect recovered* samples to the *crime scene* were, in general, slightly better separated than those that attributed the samples to *alibi site 1* (average of + 64% vs + 25%). This, along with the greater number of predictions, indicated that the *suspect recovered* samples more than likely originated from the *crime scene* rather than *alibi site 1*, however, this could not be determined with confidence.

Table 6.3 Predictions for samples in the MSP blinded case simulation validation set using a 4-PC LDA model.

Predicted Location	Suspect Recovered Samples
Alibi Site 1	7
Crime Scene	13
Alibi Site 2	0
Potential Site of Interest	0

LDA was then performed on the average MSP dataset, returning a calibration accuracy of 100% (Appendix 6.7). Despite the spread of samples within the scores

plot, and the resulting lack of discrimination between soils, all samples were correctly classified to their original source. This LDA model was then used to predict the source of the *suspect recovered* soil samples. 50% of the samples were predicted as originating from *alibi site 1*, and the other 50% were predicted as the *crime scene* (Table 6.4). The discriminant values (Appendix 6.8) indicated that while the *alibi site 1* prediction was of high confidence (+ 327%), the *crime scene* predicted sample was also very close to the centroid of the *alibi site 2* soils, and therefore demonstrated low confidence (+ 4%). While this implied that the *suspect recovered* soil more than likely originated from *alibi site 1*, the limited sample size meant that these results may not have been accurately representative; sample heterogeneity was demonstrated to cause issues in Chapter 3, and it is likely that repeated MSP analysis of this soil sample would lead to different results.

Table 6.4 Predictions for samples in the average MSP blinded case simulation validation set using a 2-PC LDA model.

Predicted Location	Suspect Recovered Samples
Alibi Site 1	1
Crime Scene	1
Alibi Site 2	0
Potential Site of Interest	0

Finally, LDA was performed on the L*a*b* values, returning a calibration accuracy of 50.0% (Appendix 6.9). *Alibi site 2* exhibited correct classification of all samples, most likely due to its tighter clustering (and hence less variation) seen within the scores plot. Despite being well clustered, the *potential site of interest* also only had one sample classified correctly, with the other attributed to the *crime scene*; while these groups were positioned relatively far apart, the *potential site of interest* soils were situated in the centre of the spread of the *crime scene* samples.

When this LDA model was used to predict the source of the *suspect recovered* samples, 100% of sample spectra were predicted as originating from the *crime scene* (Table 6.5). However, the discriminant values (Appendix 6.10) indicated that the

confidence surrounding these predictions was relatively low, with most of the other sites achieving similar values. Interestingly, while *alibi site 2* showed the best separation from *suspect recovered* sample 'a', it was the closest site to sample 'b', highlighting the large degree of variability within these soils from the same location.

Table 6.5 Predictions for samples in the L*a*b* chromaticity values blinded case simulation validation set using a 2-PC LDA model.

Predicted Location	Suspect Recovered Samples
Alibi Site 1	0
Crime Scene	2
Alibi Site 2	0
Potential Site of Interest	0

6.3.2.5 Summary of results

Overall, chemometric interpretation of MSP data highlighted the degree of microscopic colour variation within soils from the same site. As was demonstrated in Chapter 3, MSP is most valuable for the discrimination of highly coloured soils. However, all of the case simulation soils were a similar grey colour and hence, MSP was not expected to be particularly useful for their discrimination. The use of entire MSP spectra for PCA was unable to suggest any associations or discriminations between soils due to the high degree of sample heterogeneity and noise, as was encountered in Section 3.3.1. The use of average spectra improved the discrimination seen within the model, allowing the *potential site of interest* soil to be excluded as the origin of the *suspect recovered* soil. The conversion of these spectra to L*a*b* values increased the discrimination even further, allowing for both the *potential site of interest* and *alibi site 2* soils to be excluded. While the improvements demonstrated through the use of average MSP spectra aligned with the outcomes in Chapter 3, the L*a*b* results were unexpected.

MSP analysis combined with predictive LDA produced a total accuracy of 72%; predictions utilising full MSP spectra resulted in 65% correct, average MSP spectra

resulted in 50% correct, and L*a*b* values generated from average MSP spectra resulted in 100% correct. The results obtained by the full MSP spectra and L*a*b* values were unexpected, as they both exhibited much lower accuracies within the LDA models produced in Chapter 3. However, the confidence surrounding most of these predictions was relatively low in comparison to that obtained through the use of other techniques, so it is questionable whether these results would actually be consistently reproducible within casework. As all of the incorrectly predicted *suspect recovered* samples across all MSP LDA models were attributed to *alibi site 1*, which was previously indicated to be different through visual examination of the soil, these predictions could be cautiously overlooked. So while the use of MSP methods and chemometrics could not suggest an association between the *suspect recovered* soil and any of the known soils, they did allow for two exclusions to be made. Therefore, they are still worth applying to pairwise soil comparisons, as was suggested in Chapter 3.

6.3.3 Infrared spectroscopic analysis of soil samples

ATR-FTIR absorbance spectra were collected from the quartz-recovered fine fractions of the five supplied soils (Figure 6.9), as outlined in Section 2.5. These spectra exhibited many similarities, with minor variation only seen in the level of absorbance measured. All samples contained comparable levels of kaolinite, quartz, and H₂O/CO₂ organics. The two *alibi* sites exhibited higher concentrations of gibbsite, goethite, and humus in comparison to other samples, whilst the *crime scene*, *potential site of interest*, and *suspect recovered* soils displayed equally lower concentrations of these compounds. The spectrum of the *suspect recovered* soil was again most aligned with that of the *crime scene* soil, particularly evident within the gibbsite and goethite peaks at approximately 3450 and 3510 cm⁻¹. However, the *suspect recovered* samples still shared noteworthy similarities with the other spectra, and hence chemometric methods were applied to objectively identify and maximise the differentiation between samples.

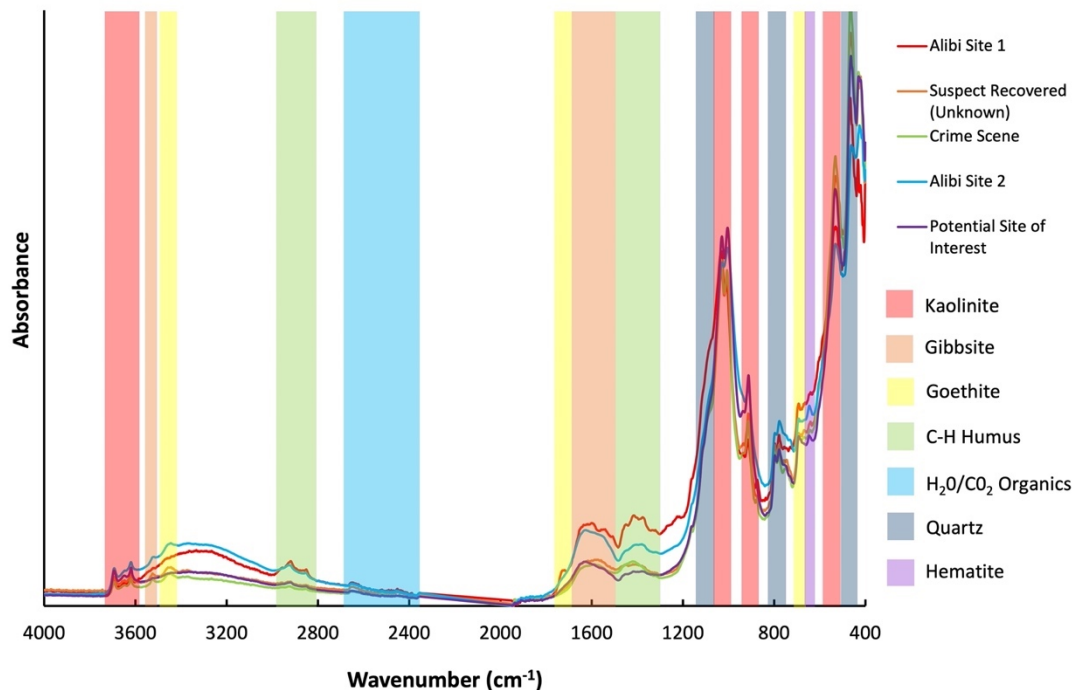


Figure 6.9 Baseline corrected and normalised ATR-FTIR absorbance spectra showing the variability in composition of the soil samples collected from differing locations representative of a blinded case simulation. Annotations were based off peak assignments outlined in Chapter 4 - Table 4.1.

6.3.3.1 Principal component analysis

Following the method demonstrated in Section 4.3.2, PCA performed on these ATR-FTIR spectra revealed that 95.0% of the total variance in the dataset could be described by the first three PCs (Appendix 6.11). Three-dimensional score plots generated using these PCs (Figure 6.10) resulted in well-defined clustering for the *suspect recovered* soil, but significant intra-site separation based on individual samples for the soils from the four known sites. The largest of this was exhibited by the *crime scene* samples; the two samples were substantially separated across PCs 1-3. Comparing the loadings plot (Figure 6.11) with representative ATR-FTIR spectra from these samples (Appendix 6.12), their positioning was influenced based on the difference in height of the kaolinite and quartz peaks within the fingerprint region ($\sim 400\text{--}550/910/1080\text{ cm}^{-1}$), as well as the gibbsite/goethite/humus peaks ($1350\text{--}1700/3200\text{--}3600\text{ cm}^{-1}$). For example, sample 'b' was situated more positively than sample 'a' along PC-1 due to smaller kaolinite and quartz peaks ($400\text{--}550\text{ cm}^{-1}$) and larger gibbsite/goethite/humus peaks (all), and more negatively along PC-2 again due

to smaller quartz peaks (all) and larger kaolinite peaks ($910/1080\text{ cm}^{-1}$). The replicates from sample 'b' of *alibi site 2* were clustered nearby to those from sample 'a' from the *potential site of interest*, and examination of their spectra showed that these soils were very similar in composition (Appendix 6.12); the only visual differences were in peak heights attributed to goethite (3450 cm^{-1}), gibbsite (3520 cm^{-1}), hematite (610 cm^{-1}), and one kaolinite peak (910 cm^{-1}).

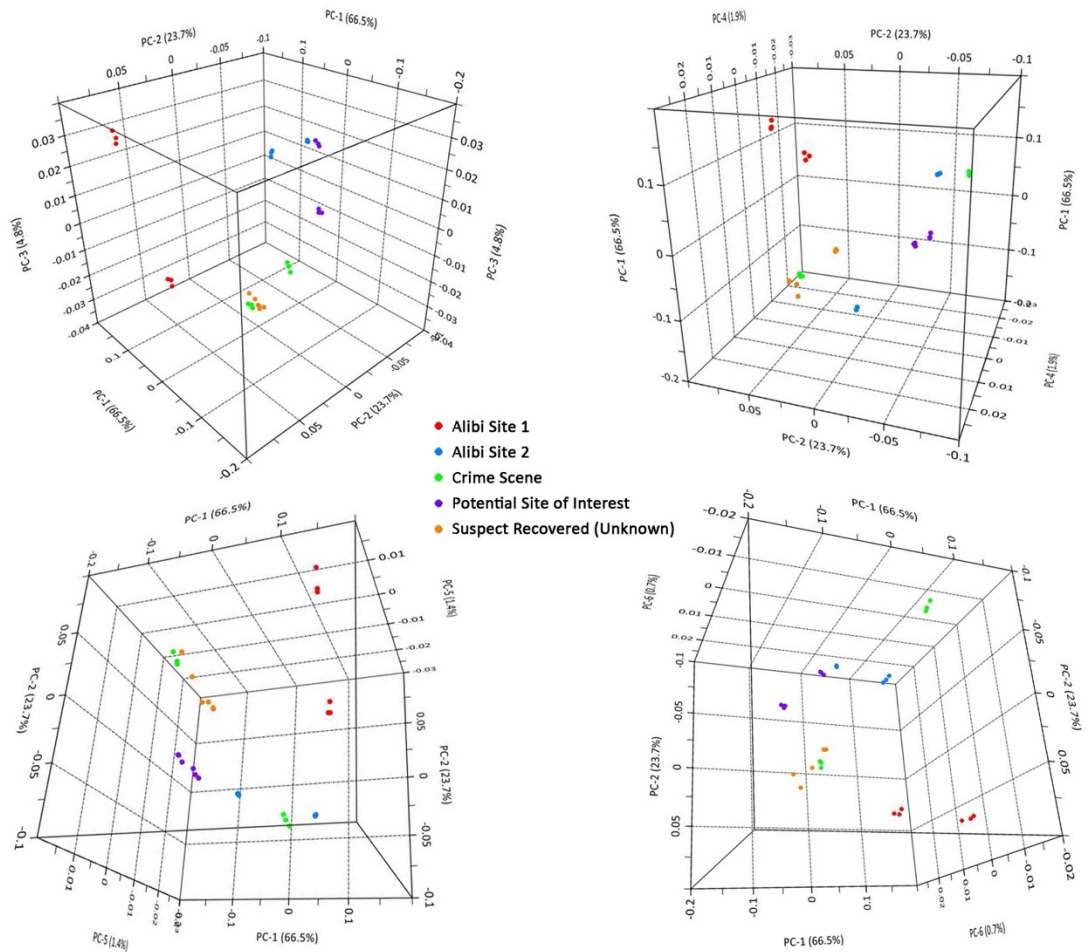


Figure 6.10 3-dimensional PCA scores plots generated using the first six PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding ATR-FTIR spectra.

Samples from *alibi site 1* were well-separated from all other soils, indicating that they were unlikely to have originated from the same location as the *suspect recovered* soil. Despite the intra-site separation within the known soils, the only overlap exhibited across PCs 1-3 within the scores plot was between the *suspect recovered* samples and those from the *crime scene*, indicating apparent similarities between

these soils. No matter which combination of PCs were used to visualise the model, the *suspect recovered* samples were consistently clustered closest to one of the *crime scene* samples (sample 'a'), highlighting the likeness of these groups.

The loadings appeared to focus more on the variation in peak height to separate spectra, especially of the larger peaks within the fingerprint region. However, the presence or absence of minerals indicated by smaller peaks, such as the goethite and gibbsite peaks at 3450/3520 cm^{-1} , could have perhaps been more indicative of the origin of the soil. The inadvertent variation in peak heights was the primary reason that intra-site samples appeared separated within the plot. This was likely due to disparities in the amount of soil on the ATR crystal, as visual inspection of the raw, unprocessed spectra revealed differences in overall absorbance (Appendix 6.13). Based entirely on the PCA results, the *suspect recovered* sample most closely resembled the *crime scene* sample.

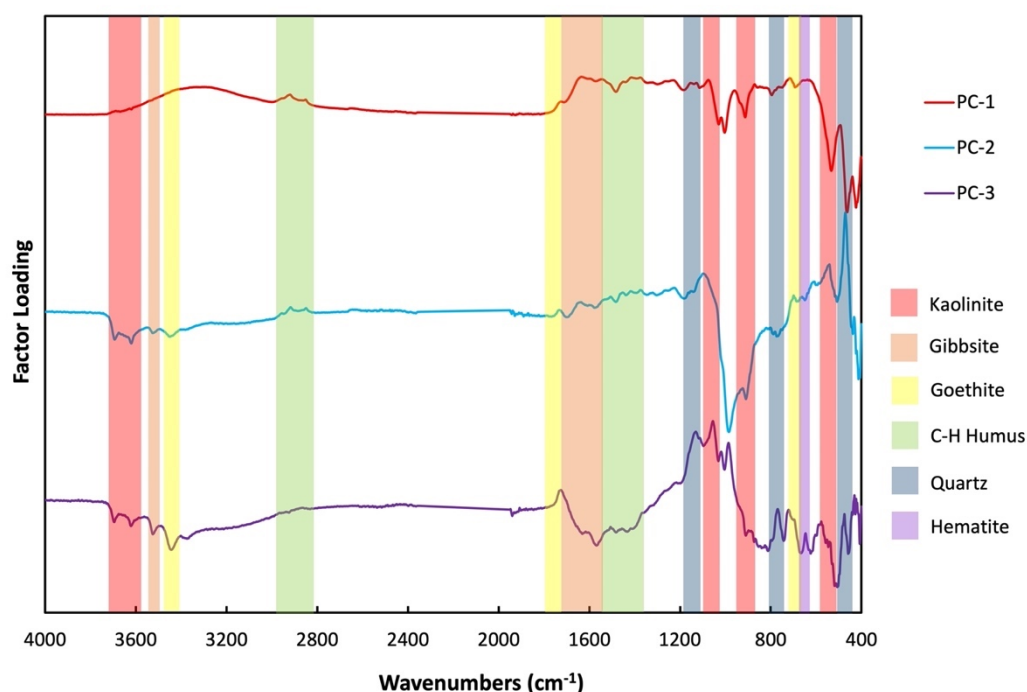


Figure 6.11 Factor loadings plot of PCs 1-3 for PCA of the blinded case simulation ATR-FTIR absorbance dataset, with the main peaks of interest highlighted and annotated with their contributing compounds. Annotations were based off peak assignments outlined in Chapter 4 - Table 4.1.

6.3.3.2 Linear discriminant analysis

LDA was performed on the ATR-FTIR dataset with each location treated as an individual class, returning a calibration accuracy of 100% (Appendix 6.14). Utilising six PCs for LDA was found to maximise discrimination between locations and increase the validation accuracy. Despite the large degree of intra-site sample separation exhibited within the scores plot, all replicates were correctly classified to their source.

This LDA model was then used to predict the source of the *suspect recovered* soil samples. 100% of sample replicates were predicted as originating from the *crime scene* (Table 6.6). The discriminant values (Appendix 6.15) showed a large degree of separation from the other sites, and hence the confidence surrounding these predictions was high. All six *suspect recovered* replicates also showed very similar discriminant values for each site; for each replicate, the *crime scene* soils were predicted as the most similar, followed by *alibi site 2* soils, then *alibi site 1* soils, and lastly soils from the *potential site of interest*. This pattern observed highlighted the increased precision within these results compared to previous models using MSP.

Table 6.6 Predictions for samples in the ATR-FTIR blinded case simulation validation set using a 6-PC LDA model.

Predicted Location	Suspect Recovered Samples
Alibi Site 1	0
Crime Scene	6
Alibi Site 2	0
Potential Site of Interest	0

6.3.3.3 Summary of results

Overall, chemometric interpretation of ATR-FTIR data was very successful for discriminating visually similar soils from different sites, as was expected based on the results of Chapter 4. The use of ATR-FTIR spectra for PCA allowed for the exclusion of *alibi site 1* as the source of the *suspect recovered* soil, and an association was

revealed between the *suspect recovered* soil and the *crime scene* soil. The model did exhibit clustering of the soils based on their individual sub-samples rather than the overall site from which they originated, indicating some variation attributed to sample preparation. This was also evident across the models produced in Chapter 4. Despite this, discrimination of all known soils was achieved.

ATR-FTIR analysis combined with predictive LDA produced a total accuracy of 100%, with a high level of confidence in these predictions. All of the *suspect recovered* replicates also exhibited very similar discriminant values for each site, highlighting the increased precision compared to previous MSP models. This outcome correlated with the results of Chapter 4, which showcased the high accuracy of ATR-FTIR when used for predictive chemometric methods. Therefore, the use of ATR-FTIR post MSP analysis allowed for enhanced discrimination between visually similar soils, by further excluding *alibi site 2* soils and indicating an association between the *suspect recovered* and *crime scene* soils, and helped to increase the confidence surrounding the previous exclusions made by MSP by reinforcing them.

6.3.4 X-ray diffractive analysis of soil samples

As outlined in Section 2.5, XRD patterns were collected from the quartz-recovered fine fractions of the five supplied soils using two different XRD instruments; first, analysis was done by the modern ChemCentre instrumentation with its settings adjusted to best replicate the previous older instrumentation utilised in Chapter 5 (Figure 6.12), followed by analysis of the same samples on the same low background plates by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) instrumentation operating at modern capabilities (Figure 6.13). Six dominant minerals were detected within these patterns – mica, kaolinite, microcline feldspar, gibbsite, quartz, and calcite. The intensities or presence/absence of these reflections accounted for the majority of the variance observed between samples. Closer visual inspection of the patterns also revealed possible smaller contributions from chloritised vermiculite/clinochlore (7.2°), bohmite (16.8° , 32.8°), goethite (24.8° , 43.2°), anatase (29.5°), and aragonite (30.5° – left shoulder of quartz reflection, 38.7° , 53.8°) in some samples, though due to the size of these reflections in

comparison to the level of noise, these were difficult to accurately confirm in the ChemCentre patterns. These smaller mineral reflections were easier to visualise within the CSIRO patterns, primarily due to the improved signal to noise ratio. These were the only visible differences between ChemCentre and CSIRO patterns; all mineral reflections and their relative intensities were comparable between the two sources of analysis.

Patterns from both of the *alibi* sites were dominated by quartz, microcline, and kaolinite, with a noticeable lack of calcite in comparison to the other known sites. The only visual difference between these soils was an increased amount of gibbsite in *alibi site 2's* XRD pattern. The soil from the *potential site of interest* was arguably the most different from the others, containing much larger amounts of mica and kaolinite. XRD patterns from the *crime scene* and the *suspect recovered* soils were visually very similar, both containing higher amounts of gibbsite and calcite than other samples. The presence of limestone was previously indicated within these samples through the visual identification of small, white rocks in Section 6.3.1, which likely accounts for this added calcite. The only visual difference between these soils' patterns was the presence of a small vermiculite/clinochlore reflection in the ChemCentre *crime scene* pattern (7.2°), and a small goethite reflection in the ChemCentre *suspect recovered* pattern (24.8°). However, these differences were not apparent in the CSIRO patterns. Based on visual examination of the XRD patterns alone, the *suspect recovered* soil was most comparable to the *crime scene* soil, indicating that they may have originated from the same location.

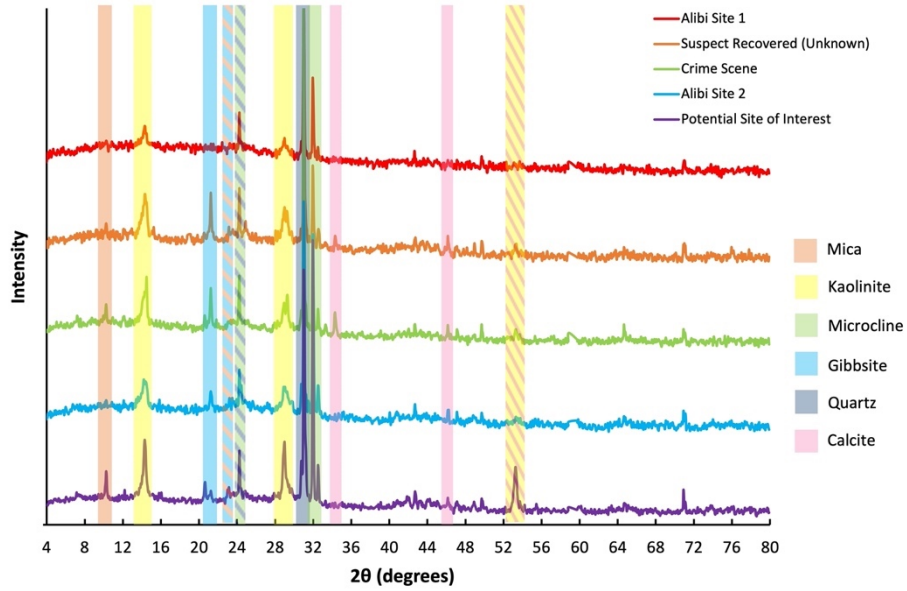


Figure 6.12 Baseline corrected and normalised XRD patterns obtained from ChemCentre, showing the variability in composition of a selection of soil samples collected from differing locations representative of a blinded case simulation. Annotations were based off reflection assignments outlined in Chapter 5 - Table 5.1.

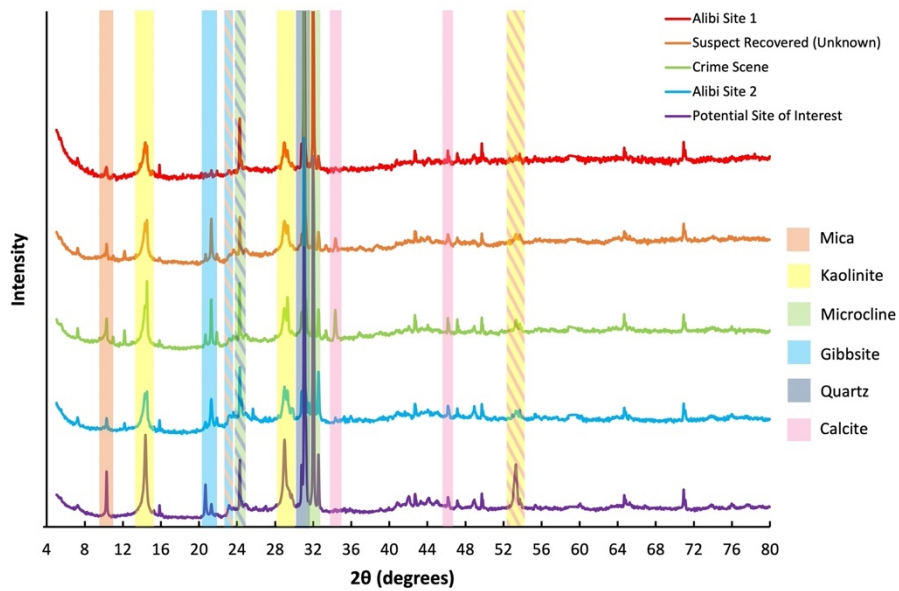


Figure 6.13 Baseline corrected and normalised XRD patterns obtained from CSIRO, showing the variability in composition of a selection of soil samples collected from differing locations representative of a blinded case simulation. Annotations were based off reflection assignments outlined in Chapter 5 - Table 5.1.

6.3.4.1 Principal component analysis utilising ChemCentre XRD patterns

Following the method applied in Section 5.3, PCA performed on the ChemCentre XRD patterns revealed that 92.8% of the total variance in the dataset could be described by the first five PCs (Appendix 6.16). Three-dimensional score plots generated using these PCs (Figure 6.14) showed clear discrimination of all of the known soil sites. The loadings (Figure 6.15) were largely influenced by variation in microcline feldspar and quartz, with contribution also evident from kaolinite and gibbsite. While mica and calcite were also detected within the loadings, their influence was minimal. There was some intra-site separation evident within the *crime scene* and *alibi site 1* soils, however, this was not significant enough to affect the clustering of these groups or cause them to overlap with others. The largest degree of intra-site variation was again seen within the *crime scene* samples, which were most significantly separated along PC-3. Comparing the loadings plot with XRD patterns from these samples (Appendix 6.17), their positioning along PC-3 was influenced based on the height of the quartz (31.0°), microcline (32.0°), and kaolinite (28.9°) reflections.

The *potential site of interest* soils were consistently the best separated from all other sites, presumably due to the substantial differences in underlying mineralogy associated with the original dune system; all of the known soils came from the Spearwood dune system, except for the *potential site of interest* which came from the Pinjarra Plain. The loadings also showed that increasing amounts of noise were detected with each additional PC, indicating that a portion of the separation across PCs 3 and above may have been due to this and hence unreliable. Despite this, the *suspect recovered* samples were clustered closely and overlapping with the *crime scene* samples in the scores plot, indicating that these likely originated from the same location. No matter which combination of PCs were used to visualise the model, the *suspect recovered* samples were consistently clustered closest to the *crime scene* samples.

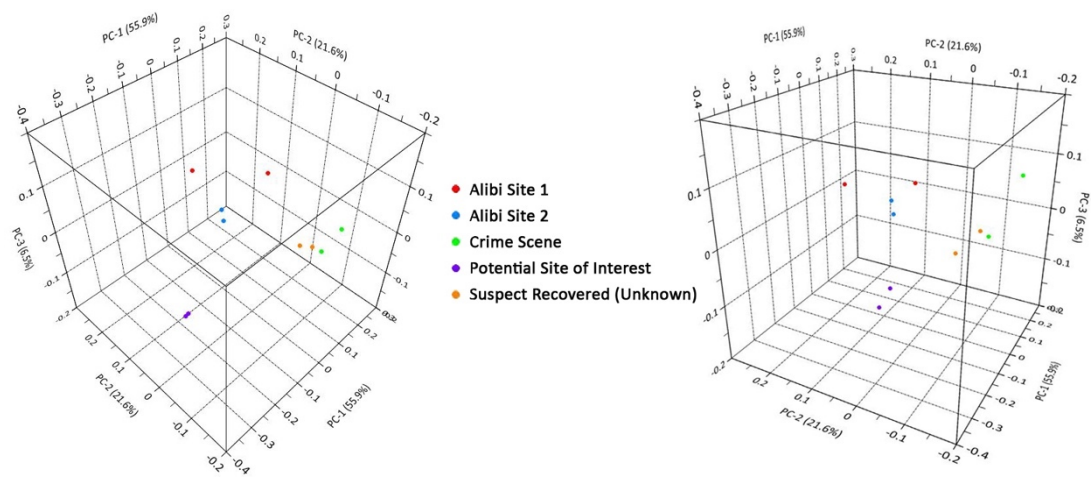


Figure 6.14 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding ChemCentre XRD patterns.

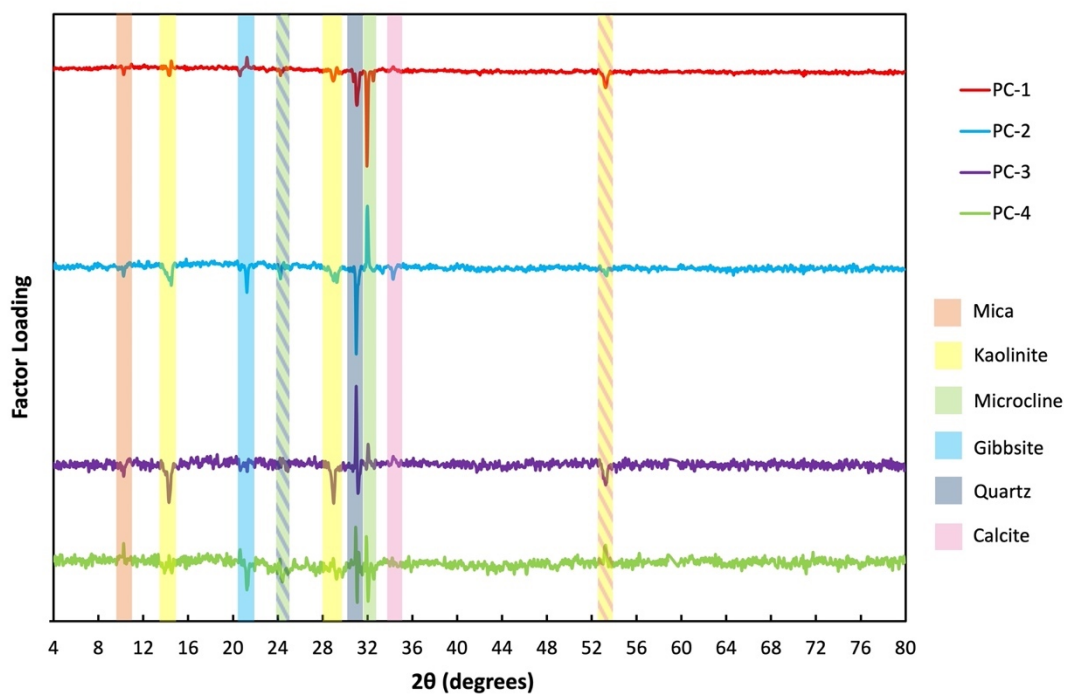


Figure 6.15 Factor loadings plot of PCs 1-4 for PCA of the blinded case simulation ChemCentre XRD dataset, with the main reflections of interest highlighted and annotated with their contributing compounds. Annotations were based off reflection assignments outlined in Chapter 5 - Table 5.1.

6.3.4.2 Principal component analysis utilising CSIRO XRD patterns

PCA performed on the CSIRO XRD patterns revealed that 96.9% of the total variance in the dataset could be described by the first five PCs (Appendix 6.18). Three-dimensional score plots generated using these PCs (Figure 6.16) showed clear discrimination of all of the known soil sites, albeit some intra-site sample separation for the soils from the four known sites. The loadings (Figure 6.17) showed influence mainly from microcline feldspar and quartz, with contribution also evident from kaolinite, gibbsite, and mica. While calcite was also detected within the loadings, its influence was again minimal. In contrast to the ChemCentre XRD model, these CSIRO XRD loadings did not encapsulate any detection of noise, exhibiting relatively smooth spectra. However, as these were the same minerals detected within the ChemCentre loadings, the CSIRO model exhibited an equivalent degree of clustering and differentiation between soils, with just increased separation between clusters due to the improved sensitivity of the instrumentation. The *suspect recovered* samples were again consistently clustered closely with the *crime scene* samples in the scores plot, irrespective of which PCs were used to visualise the model. Once more, this suggested that the *suspect recovered* soils likely originated from the *crime scene*.

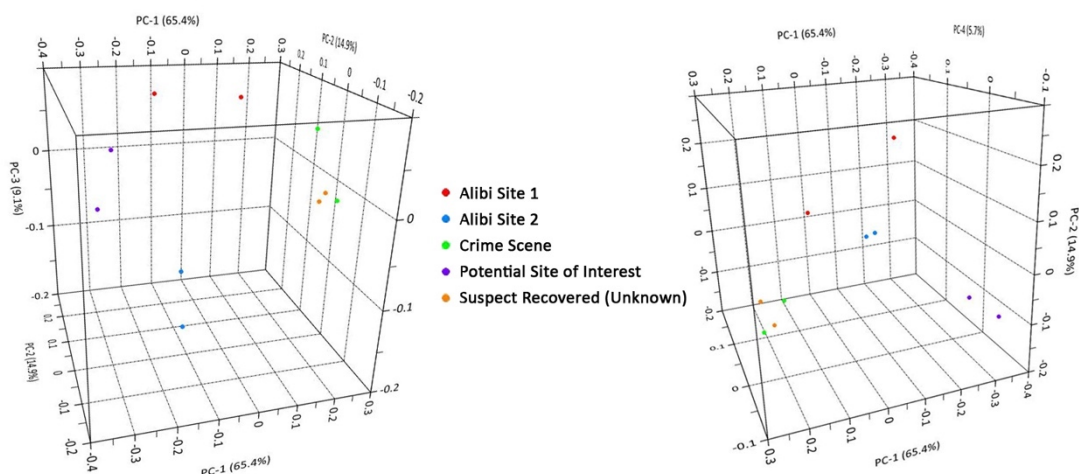


Figure 6.16 3-dimensional PCA scores plots generated using the first four PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding CSIRO XRD patterns.

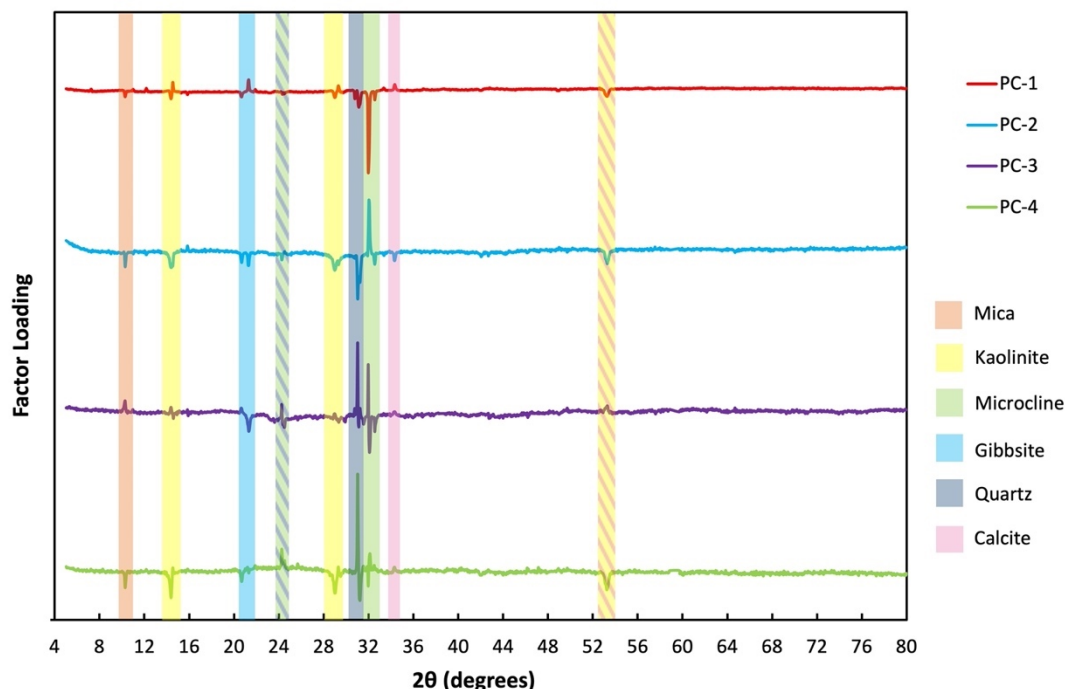


Figure 6.17 Factor loadings plot of PCs 1-4 for PCA of the blinded case simulation CSIRO XRD dataset, with the main reflections of interest highlighted and annotated with their contributing compounds. Annotations were based off reflection assignments outlined in Chapter 5 - Table 5.1.

6.3.4.3 Linear discriminant analysis

LDA was performed on the ChemCentre XRD dataset with each location treated as an individual class, returning a calibration accuracy of 100% (Appendix 6.19). Due to the limited sample size, a maximum of only two PCs could be utilised to build the model. This was a disadvantage as a substantial amount of the variation (22.5%) was captured within PCs 3 and above, as evident in the scree plot. Conversely, this did allow for the majority of the variation attributed to noise to be disregarded. When reduced to a 2-PC model (Appendix 6.20), there was a slight decrease in the degree of separation between known samples, however, this model still allowed for the discrimination of all of the known soils. Despite the intra-site separation exhibited by the *crime scene* and *alibi site 1* samples, all replicates were correctly classified to their source, due to their clear discrimination.

This LDA model was then used to predict the source of the *suspect recovered* soil samples. 50% of sample spectra were predicted as originating from *alibi site 1*, and

the other 50% were predicted as the *crime scene* (Table 6.7). The discriminant values (Appendix 6.21) indicated that while the *crime scene* prediction was of moderately high confidence (+109%), the *alibi site 1* predicted sample was very close to the centroid of the *alibi site 2* soils (+10%). While this implied that the *suspect recovered* soil more than likely originated from the *crime scene*, the limited sample size meant that these results may not have been accurately representative.

Table 6.7 Predictions for samples in the ChemCentre XRD blinded case simulation validation set using a 2-PC LDA model.

Predicted Location	Suspect Recovered Samples
Alibi Site 1	1
Crime Scene	1
Alibi Site 2	0
Potential Site of Interest	0

LDA was then performed on the CSIRO XRD dataset, returning a calibration accuracy of 100% (Appendix 6.22). As seen in the ChemCentre LDA model, there was a slight decrease in the degree of separation between known samples when reduced to a 2-PC model (Appendix 6.23), and the intra-site separation within the *alibi site 1* soil was made more prominent. Despite this, the 2D model still allowed for the discrimination of all of the known soils, and hence, all replicates were correctly classified to their source.

When this LDA model was used to predict the source of the *suspect recovered* soil samples, 100% of sample spectra were predicted as originating from the *crime scene* (Table 6.8). The discriminant values (Appendix 6.24) showed an adequate degree of separation from the other sites, and hence the confidence surrounding these predictions was moderately high. Both *suspect recovered* samples also showed similar discriminant values for each site; the *crime scene* soils were predicted as the most similar, followed by *alibi site 1* soils, then *alibi site 2* soils, and lastly soils from the *potential site of interest*. This pattern observed highlighted the increased precision within these results compared to the previous ChemCentre XRD model,

which showed substantially different discriminant values between samples for the three most similar sites.

Table 6.8 Predictions for samples in the CSIRO XRD blinded case simulation validation set using a 2-PC LDA model.

Predicted Location	Suspect Recovered Samples
Alibi Site 1	0
Crime Scene	2
Alibi Site 2	0
Potential Site of Interest	0

6.3.4.4 Summary of results

Overall, the use of XRD and chemometrics was demonstrated to be valuable for the discrimination of soils in forensic casework, particularly when utilised in a sequence with other spectroscopic techniques. PCA using both simulated 'older' ChemCentre data and 'newer' CSIRO data allowed for the association between the *suspect recovered* soil and the *crime scene* soil to be confirmed, with discrimination of all other sites. The CSIRO model exhibited an enhanced level of separation between sites due to the improved sensitivity and signal-to-noise ratio of the modern instrumentation, as demonstrated in Chapter 5. This reinforced the notion that the same conclusions can be inferred from results produced through the use of both outdated and modern instrumentation.

XRD analysis combined with predictive LDA produced a total accuracy of 75%; simulated 'older' ChemCentre instrumentation resulted in 50% correct, while modern CSIRO instrumentation achieved 100% correct. Additionally, the confidence surrounding the correct predictions using XRD was relatively high in comparison to the predictions that were incorrect, increasing the confidence in the correct results. These models displayed improved accuracies in comparison to the LDA model presented in Chapter 5, in which ChemCentre and CSIRO data were combined into one dataset for LDA. By keeping the two datasets separate for this case simulation,

each model no longer encapsulated variation attributed to instrumental analysis, which improved the results. Both *suspect recovered* samples within the CSIRO model also showed similar discriminant values for each site, highlighting the increased precision compared to the simulated 'older' ChemCentre model. It was previously demonstrated in Chapter 5 that predictive LDA was more affected by the instrument capabilities than unsupervised methods of analysis, and this was reflected in these results too, highlighting the advantage that more advanced instrumentation offers. In this case simulation however, LDA was performed alongside PCA and the results were interpreted together with those obtained through the use of other techniques, so the LDA data was primarily used to confirm previous associations rather than introduce new information. For use within forensic casework, XRD was shown to perform well when utilised as the final step of an analysis sequence; it was able to confirm the exclusions and associations previously indicated through MSP and ATR-FTIR, therefore providing an increased level of confidence surrounding the results.

6.3.5 Comparison of results with true source

As previously indicated, the *suspect recovered* soil was revealed to have originated from the *crime scene* site post-analysis and interpretation. Therefore, the sequence utilised within this case simulation was able to correctly predict the source of the *suspect recovered* soil from analysis of its quartz-recovered fine fraction. Additionally, due to the discrimination of different known soils at each successive stage, 100% of samples were able to be differentiated from the *suspect recovered* and *crime scene* soils by the end of the sequence. While the results achieved by each individual technique correlate with those obtained in previous chapters, i.e., ATR-FTIR methods were most accurate, followed by XRD, and lastly MSP, the LDA accuracies achieved within this blinded case simulation were consistently higher across all techniques. Whilst this is generally the case when utilising smaller sample populations, the level of similarity between these samples was much higher than in previous chapters, which was expected to increase the difficulty of their discrimination.

Therefore, this sequence has been shown to be particularly useful in forensic casework for differentiating soils that are similar in appearance, but contain subtle differences in composition based on their location. It has illustrated the value in full compositional analysis of soils, by utilising several methods in combination to ensure all forms of variation across the samples can be accurately identified and exploited for their discrimination. Additionally, this allowed for the overall effects of sample preparation, instrumental variation, and anthropogenic influences to be better recognised and minimised accordingly. The integration of chemometric analysis within the sequence also provided strong numerical support to back up the conclusions reached; in some instances, chemical analysis without chemometrics suggested the same results, but these were able to be reinforced using a documented statistical backing. This could assist with presentation of this evidence in court by providing a quantitative basis for decision-making, and addressing the concerns of subjectivity often highlighted in feature comparison disciplines. This approach could be readily integrated into existing analysis workflows. To reinforce confidence in the conclusions even further, forensic practitioners can combine this sequence with other techniques that collect data on the soils in question, such as ICP-MS for elemental analysis, detection of heavy minerals, or pollen / microbiome analysis methods, or results from analysis of other forms of trace evidence found within the samples, such as glass.

6.4 Conclusions

In this chapter, the use of multiple spectroscopic techniques with chemometrics combined into a sequence successfully proved an association between samples in a blinded case simulation. MSP, ATR-FTIR spectroscopy, and XRD were used to characterise the quartz-recovered fine fraction of Swan Coastal Plain soils. Applying chemometric methods (PCA and LDA) to data obtained from analysis of the quartz-recovered fine fraction was successful at discriminating between soils from different known locations, and predicting the source of an unknown sample. This not only highlighted the value of the technique for application to sandy forensic soils, but also

the strength of chemometrics for objectively identifying associations where only subtle differences occur.

Overall, the use of ATR-FTIR spectra for chemometrics was the most successful method, due to the high prediction accuracy and high confidence of the results. Presumably, this is due to the ability of IR techniques to detect both inorganic and organic compounds within the soil samples, allowing for the representation of a higher proportion of variation between samples. If forensic experts are limited by time restrictions, ATR-FTIR is suggested for the efficient comparison of soil samples. However, if possible, it is recommended to utilise all available methods within a sequence (including visual examination), in order to increase the knowledge gained and heighten the confidence in results. From this study, it is clear that different methods gave varied results, but when taken holistically, these results provided complementary information. For example, the soil from *alibi site 1* could visually be discriminated from the *suspect recovered* soil, however, all further incorrect LDA predictions were assigned as *alibi site 1*, indicating that they were likely to be incorrect. Therefore, this study not only highlighted that the value of the method was increased through use of the whole sequence, but also that performing chemometrics alongside each of these steps allowed for greater confidence in the end result.

Additional points of differentiation could be utilised to enhance the discrimination of soil by conducting LDA based on other attributes of the known sites, such as the original dune system or type of location, however, in this instance these were consistent between most sites. This was intentional for two reasons; to test the full capabilities of the developed method for discriminating between soils that are very similar, and to imitate as closely as possible the likely types of samples provided in a real forensic examination. It was also indicated that all techniques within the sequence were able to detect variation across the soils that was related to different features of the locations involved. For example, MSP was able to discriminate soils based on colour, and both ATR-FTIR and XRD analysis could detect compounds within the soil samples in quantitative ratios that were characteristic of their dune system.

While this suggests potential for application of this method to a database matching approach, it is understood that the quartz-recovered fine fraction cannot provide the required amount of holistic information on the soil samples for this to be implemented with confidence. Hence, this sequence is more appropriate for the differentiation of samples, and is primarily intended to be incorporated within standard questioned versus known comparisons. In this instance, it is important to ensure the results generated through chemometric analyses are verified and further interpreted by the forensic examiner, to ensure that they are supported by the numerical data and not just suggested due to 'best fit'. The number of replicates used within this study was also limited due to sample size restrictions commonly encountered in casework; the inclusion of additional replicates would be expected to increase the accuracy and precision of results.

Chapter 7. Conclusions and Future Work

Portions of this chapter have been published in the following articles:

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics for the forensic analysis of Western Australian sandy soils." *Forensic Chemistry*, 2022. 28: 100412.

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics: Application to simulated forensic soil casework." *Forensic Chemistry*, 2023. 33: 100481.

7.1 Conclusions

The aim of this thesis was to develop a multi-faceted approach to the analysis of the quartz-recovered fine fraction of soils that utilises chemometrics to demonstrate objective characterisation and differentiation of arid, sandy soil samples for forensic purposes. Analysis of these soils is especially challenging due to their very low levels of clay and organic matter, which is further compounded by the trace quantities of soil commonly encountered in forensic casework. Common soil analysis techniques are unable to be utilised with these quartz-dominated soils, and novel methods focused on the detection of organics within the sample do not allow for their discrimination (94). The majority of the variation within these sandy soils is contained within the clay fine fraction, typically found as thin coatings on the surfaces of the quartz grains (94). This fraction is estimated to make up only 1 – 5% of the soil mass, and when encountered in trace quantities as forensic evidence, translates to sample sizes of approximately 1 – 2 mg (94). It is therefore important to have reliable, validated methods for the isolation and analysis of this quartz-recovered fine fraction, as well as for the interpretation and communication of the results generated. This will ensure these types of soils can be used as a form of forensic evidence, and the value of this evidence is not overlooked.

A method for isolation of the quartz-recovered fine fraction was previously developed by Pitts and Clarke (94) and applied to soil samples throughout this thesis, which is an expansion of this earlier work. The chemical characteristics of the soils were measured using a variety of spectroscopic techniques and the collected data was analysed using chemometric approaches, which allowed for interpretation of the variation and subsequent discrimination between them. The multi-variate statistical methods explored in this study not only provide a more objective interpretation of the examination of forensic soils, of which the significance has been highlighted many times within the forensic community, but also offer a statistical foundation to determine the limits of performance (99, 101). On the basis of this, an analytical sequence combining spectroscopy and chemometrics was developed for application to forensic casework and this was subsequently tested by undertaking a blinded case

simulation. Additionally, by utilising a particularly challenging set of samples, detailed data has been collected on forensic soil as a form of forensic trace evidence in Western Australia, underpinning the interpretation of findings that might come from future analysis (23). In a broader forensic context, these methods developed are applicable to other national and international jurisdictions that experience arid, sandy soils, or soils where the bulk material is dominated by quartz.

7.1.1 Investigations into spectroscopic techniques in combination with chemometrics

Chapters 3 to 5 investigated the chemical characterisation and discrimination of a selection of sandy soils from the Swan Coastal Plain in Perth, Western Australia. Samples were collected from a range of differing location types, and dune and plant systems, representative of the whole plain. The quartz-recovered fine fractions were extracted from these soils and analysed using several complementary spectroscopic techniques, in combination with chemometric methods.

Chapter 3 demonstrated the use of microspectrophotometry (MSP) for the analysis of the quartz-recovered fine fraction of soils, which had previously never been explored. When combined with chemometric interpretation however, this technique provided minimal discrimination between soil samples from different locations, as the noise within the reflectance spectra was accentuated by the complex statistical analyses. Soils that were distinctly coloured were able to be differentiated from the rest of the population with confidence, but only a few soils that were visually similar achieved the same result; MSP spectra essentially provided a more precise and objective determination of the colour of the soil sample. However, this colour was not always representative of the bulk colour of the soil and was affected by inconsistency in the thickness of the samples. Therefore, the increased precision became a disadvantage when paired with the high degree of variability that exists within soil, producing chemometric models that were only sometimes able to differentiate soils based on their original locations. When MSP spectra were converted to $L^* a^* b^*$ colorimetric values, detailed information was lost and the method was unable to provide any beneficial separation over visual examination of

the soil, so is not favourable as an interpretation method in this instance. Nevertheless, MSP analysis combined with chemometrics is still valuable for forensic pairwise comparisons to detect subtle differences in colour not visible to the naked eye, and to quickly rule out samples that are highly coloured and appear visually distinctive using a statistical measure of differentiation.

In Chapter 4, vibrational spectroscopy was utilised for analysis of the quartz-recovered fine fraction obtained from soil samples. Raman spectroscopy has recently been applied to soil data, however, showed poor results due to strong levels of fluorescence that swamped the spectra (77, 121-123). This thesis aimed to expand on that work by utilising the quartz-recovered fine fraction of soils and different instrumental conditions in an attempt to reduce fluorescence and improve results. Raman spectroscopy was unsuccessful at providing any chemical information on the samples due to aforementioned fluorescence and attempts at background correction showed no beneficial effect on the spectra, so no further studies were carried out using this technique.

Attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy has never before been applied to the quartz-recovered fine fraction of sandy soils. ATR-FTIR spectroscopy successfully detected many inorganic and organic components within the soils that allowed for their differentiation. Chemometric methods applied to this data resulted in the discrimination of all soil samples based on their original locations, and the classification of nearly all samples (93%) to their correct locations. All soils did show varying degrees of intra-location separation based on individual samples rather than overall location, indicating potential variation due to sampling, however, this was generally not significant enough to alter the discrimination achieved. It was also apparent that the compounds responsible for variance within the ATR-FTIR chemometric models were associated with the visual colour of the soils. The combined use of ATR-FTIR with chemometrics was the best performing method for detecting the subtle variation within quartz-recovered fine fractions and showed great potential for differentiating visually similar sandy soils based on their locations.

Chapter 5 demonstrated the use of combined X-ray diffraction (XRD) and chemometrics for the analysis of the quartz-recovered fine fraction of soils, and also incorporated a reproducibility study of results obtained through analysis conducted on three different XRD instruments. Previous work by Pitts and Clarke demonstrated the successful use of XRD on quartz-recovered fine fractions of soil, however, this only made use of selective percentage intensities and lacked any chemometric interpretation of results (94). Whilst there are several other published studies that do use XRD soil data with chemometrics, these did not utilise the quartz-recovered fine fraction and also involved 'peak picking' or profile fitting prior to multivariate analysis (19, 20, 72). This thesis made use of the whole XRD pattern for chemometric analysis, to capture any minor variations in mineral content or shifts away from the 'standard' reflection positions.

Results showed the discrimination of most soil samples based on their locations, however, some sites showed significant intra-site variability potentially due to anthropogenic influence. Predictive models generated were not accurate at classifying soils to their original locations despite the degree of discrimination achieved, due to limitations on the number of samples available to build these models. The variation detected in the mineral content of the soil correlated with the dune systems that these samples originated from, which may indicate potential for development as a predictive screening tool that could allow detectives to focus on areas of interest. Most importantly, the XRD models were able to better discriminate different soils to those that were easily discriminated using the ATR-FTIR model, demonstrating how the soil examination process would benefit from a multi-model approach.

The reproducibility study made use of multiple XRD instruments to analyse the same collection of soil samples; a three decades-old instrument, a newer instrument operating at more modern capabilities, and another modern instrument with settings adjusted to closely mimic those used on the older instrumentation. It is recognised that these diffractometers are expensive capital items, and some forensic laboratories may only have access to older instrumentation. Results showed

variations in the quality of data collected that did affect the results, however, all instrumentation achieved similar overall discrimination of soils and predictive accuracies. This highlighted the robustness of the methods used. Ideally, the analysis of all samples should be conducted using the same instrumentation where possible for best results, however, it is possible to use multiple different instruments if the variations in quality are accounted for when interpreting the results.

The approach developed provides good evidence that selected spectroscopic methods, utilised in combination, and with chemometrics performed alongside each stage of the sequence, can maximise the differentiation observed by forensic practitioners when examining soil samples (Figure 7.1). In addition, examination of loadings plots provided key information on the chemical characteristics that allowed for the differentiation between samples. Sequenced analysis can be especially useful when samples are relatively similar and available instrumentation is limited, as different techniques can inform in complementary ways when paired with chemometric analysis. In this thesis, using several non-destructive techniques in sequence allowed for more accurate identification of minerals, and enhanced discrimination of soils from similar locations. Chemometrics performed on MSP data initially discriminated distinctly coloured soils from the rest of the population. ATR-FTIR spectroscopy with chemometrics allowed for the discrimination and classification of the majority of the remaining soils. Finally, chemometrics performed on XRD data was able to confirm the mineral associations made through the ATR-FTIR loadings and enhance the discrimination of some samples that were not well-separated within the previous ATR-FTIR model. These methods each provided information that built on knowledge obtained from the previous technique. Taken in isolation, each analysis method provided only part of the entire picture. The ATR-FTIR and XRD methods also provided additional forensic intelligence on the visual appearances and dune systems associated with the soil samples being analysed, which could prove especially useful for soil provenancing.

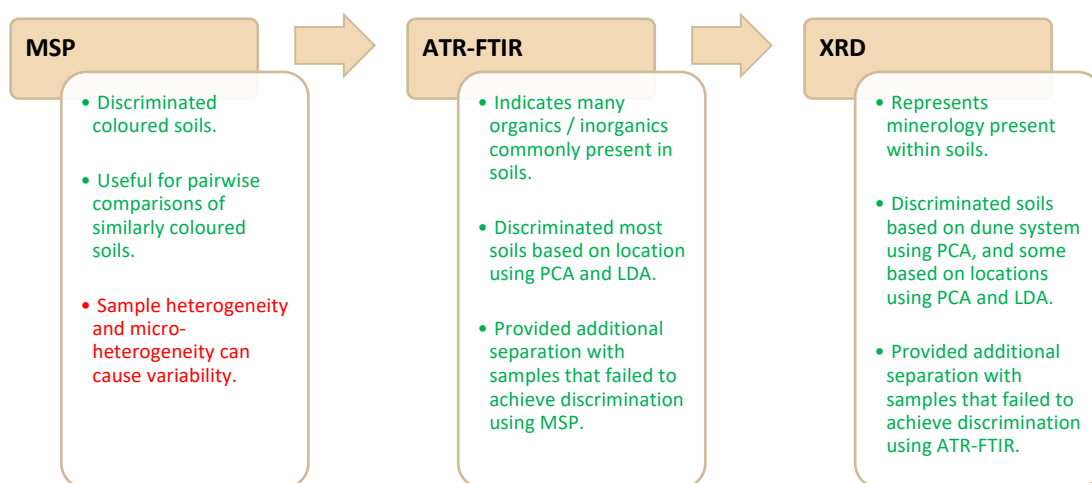


Figure 7.1 Analytical sequence for examination of the quartz-recovered fine fraction of sandy forensic soil samples from the Swan Coastal Plain in Perth, Western Australia, with the information obtained through analysis at each stage.

7.1.2 Application to blinded case simulation and forensic casework

Chapter 6 applied the proposed analytical sequence developed in the previous chapters to a blinded case simulation to evaluate its capabilities and showcase its suitability for use within the context of forensic casework. This simulation was generated by a third-party for objectivity and was designed to be challenging by utilising five soils that were very similar, imitating as closely as possible the likely types of samples provided in a real forensic examination. All of the soils provided were similar in colour, with only small differences in their bulk appearance, and most were collected from locations that were situated close by to each other (within the same dune system) and subject to the same external factors (location type and usage). The aim was to determine any associations for a *suspect recovered* sample by either associating it with or excluding it from any or all of the four known locations, represented by their own reference soil samples; *alibi site 1*, *alibi site 2*, *crime scene*, or *potential site of interest*. The source of the *suspect recovered* soil was the *crime scene* site, however, this was unknown to this author until all analysis and interpretation had been completed.

The results supported the conclusions made in previous chapters; different analysis methods gave varied results, but when taken holistically, these results provided complementary information that was able to be used collectively to discriminate between the known soils and identify the source of the *suspect recovered* soil with greater confidence (Figure 7.2). The use of MSP with chemometric methods was able to discriminate the *potential site of interest* and *alibi site 2* soils from the rest of the population, tentatively excluding these sites as origins for the *suspect recovered* soil. Chemometric methods performed on ATR-FTIR data showed a strong association between the *suspect recovered* soil and the *crime scene* soil, confirmed the exclusions made by MSP, and also discriminated *alibi site 1* from both of these sites. Finally, the use of XRD and chemometrics confirmed all these results by again showing an association between the *suspect recovered* soil and the *crime scene* soil and discriminating all other soils from these sites. Using multi-variate statistical methods to interpret the data collected by each analysis technique was beneficial for providing an objective, statistical measure of the similarities and/or variation between samples.

This approach can be readily incorporated into forensic casework to provide discrimination between similar soil samples in an objective manner, with the conclusions reinforced by a scientific measure of similarity that can be presented as weighting for evidence in court. While this method does show some potential for use within a database matching approach, it is primarily intended to be incorporated within standard questioned versus known comparisons. This is because the quartz-recovered fine fraction cannot provide the holistic information on the soil samples that would be required to construct a database, and hence is more appropriate for differentiation between samples. However, it was indicated that both ATR-FTIR and XRD analysis can detect compounds within the soil samples in quantitative ratios that are characteristic of their dune system, and therefore could potentially provide information to forensic investigators concerning general regions that a soil sample may have originated from. Caution should surround this approach as soils are a constantly changing form of evidence, so their use for forensic intelligence is always going to be challenging. It is important to therefore understand the limitations to this

prior to use and communicate these appropriately when reporting the findings to end users of the information.

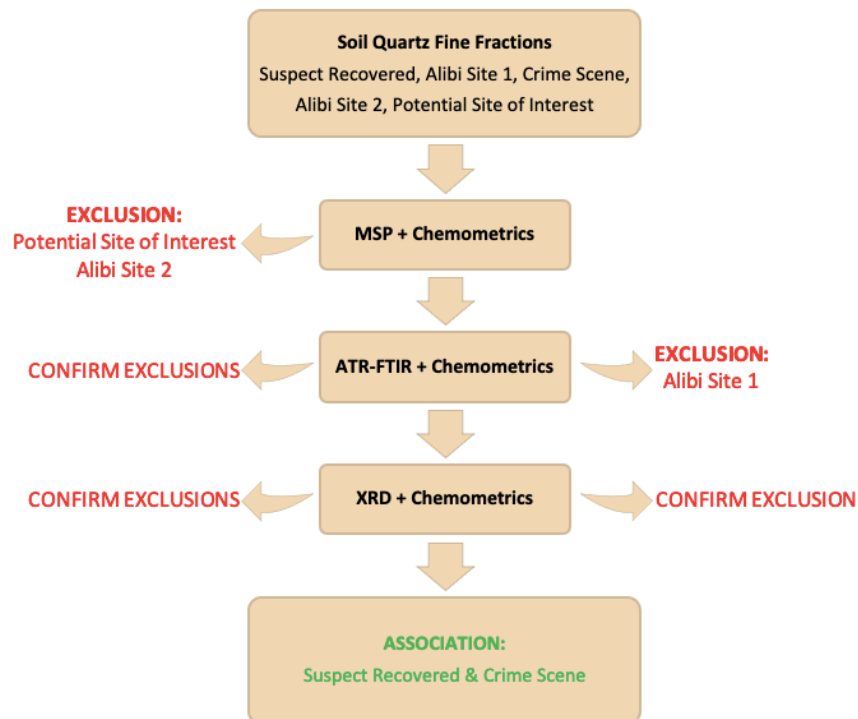


Figure 7.2 Flowchart illustrating how each stage of analysis within the sequence allowed for an association of exclusion to be made between the suspect recovered soil and the reference soils in the blind case simulation in Chapter 6.

7.2 Future work

There is a need for further research in this space to reinforce the proposition that soil analysis can provide extremely useful information in the context of a forensic investigation, even when soil samples are sandy, minute, and fractionated. Whilst the dataset for this study was limited by the number of samples per location, the intention was to imitate sample size requirements from a forensic case work scenario as closely as possible. In these situations, there is usually minimal sample available for analysis. However, validation of this approach should be explored further by utilising a larger sample population and a more complex range of soils. Investigations should also be expanded into sub-surface soil samples, to assess the suitability of the technique for processing burial sites; sub-surface samples were initially collected

along with the surface samples used throughout this study but have not been utilised for this purpose as of yet. The database matching capabilities of the approach could also be explored by combining the method development and case simulation datasets into one singular chemometric model, to assess how well the approach can provide an association within a more complex dataset. Finally, consideration should be given as to how to effectively present these results in court. As previously shown with other forms of forensic evidence, e.g., fibres, this chemometric approach is suitable for use with other statistical interpretation methods, such as Bayesian interpretation (100, 107, 108). Future research should examine how chemometric methods can be integrated within these other statistical frameworks, in order to gain acceptance in the legal system.

7.3 Summary

In summary, this thesis demonstrates the value that arid, sandy soils hold as a form of trace evidence that can be utilised as a part of forensic investigations. Examination of the quartz-recovered fine fraction of Perth soils has demonstrated that chemometrics can be used successfully in combination with spectroscopic techniques to objectively discriminate between these challenging soils. Unlike other forms of forensic trace evidence, soil is in a constant state of change all around the world and therefore requires many different approaches for its analysis dependent on the type of soil present. With so many distinctly different soils found internationally, and many of these showcasing extreme characteristics, it is important to make sure that a range of relevant, validated methods are available for selection of an approach that is most appropriate for the soil type encountered (54). This thesis provides a reliable method for obtaining as much discriminatory information as possible out of dry, quartz-dominated soils. This research could assist in allowing the introduction of soil analysis back into forensic laboratories and improve the capability of those already undertaking it, and enhance its reputation in the courts as a reliable form of trace evidence (38, 62). The results generated may also allow soil evidence to be utilised in the midst of criminal investigations to provide

investigative information to police, rather than just as a reconstructive tool in court, assisting them in locating burial sites or other geological areas of interest.

References

1. Petraco N, Kubic TA, Petraco ND. Case studies in forensic soil examinations. *Forensic Science International*. 2008; 178: p. e23-7.
2. Demaneche S, Schauser L, Dawson L, Franqueville L, Simonet P. Microbial soil community analyses for forensic science: Application to a blind test. *Forensic Science International*. 2017; 270: p. 153-8.
3. Horswell J, Cordiner SJ, Maas EW, Martin TM, Sutherland KBW, Speir TW, et al. Forensic Comparison of Soils by Bacterial Community DNA Profiling. *Journal of Forensic Sciences*. 2002; 47(2): p. 350-3.
4. Woods B, Lennard C, Kirkbride KP, Robertson J. Soil examination for a forensic trace evidence laboratory - Part 1: Spectroscopic techniques. *Forensic Science International*. 2014; 245: p. 187-94.
5. Testoni SA, Melo VF, Dawson LA, Salvador FAdS, Kunii PA. Validation of a Standard Operating Procedure (SOP) for Forensic Soils Investigation in Brazil. *Rev Bras Cienc Solo*. 2019; 43(e0190010).
6. Lee CS, Sung TM, Kim HS, Jeon CH. Classification of forensic soil evidences by application of THM-PyGC/MS and multivariate analysis. *Journal of Analytical and Applied Pyrolysis*. 2012; 96: p. 33-42.
7. Ruffell A, Wiltshire P. Conjunctive use of quantitative and qualitative X-ray diffraction analysis of soils and rocks for forensic analysis. *Forensic Science International*. 2004; 145(1): p. 13-23.
8. Carvalho A, Ribeiro H, Mayes R, Guedes A, Abreu I, Noronha F, et al. Organic matter characterization of sediments in two river beaches from northern Portugal for forensic application. *Forensic Science International*. 2013; 233(1-3): p. 403-15.

9. McCulloch G, Dawson LA, Brewer MJ, Morgan RM. The identification of markers for Geoforensic HPLC profiling at close proximity sites. *Forensic Science International*. 2017; 272: p. 127-41.
10. Reidy L, Bu K, Godfrey M, Cizdziel JV. Elemental fingerprinting of soils using ICP-MS and multivariate statistics: a study for and by forensic chemistry majors. *Forensic Science International*. 2013; 233(1-3): p. 37-44.
11. Jantzi SC, Almirall JR. Characterization and forensic analysis of soil samples using laser-induced breakdown spectroscopy (LIBS). *Anal Bioanal Chem*. 2011; 400: p. 3341–51.
12. Habtom H, Demaneche S, Dawson L, Azulay C, Matan O, Robe P, et al. Soil characterisation by bacterial community analysis for forensic applications: A quantitative comparison of environmental technologies. *Forensic Science International: Genetics*. 2017; 26: p. 21-9.
13. Ritz K, Dawson L, Miller D, editors. *Criminal and Environmental Soil Forensics*. Netherlands: Springer Science & Business Media; 2009.
14. de Caritat P, Woods B, Simpson T, Nichols C, Hoogenboom L, Ilheo A, et al. Forensic soil provenancing in an urban/suburban setting: A sequential multivariate approach. *Journal of Forensic Sciences*. 2021; 66(5): p. 1679-96.
15. Guedes A, Ribeiro H, Valentim B, Rodrigues A, Sant'Ovaia H, Abreu I, et al. Characterization of soils from the Algarve region (Portugal): A multidisciplinary approach for forensic applications. *Science & Justice*. 2011; 51: p. 77-82.
16. Jantzi SC, Almirall JR. Elemental Analysis of Soils Using Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) and Laser-Induced Breakdown Spectroscopy (LIBS) with Multivariate Discrimination: Tape Mounting as an Alternative to Pellets for Small Forensic Transfer Specimens. *Journal of Applied Spectroscopy*. 2014; 68(9): p. 963-74.

17. Dawson LA, Hillier S. Measurement of soil characteristics for forensic applications. *Surface and Interface Analysis*. 2010; 42: p. 363–77.
18. Fløjgaard C, Frøslev TG, Brunbjerg AK, Bruun HH, Moeslund J, Hansen AJ, et al. Predicting provenance of forensic soil samples: Linking soil to ecological habitats by metabarcoding and supervised classification. *PLOS ONE*. 2019; 14(7): p. e0202844.
19. Melo VF, Testoni SA, Dawson L, de Lara AG, Salvador FAdS. Can analysis of a small clod of soil help to solve a murder case? *Science & Justice*. 2019; 59: p. 667-77.
20. Corrêa RS, Melo VF, Abreu GGF, Sousa MH, Chaker JA, Gomes JA. Soil forensics: How far can soil clay analysis distinguish between soil vestiges? *Science & Justice*. 2018; 58: p. 138–44.
21. Cengiz S, Karaca AC, Cakir I, Bulent Uner H, Sevindik A. SEM-EDS analysis and discrimination of forensic soil. *Forensic Science International*. 2004; 141(1): p. 33-7.
22. Profumo A, Gorrioni A, Guarnieri SA, Mellerio GG, Cucca L, Merli D. GC-MS qualitative analysis of the volatile, semivolatile and volatilizable fractions of soil evidence for forensic application: A chemical fingerprinting. *Talanta*. 2020; 219(121304).
23. National Institute of Forensic Science Australia New Zealand. Research and Innovation Roadmap. Australia & New Zealand: Australia New Zealand Policing Advisory Agency (ANZPAA); 2020.
24. Fitzpatrick RW. Soil: Forensic Analysis. In: Jamieson A, Moenssens A, editors. *Wiley Encyclopedia of Forensic Science*. United Kingdom: John Wiley & Sons, Ltd.; 2009. p. 2377-88.

25. Powell R, Bronswijk Wv, Coumbaros J. Enhancing the evidential value of textile fibres. Part 1: Development of a spectral database and evaluative comparison strategy. *Forensic Science International*. 2018; 287: p. 54-62.
26. Robertson JR, Roux C, Wiggins K, Grieve M. Interpretation of Fibres Evidence. *Forensic Examination of Fibres*. 2nd ed: Hoboken: Taylor and Francis; 2002.
27. Madureira-Carvalho Á, Ribeiro H, Newman G, Brewer MJ, Guedes A, Abreu I, et al. Geochemical analysis of sediment samples for forensic purposes: characterisation of two river beaches from the Douro River, Portugal. *Australian Journal of Forensic Sciences*. 2020; 52(2): p. 222–34.
28. Nanzyo M, Kanno H. *Inorganic Constituents in Soil: Basics and Visuals*. 1st ed: Springer Open; 2018.
29. Kammrath BW, Koutrakos A, Castillo J, Langley C, Huck-Jones D. Morphologically-directed Raman spectroscopy for forensic soil analysis. *Forensic Science International*. 2018; 285: p. e25–33.
30. Sharma V, Chauhan R, Kumar R. Spectral characteristics of organic soil matter: A comprehensive review. *Microchemical Journal*. 2021; 171(106836).
31. Mitchell RJ, Keith AM, Potts JM, Ross J, Reid E, Dawson LA. Overstory and understory vegetation interact to alter soil community composition and activity. *Plant Soil*. 2012; 352: p. 65-84.
32. Bommarito CR, Sturdevant AB, Szymanski DW. Analysis of forensic soil samples via high-performance liquid chromatography and ion chromatography. *Journal of Forensic Sciences*. 2007; 52(1): p. 24-30.
33. Cox RJ, Peterson HL, Young J, Cusik C, Espinoza EO. The forensic analysis of soil organic by FTIR. *Forensic Science International*. 2000; 108: p. 107-16.
34. Wald C. Forensic science - The soil sleuth. *Nature*. 2015: p. 422-4.

35. Renella G, Ogunseitan O, Giagnoni L, Arenella M. Environmental proteomics: A long march in the pedosphere. *Soil Biology and Biochemistry*. 2014; 69: p. 34-7.
36. Bastian LV. Residual soil mineralogy and dune subdivision, Swan Coastal Plain, Western Australia. *Australian Journal of Earth Sciences*. 1996; 43(1): p. 31-44.
37. Aitkenhead MJ, Coull MC, Dawson LA. Predicting Sample Source Location from Soil Analysis Using Neural Networks. *Environmental Forensics*. 2014; 15: p. 281–92.
38. Woods B, Lennard C, Kirkbride KP, Robertson J. Soil examination for a forensic trace evidence laboratory - Part 3: A proposed protocol for the effective triage and management of soil examinations. *Forensic Science International*. 2016; 262: p. 46-55.
39. Bull PA, Parker A, Morgan RM. The forensic analysis of soils and sediment taken from the cast of a footprint. *Forensic Science International*. 2006; 162(1-3): p. 6-12.
40. Pasternak Z, Luchibia AO, Matan O, Dawson L, Gafny R, Shpitzen M, et al. Mitigating temporal mismatches in forensic soil microbial profiles. *Australian Journal of Forensic Sciences*. 2019; 51(6): p. 685-94.
41. Turner BL, Laliberté E. Soil Development and Nutrient Availability Along a 2 Million-Year Coastal Dune Chronosequence Under Species-Rich Mediterranean Shrubland in Southwestern Australia. *Ecosystems*. 2015; 18(2): p. 287-309.
42. McArthur WM, Bartie GA. Landforms and Soils as an Aid to Urban Planning in the Perth Metropolitan Northwest Corridor, Western Australia. *Land Resources Management Series*. 1980; 5: p. 1-14.
43. McArthur WM, Bettenay E. The development and distribution of the soils of the Swan Coastal Plain, Western Australia. Commonwealth Scientific and Industrial Research Organisation (CSIRO); 1974.

44. Smolinski H, Scholz GG. Soil assessment of the west Gingin area. Western Australia: Department of Primary Industries and Regional Development; 1997.
45. Western Australian Planning Commission. Perth Coastal Planning Strategy. Perth, Western Australia: Department of Planning; 2008.
46. Pitts KM, Lewis SW, Newland TG. Geochemistry | Soil and Mineralogical Analysis. In: Worsfold P, Poole C, Townshend A, Miró M, editors. Encyclopedia of Analytical Science. 3rd ed: Oxford: Academic Press; 2019. p. 292-301.
47. Semeniuk V, Cresswell ID, Wurm PAS. The Quindalup Dunes: The regional system, physical framework and vegetation habitats. Journal of the Royal Society of Western Australia. 1989; 71: p. 23-47.
48. Semeniuk V, Semeniuk CA. Sedimentary fill of basin wetlands, central Swan Coastal Plain, southwestern Australia. Part 1: sediment particles, typical sediments, and classification of depositional systems. Journal of the Royal Society of Western Australia. 2004; 87: p. 139-86.
49. Middle G. Geomorphology of Swan Coastal Plain: Garry Middle (VisionEnvironment); 2014 [Available from: <http://www.garrymiddle.net/geomorphology-of-swan-coastal-plain/>].
50. Fitzpatrick RW, Schwertmann U. Al-substituted goethite—An indicator of pedogenic and other weathering environments in South Africa. Geoderma. 1982; 27(4): p. 335-47.
51. Salama RB, Silberstein R, Pollock D. Soils characteristics of the Bassendean and Spearwood sands of the Gnangara Mound (Western Australia) and their controls on recharge, water level patterns and solutes of the superficial aquifer. Water, Air, and Soil Pollution: Focus. 2005; 5: p. 3-26.
52. Semeniuk V, Glassford DK. Bassendean and Spearwood Dunes: their geomorphology, stratigraphy and soils as a basis for habitats of Banksia

- woodlands. *Journal of the Royal Society of Western Australia*. 1989; 71(4): p. 87-8.
53. Earth Science Western Australia. Swan Coastal Plain. 2016.
 54. Melo VF, Testoni SA, Dawson LA, Salvador FAdS. Sand fraction is not suitable for forensic investigations in subtropical soils. *Rev Bras Cienc Solo*. 2020; 44(e0190174).
 55. Turner BL, Hayes PE, Laliberté E. A climosequence of chronosequences in southwestern Australia. *European Journal of Soil Science*. 2018; 69: p. 69-85.
 56. Fitzpatrick RW, Raven MD. How Pedology and Mineralogy Helped Solve a Double Murder Case: Using Forensics to Inspire Future Generations of Soil Scientists. *Soil Horizons*. 2012; 53(5): p. 14-29.
 57. Pye K, Blott SJ, Croft DJ, Carter JF. Forensic comparison of soil samples: Assessment of small-scale spatial variability in elemental composition, carbon and nitrogen isotope ratios, colour, and particle size distribution. *Forensic Science International*. 2006; 163(1-2): p. 59-80.
 58. de Caritat P, Simpson T, Woods B. Predictive Soil Provenancing (PSP): An Innovative Forensic Soil Provenance Analysis Tool. *Journal of Forensic Sciences*. 2019; 64(5): p. 1359-69.
 59. McCulloch G, Dawson LA, Ross JM, Morgan RM. The discrimination of geoforensic trace material from close proximity locations by organic profiling using HPLC and plant wax marker analysis by GC. *Forensic Science International*. 2018; 288: p. 310-26.
 60. Fitzpatrick RW, Raven MD, Forrester ST. A Systematic Approach to Soil Forensics: Criminal Case Studies Involving Transference from Crime Scene to Forensic Evidence. In: Ritz K, Dawson L, Miller D, editors. *Criminal and Environmental Soil Forensics*. 1st ed: Springer; 2009. p. 105-27.
 61. Pye K. *Geological and Soil Evidence: Forensic Applications*: CRC Press; 2007.

62. Woods B, Kirkbride P, Lennard C, Robertson J. Soil examination for a forensic trace evidence laboratory - Part 2: Elemental analysis. *Forensic Science International*. 2014; 245: p. 195-201.
63. Edmond G, Towler A, Grouns B, Ribeiro G, Found B, White D, et al. Thinking forensics: Cognitive science for forensic practitioners. *Science and Justice*. 2017; 57(2): p. 144-54.
64. Prandel LV, Melo VF, Testoni SA, Brinatti AM, Saab SdC, Dawson LA. Spectroscopic techniques applied to discriminate soils for forensic purposes. *Soil Research*. 2020; 58: p. 151–60.
65. Xu X, Du C, Ma F, Shen Y, Zhou J. Forensic soil analysis using laser-induced breakdown spectroscopy (LIBS) and Fourier transform infrared total attenuated reflectance spectroscopy (FTIR-ATR): Principles and case studies. *Forensic Science International*. 2020; 310(110222).
66. Young JM, Weyrich LS, Breen J, Macdonald LM, Cooper A. Predicting the origin of soil evidence: High throughput eukaryote sequencing and MIR spectroscopy applied to a crime scene scenario. *Forensic Science International*. 2015; 251: p. 22–31.
67. Ruffell A, McKinley J. Forensic geoscience: applications of geology, geomorphology and geophysics to criminal investigations. *Earth-Science Reviews*. 2005; 69: p. 235–47.
68. Khajuria H, Gupta S, Nayak BP. Introduction to Polarized Light Microscope. In: Shukla RK, Kapoor N, Badiye A, editors. *Forensic Microscopy: Truth Under the Lenses*. 1st ed: CRC Press; 2022.
69. Agarwal R, Chand A. Forensic Applications of Polarized Light Microscope. In: Shukla RK, Kapoor N, Badiye A, editors. *Forensic Microscopy: Truth Under the Lenses*. 1st ed: CRC Press; 2022.

70. Locard E, Larson DJ. The Analysis of Dust Traces. Part II. The American Journal of Police Science. 1930; 1(4): p. 401-18.
71. Kars H, van den Eijkel L. Proceedings of the Soil Forensics Special, 6th European Academy of Forensic Science Conference. Soil in Criminal and Environmental Forensics. 2016.
72. Prandel LV, Melo VdF, Brinatti AM, Saab SdC, Salvador FAS. X-ray Diffraction and Rietveld Refinement in Deferrified Clays for Forensic Science. Journal of Forensic Sciences. 2018; 63(1): p. 251-7.
73. Pye K, Blott SJ, Wray DS. Elemental analysis of soil samples for forensic purposes by inductively coupled plasma spectrometry - precision considerations. Forensic Science International. 2006; 160(2-3): p. 178-92.
74. Pye K, Croft D. Forensic analysis of soil and sediment traces by scanning electron microscopy and energy-dispersive X-ray analysis: an experimental investigation. Forensic Science International. 2007; 165(1): p. 52-63.
75. Islam K, McBratney AB, Singh B. Estimation of soil colour from visible reflectance spectra. SuperSoil 2004: 3rd Australian New Zealand Soils Conference; University of Sydney, Australia. 2004.
76. Chauhan R, Kumar R, Sharma V. Soil forensics: A spectroscopic examination of trace evidence. Microchemical Journal. 2018; 139: p. 74-84.
77. Edwards H, Munshi T, Scowen I, Surtees A, Swindles GT. Development of oxidative sample preparation for the analysis of forensic soil samples with near-IR Raman spectroscopy. Journal of Raman Spectroscopy. 2012; 43: p. 323-5.
78. Penido CAFdO, Pacheco MTT, Lednev IK, Silveira L. Raman spectroscopy in forensic analysis: identification of cocaine and other illegal drugs of abuse. Journal of Raman Spectroscopy. 2016; 47(1): p. 28-38.
79. Stewart SP, Bell SEJ, Fletcher NC, Bouazzaoui S, Ho YC, Speers SJ, et al. Raman spectroscopy for forensic examination of β -ketophenethylamine "legal highs":

- Reference and seized samples of cathinone derivatives. *Analytica Chimica Acta*. 2012; 711: p. 1-6.
80. Maric M, van Bronswijk W, Pitts K, Lewis SW. Characterisation and classification of automotive clear coats with Raman spectroscopy and chemometrics for forensic purposes. *Journal of Raman Spectroscopy*. 2016; 47(8): p. 948-55.
 81. Stewart SP, Bell SEJ, Armstrong WJ, Kee G, Speers SJ. Forensic examination of multilayer white paint by lateral scanning Raman spectroscopy. *Journal of Raman Spectroscopy*. 2012; 43(1): p. 131-7.
 82. Thomas J, Buzzini P, Massonnet G, Reedy B, Roux C. Raman spectroscopy and the forensic analysis of black/grey and blue cotton fibres. Part 1. Investigation of the effects of varying laser wavelength. *Forensic Science International*. 2005; 152: p. 189-97.
 83. Kavkler K, Demšar A. Examination of cellulose textile fibres in historical objects by micro-Raman spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 2011; 78(2): p. 740-6.
 84. Chalmers JM, Edwards HGM, Hargreaves MD, editors. *Infrared and Raman Spectroscopy in Forensic Science*: John Wiley & Sons; 2012.
 85. Larkin P. *Infrared and Raman Spectroscopy: Principles and Spectral Interpretation*: Elsevier; 2011.
 86. Maric M, Bronswijk Wv, Lewis SW, Pitts K, Martin DE. Characterisation of chemical component migration in automotive paint by synchrotron infrared imaging. *Forensic Science International*. 2013; 228(1-3): p. 165-9.
 87. D'Uva JA, DeTata D, Lewis SW. Source determination of homemade ammonium nitrate using ATR-FTIR spectroscopy, trace elemental analysis and chemometrics. *Forensic Chemistry*. 2022; 28(100411).

88. Wong JXW, Sauzier G, Lewis SW. Forensic discrimination of lipsticks using visible and attenuated total reflectance infrared spectroscopy. *Forensic Science International*. 2019; 298: p. 88–96.
89. McGann J, Willans M, Sauzier G, Hackett MJ, Lewis SW, McGinn J, et al. Investigating diversity in polymer-based identity cards using ATR-FTIR spectroscopy and chemometrics. *Forensic Science International: Reports*. 2020; 2(100149).
90. *The Infrared Spectra of Minerals*. Farmer VC, editor. London: Mineralogical Society; 1974.
91. Ng W, Malone BP, Minasny B. Rapid assessment of petroleum-contaminated soils with infrared spectroscopy. *Geoderma*. 2017; 289: p. 150-60.
92. Caddy B, editor. *Forensic Examination of Glass and Paint*: Taylor & Francis; 2001.
93. Griffiths PR, de Haseth JA, editors. *Fourier Transform Infrared Spectrometry*. 2nd ed: John Wiley & Sons; 2007.
94. Pitts KM, Clarke RM. The forensic discrimination of quartz sands from the Swan Coastal Plain, Western Australia. *Forensic Science International: Reports*. 2020; 2(100130).
95. Pirrie D, Ruffell A, Dawson L. Chapter 5: Geological Evidence Recovery from Exhibits. In: Donnelly LJ, Pirrie D, Harrison MA, Ruffell A, Dawson L, editors. *A Guide to Forensic Geology*. London: Geological Society; 2021. p. 111-28.
96. Marumo Y, Nagatsuka S, Oba Y. *Rapid Clay Mineralogical Analysis for Forensic Science Investigation - Clay Mineralogy Over the Short Distances*. 1988.
97. Clarke R, Bastian L. *The Forensic Comparison of Sandy Soils Using Associated Clay and Heavy Mineral Assemblages as Indicators of Provenance*. Chemistry Centre of Western Australia for the National Institute of Forensic Science. 2004.

98. Pitts KM. The Use of Raman Spectroscopy and X-ray Diffraction for the Examination of Associated Clay and Silt Coatings of Forensically Important Sandy Soils from the Swan Coastal Plain, Western Australia: Curtin University; 2022.
99. Committee on Strengthening Forensic Science at the National Institute of Justice, Committee on Law and Justice, Division of Behavioral and Social Sciences and Education, National Academies of Sciences Engineering and Medicine. Support for Forensic Science Research: Improving the Scientific Role of the National Institute of Justice. Washington, DC; 2015.
100. Norgaard A, Rasmusson B. The likelihood ratio as value of evidence - more than a question of numbers. *Law, Probability, and Risk*. 2012; 11: p. 303-15.
101. Science and Technology Select Committee. Forensic science and the criminal justice system: a blueprint for change. United Kingdom: House of Lords; 2019.
102. Roux C, Bucht R, Crispino F, De Forest P, Lennard C, Margot P, et al. The Sydney declaration – Revisiting the essence of forensic science through its fundamental principles. *Forensic Science International*. 2022; 332(111182).
103. van Straalen EK, dePoot CJ, Malsch M, Elffers H. The interpretation of forensic conclusions by criminal justice professionals: The same evidence interpreted differently. *Forensic Science International*. 2020; 313(110331).
104. Stern HS. Statistical Issues in Forensic Science. *Annual Review of Statistics and Its Application*. 2017; 4: p. 225-44.
105. Sauzier G, Bronswijk Wv, Lewis SW. Chemometrics in forensic science: approaches and applications. *Analyst*. 2021; 146(8): p. 2415-48.
106. President’s Council of Advisors on Science and Technology (PCAST). Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. USA; 2016.

107. Sauzier G, Reichard E, van Bronswijk W, Lewis SW, Goodpaster JV. Improving the confidence of “questioned versus known” fiber comparisons using microspectrophotometry and chemometrics. *Forensic Chemistry*. 2016; 2: p. 15-21.
108. Garbolino P, Taroni F. Evaluation of scientific evidence using Bayesian networks. *Forensic Science International*. 2002; 125: p. 149-55.
109. Pitts K. Personal communication. 2022.
110. Bastian LV. The dune systems of the Swan Coastal Plain: Subdivision based on mineral trends in the surface soils. Chemistry Centre of WA. 1994.
111. Blott SJ, Croft DJ, Pye K, Saye SE, Wilson HE. Particle size analysis by laser diffraction. In: Pye K, Croft DJ, editors. *Forensic Geoscience: Principles, Techniques and Applications*. London: Geological Society Publishing House; 2004. p. 63–73.
112. Post DF, Bryant RB, Batchily AK, Huete AR, Levine SJ, Mays MD, et al. *Correlations Between Field and Laboratory Measurements of Soil Color*: Soil Science Society of America; 1993.
113. Yélamos O, Garcia R, D’Alessandro B, Thomas M, Patwardhan S, Malveyh J. *Understanding Color*. In: Pasquali P, editor. *Photography in Clinical Medicine*. 1st ed. Switzerland: Springer Cham; 2020.
114. Stoecklein W. The role of colour and microscopic techniques for the characterisation of paint fragments. In: Caddy B, editor. *Forensic Examination of Glass and Paint*. London: Taylor & Francis; 2001. p. 143-63.
115. MUNSSELL COLOUR ©. How to Read a Munsell Color Chart: X-Rite, Incorporated; [Available from: <https://munsell.com/about-munsell-color/how-color-notation-works/how-to-read-color-chart/>].
116. Zięba-Palus J, Trzcińska BM. Comparing the Color of Forensic Traces. *Analytical Letters*. 2012; 45(11): p. 1333-46.

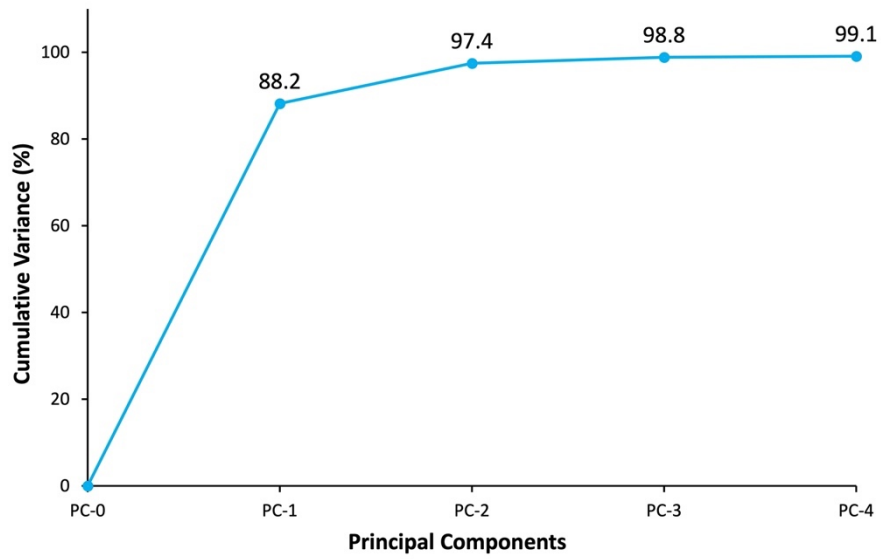
117. Duarte I, Rotter A, Malvestiti A, Silva M. The role of glass as a barrier against the transmission of ultraviolet radiation: An experimental study. *Photodermatology, Photoimmunology and Photomedicine*. 2009; 25(4): p. 181-4.
118. Scientific Working Group on Materials Analysis. *Forensic Paint Analysis and Comparison Guidelines*. 2000.
119. Bengtsson S, Berglof T, Kylin H. Near Infrared Reflectance Spectroscopy as a Tool to Predict Pesticide Sorption in Soil. *Bulletin of Environmental Contamination and Toxicology*. 2007; 78: p. 295-8.
120. Dawson L. Soil organic characterisation in forensic case work. *International Union of Geological Sciences, Initiative on Forensic Geology (IUGS-IFG)*. 2017; 40(2): p. 157-65.
121. Ewald M, Bell C, Berger P. Corrected Fluorescence Spectra of Fulvic Acids Isolated from Soil and Water. *Environmental Science & Technology*. 1983; 17(8).
122. Lanfranco AM, Schofield PF, Murphy PJ, Hodson ME, Mosselmans JFW, Valsami-Jones E. Characterization and identification of mixed-metal phosphates in soils: the application of Raman spectroscopy. *Mineralogical Magazine*. 2003; 67(6): p. 1299–316.
123. Xing Z, Du C, Zeng Y, Ma F, Zhou J. Characterizing typical farmland soils in China using Raman spectroscopy. *Geoderma*. 2016; 268: p. 147-55.
124. Vogel C, Ramsteiner M, Sekine R, Doolette A, Adam C. Characterization of phosphorus compounds in soils by deep ultraviolet (DUV) Raman microspectroscopy. *Journal of Raman Spectroscopy*. 2017; 48: p. 867–71.
125. Ramaswamy S, Raghavan P. Significance of Impurity Mineral Identification in the Value Addition of Kaolin – A Case Study with Reference to an Acidic Kaolin

- from India. *Journal of Minerals and Materials Characterization and Engineering*. 2011; 10(11): p. 1007-25.
126. Young JM, Weyrich LS, Cooper A. Forensic soil DNA analysis using high-throughput sequencing: A comparison of four molecular markers. *Forensic Science International: Genetics*. 2014; 13: p. 176-84.
 127. Dor EB, Ong C, Lau IC. Reflectance measurements of soils in the laboratory: Standards and protocols. *Geoderma*. 2015; 245-246: p. 112-24.
 128. National Association of Testing Authorities Australia. Technical Note 17: Guidelines for the validation and verification of quantitative and qualitative test methods. Australia. 2012.
 129. Di Maggio RM, Donnelly LJ, Naimi KSA, Barone PM, Salvador FADS, Dawson L, et al. Global developments in forensic geology. *Journal of International Geoscience: Episodes*. 2017; 40(2): p. 120-31.
 130. Brathwaite RL, Christie AB, Faure K, Townsend MG, Terlesk S. Origin of the Matauri Bay halloysite deposit, Northland, New Zealand. *Mineralium Deposita*. 2012; 47(8): p. 897-910.
 131. Saigusa M, Shoji S, Kato T. Origin and nature of halloysite in Ando soils from Towada tephra, Japan. *Geoderma*. 1978; 20(2): p. 115-29.
 132. Young SL, Wilson MJ, Hillier S, Delbos E, Ali SM, Stoltzfus RJ. Differences and Commonalities in Physical, Chemical and Mineralogical Properties of Zanzibari Geophagic Soils. *Journal of Chemical Ecology*. 2010; 36: p. 129-40.
 133. Liu H, Chen T, Frost RL, Chang D, Qing C, Xie Q. Effect of aging time and Al substitution on the morphology of aluminous goethite. *Journal of Colloid and Interface Science*. 2012; 385(1): p. 81-6.
 134. Churchward HM, Bettenay E. The physiographic significance of conglomeratic sediments and associated laterites in valleys of the Darling Plateau, near

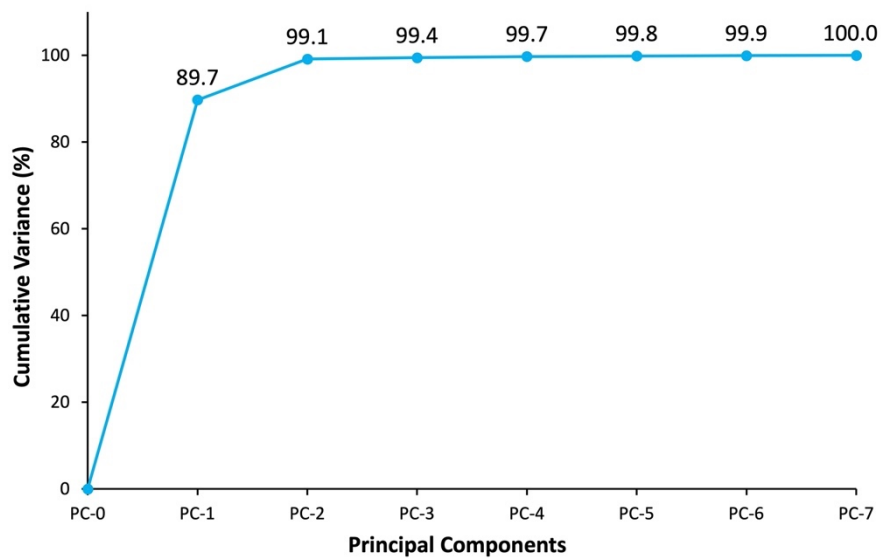
- Harvey, Western Australia. *Journal of the Geological Society of Australia*. 1973; 20(3): p. 309-17.
135. Guinea JG, Correcher V, Valle-Fuentes FJ. Thermoluminescence of kaolinite. *Radiation Protection Dosimetry*. 1999; 84(1-4): p. 507-10.
136. Leonardi A, Bish DL. Understanding Powder X-ray Diffraction Profiles from Layered Minerals: The Case of Kaolinite Nanocrystals. *Inorganic Chemistry*. 2020; 59: p. 5357-67.
137. Sauzier G, McGinn J, Trubshoe T, Lewis SW. In situ examination of handwritten blue ballpoint inks using video spectral comparison with chemometrics. *Forensic Science International: Reports*. 2019; 1(100021).
138. National Association of Testing Authorities. NATA procedures for accreditation. Australia. 2022.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Appendices



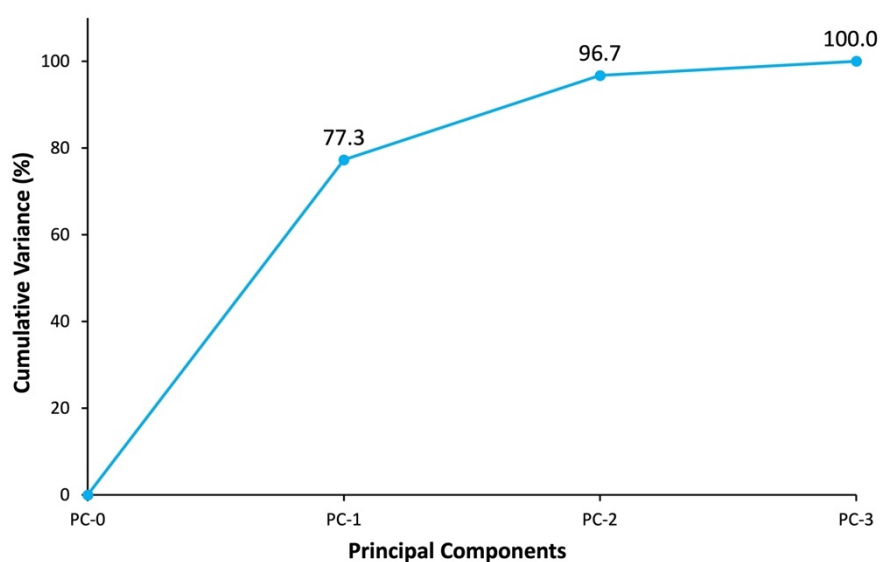
Appendix 6.1 Scree plot depicting the cumulative variance retained by each PC in the blinded case simulation full MSP dataset.



Appendix 6.2 Scree plot depicting the cumulative variance retained by each PC in the blinded case simulation average MSP dataset.

Appendix 6.3 The blinded case simulation soil samples analysed with MSP and their associated $L^*a^*b^*$ chromaticity values generated from their average MSP spectra (to five decimal places).

Soil sample	L^* value	a^* value	b^* value
BS1a – Alibi Site 1	68.06082	6.27448	15.24703
BS1b – Alibi Site 1	70.07555	4.70414	11.43255
BS2a – Suspect Recovered	69.62115	6.17024	12.29732
BS2b – Suspect Recovered	70.36893	3.56538	10.55008
BS3a – Crime Scene	73.62930	5.42913	14.10167
BS3b – Crime Scene	65.01199	4.69434	12.01520
BS4a – Alibi Site 2	64.14069	4.66415	11.07226
BS4b – Alibi Site 2	64.34891	3.82982	10.51401
BS5a – Potential Site of Interest	69.04498	4.45102	14.52991
BS5b – Potential Site of Interest	69.32379	4.60865	13.45431



Appendix 6.4 Scree plot depicting the cumulative variance retained by each PC in the blinded case simulation $L^*a^*b^*$ chromaticity dataset.

Appendix 6.5 Number of correct vs incorrect classifications for samples in the MSP blinded case simulation calibration set using a 4-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Classified	% Correct
Alibi Site 1	14	6	Alibi Site 2	70
Crime Scene	19	1	Alibi Site 1	95
Alibi Site 2	10	10	Alibi Site 1 (8), Crime Scene (1), Potential Site of Interest (1)	50
Potential Site of Interest	18	2	Crime Scene	90
				% Total Correct
				76

Appendix 6.6 Discriminant values of the suspect recovered soil's replicates from the MSP blinded case simulation validation dataset (rounded to three decimal places), with predictions shaded green. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

Suspect Recovered Replicate	Discriminant Values				Predicted Location	Closest Value Within (%)
	Alibi Site 1	Crime Scene	Alibi Site 2	Potential Site of Interest		
Sample A (1)	-7.017	-2.315	-5.334	-4.217	Crime Scene	82
Sample A (2)	-5.335	-1.619	-4.987	-5.883	Crime Scene	208
Sample A (3)	-4.607	-13.932	-7.240	-21.674	Alibi Site 1	57
Sample A (4)	-7.255	-2.015	-6.532	-5.492	Crime Scene	173
Sample A (5)	-1.745	-3.658	-2.006	-8.249	Alibi Site 1	15
Sample A (6)	-1.835	-3.777	-2.494	-8.948	Alibi Site 1	36
Sample A (7)	-5.619	-14.090	-8.666	-22.538	Alibi Site 1	54
Sample A (8)	-4.689	-3.951	-5.097	-10.542	Crime Scene	19
Sample A (9)	-6.012	-5.144	-7.378	-12.840	Crime Scene	17
Sample A (10)	-6.801	-4.632	-6.568	-7.973	Crime Scene	42
Sample B (1)	-6.144	-3.576	-5.499	-8.263	Crime Scene	54
Sample B (2)	-11.059	-9.540	-12.772	-20.484	Crime Scene	16
Sample B (3)	-8.194	-9.894	-8.794	-17.731	Alibi Site 1	7
Sample B (4)	-5.144	-2.534	-4.419	-5.369	Crime Scene	74
Sample B (5)	-5.961	-5.246	-5.267	-9.683	Crime Scene	85
Sample B (6)	-5.999	-7.823	-6.150	-14.066	Alibi Site 1	3
Sample B (7)	-10.283	-8.244	-10.409	-15.825	Crime Scene	25
Sample B (8)	-12.026	-13.210	-12.542	-21.645	Alibi Site 1	4
Sample B (9)	-6.887	-4.419	-5.374	-7.091	Crime Scene	22
Sample B (10)	-5.136	-3.119	-3.792	-5.524	Crime Scene	22

Appendix 6.7 Number of correct vs incorrect classifications for samples in the average MSP blinded case simulation calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Classified	% Correct
Alibi Site 1	2	0	-	100
Crime Scene	2	0	-	100
Alibi Site 2	2	0	-	100
Potential Site of Interest	2	0	-	100
% Total Correct				100

Appendix 6.8 Discriminant values of the suspect recovered soil's replicates from the average MSP blinded case simulation validation dataset (rounded to three decimal places), with predictions shaded green. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

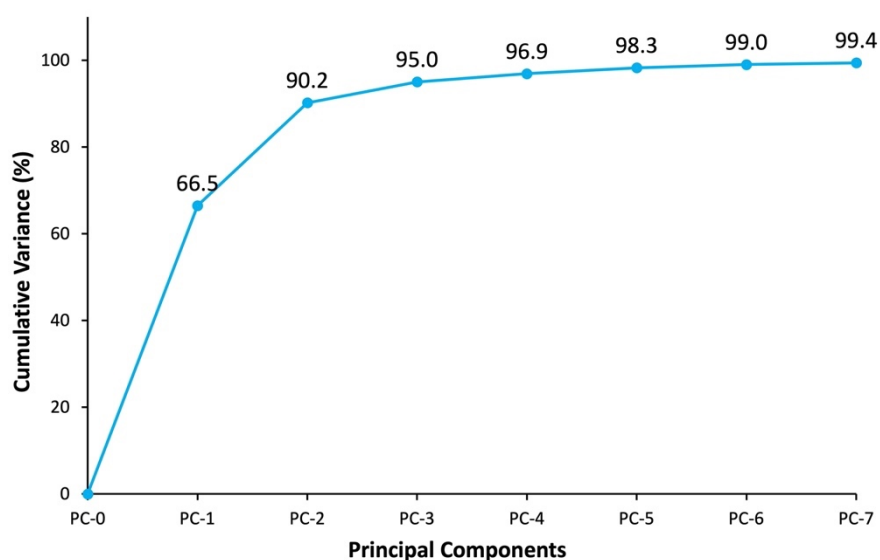
Suspect Recovered Sample	Discriminant Values				Predicted Location	Closest Value Within (%)
	Alibi Site 1	Crime Scene	Alibi Site 2	Potential Site of Interest		
Sample A	-1.390	-14.907	-5.940	-69.089	Alibi Site 1	327
Sample B	-11.089	-5.196	-5.422	-39.479	Crime Scene	4

Appendix 6.9 Number of correct vs incorrect classifications for samples in the L*a*b* chromaticity values blinded case simulation calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).

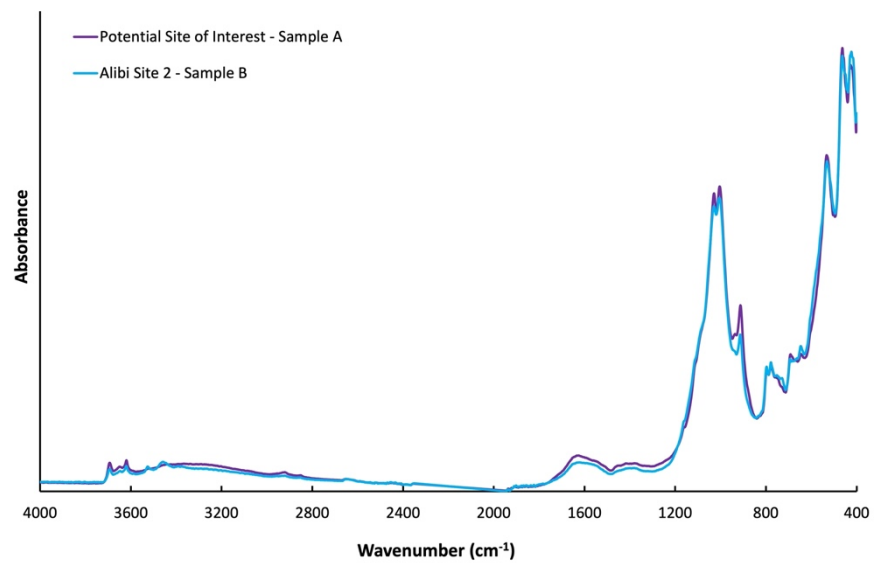
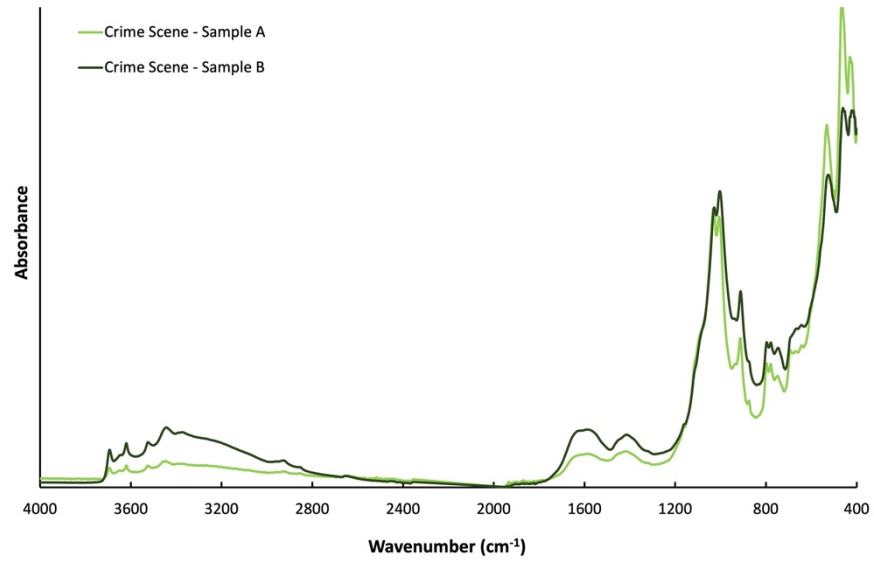
Location	Correct	Incorrect	Classified	% Correct
Alibi Site 1	0	2	Crime Scene & Potential Site of Interest	0
Crime Scene	1	1	Alibi Site 2	50
Alibi Site 2	2	0	-	100
Potential Site of Interest	1	1	Crime Scene	50
% Total Correct				50

Appendix 6.10 Discriminant values of the suspect recovered soil's replicates from the L*a*b* blinded case simulation validation dataset (rounded to three decimal places), with predictions shaded green. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

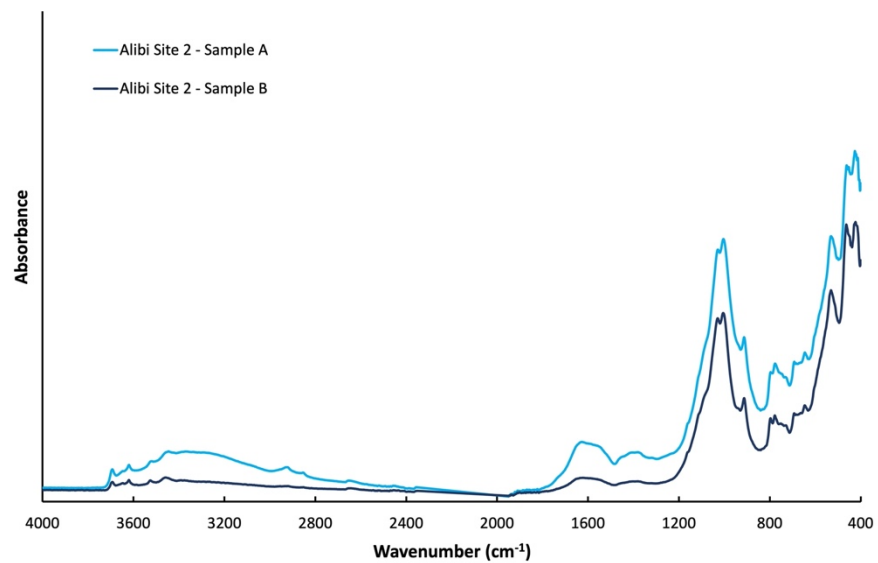
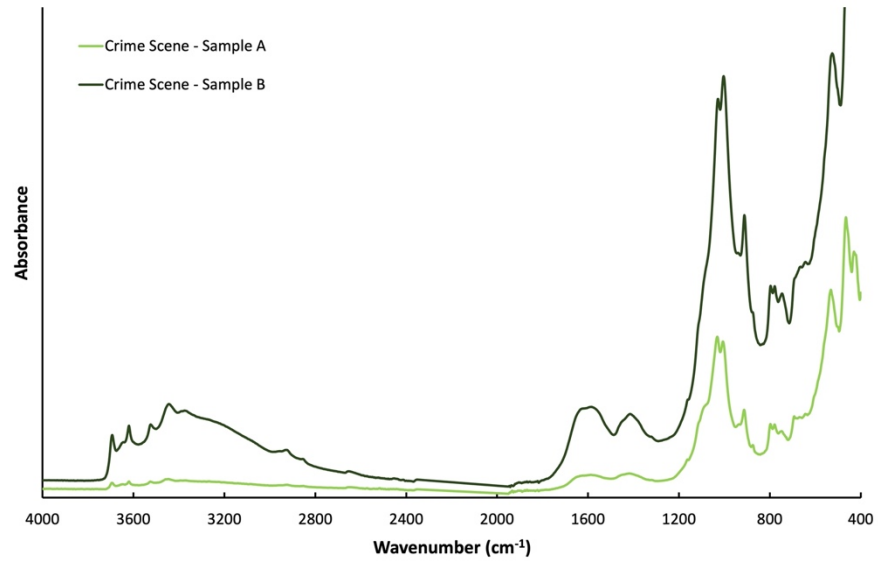
Suspect Recovered Sample	Discriminant Values				Predicted Location	Closest Value Within (%)
	Alibi Site 1	Crime Scene	Alibi Site 2	Potential Site of Interest		
Sample A	-1.531	-1.421	-3.172	-1.632	Crime Scene	8
Sample B	-3.606	-3.067	-3.588	-3.977	Crime Scene	17



Appendix 6.11 Scree plot depicting the cumulative variance retained by each PC in the blinded case simulation ATR-FTIR dataset.



Appendix 6.12 Representative ATR-FTIR spectra obtained from the crime scene soil samples (top), illustrating the variation in spectra from different samples, and from alibi site 2 and the potential site of interest (bottom), illustrating the similarities between these samples.



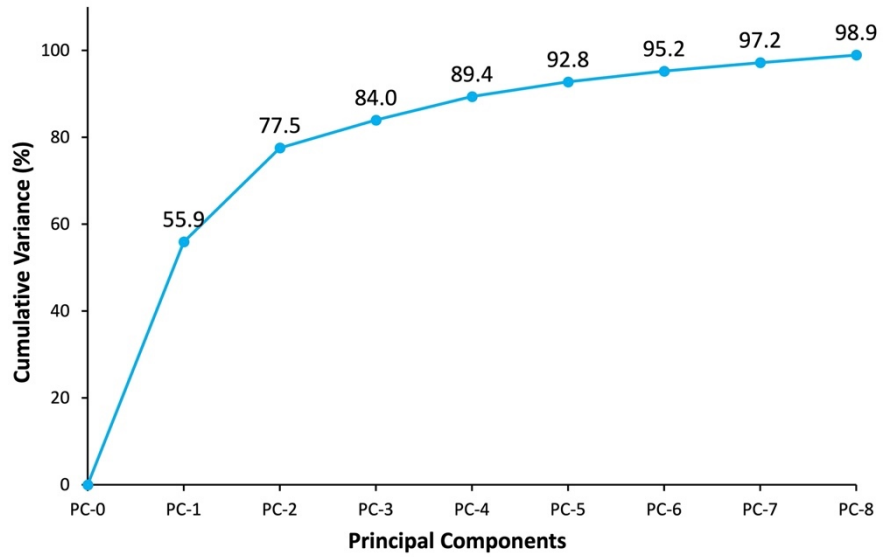
Appendix 6.13 Raw, unprocessed ATR-FTIR spectra obtained from the crime scene soil samples (top) and alibi site 2 samples (bottom), illustrating the variation in overall absorbance of different samples potentially due to amount of sample on the ATR crystal.

Appendix 6.14 Number of correct vs incorrect classifications for samples in the ATR-FTIR blinded case simulation calibration set using a 6-PC LDA model (percentages rounded to nearest whole number).

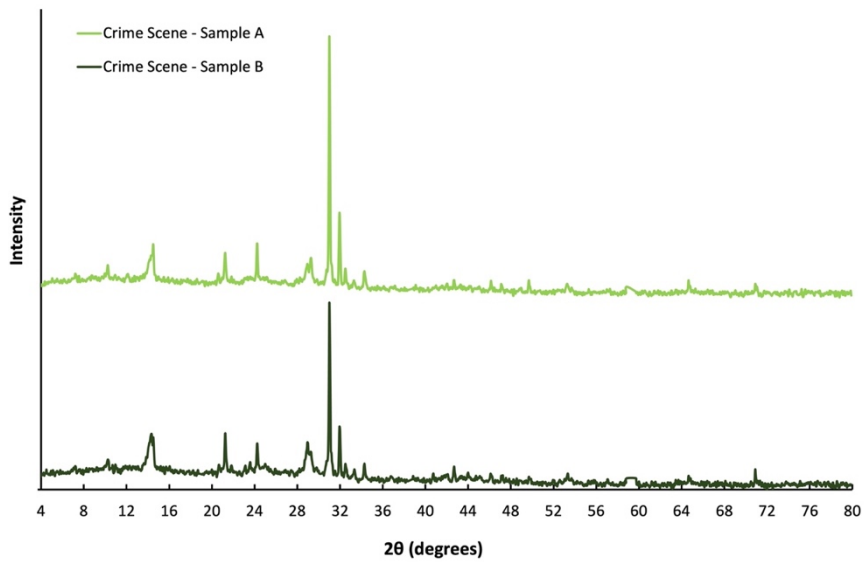
Location	Correct	Incorrect	Classified	% Correct
Alibi Site 1	6	0	-	100
Crime Scene	6	0	-	100
Alibi Site 2	6	0	-	100
Potential Site of Interest	6	0	-	100
% Total Correct				100

Appendix 6.15 Discriminant values of the suspect recovered soil's replicates from the ATR-FTIR blinded case simulation validation dataset (rounded to three decimal places), with predictions shaded green. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

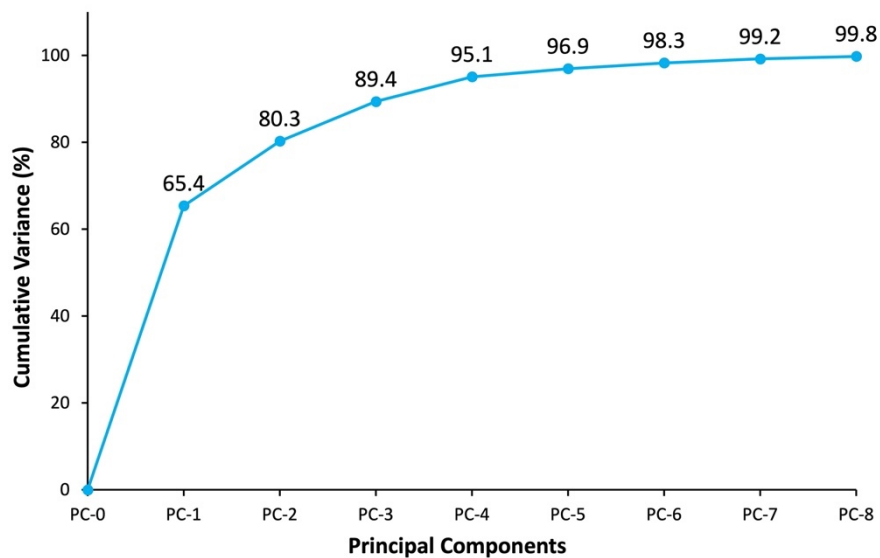
Suspect Recovered Replicate	Discriminant Values				Predicted Location	Closest Value Within (%)
	Alibi Site 1	Crime Scene	Alibi Site 2	Potential Site of Interest		
Sample A (1)	-2157.343	-9.947	-1267.209	-3360.959	Crime Scene	12640
Sample A (2)	-2386.464	-13.185	-1156.206	-3050.883	Crime Scene	8669
Sample A (3)	-1977.601	-38.556	-891.365	-2851.056	Crime Scene	2212
Sample B (1)	-1838.478	-116.671	-640.415	-2465.914	Crime Scene	449
Sample B (2)	-1864.081	-124.767	-615.409	-2408.985	Crime Scene	393
Sample B (3)	-1931.865	-102.943	-657.131	-2436.571	Crime Scene	538



Appendix 6.16 Scree plot depicting the cumulative variance retained by each PC in the blinded case simulation ChemCentre XRD dataset.



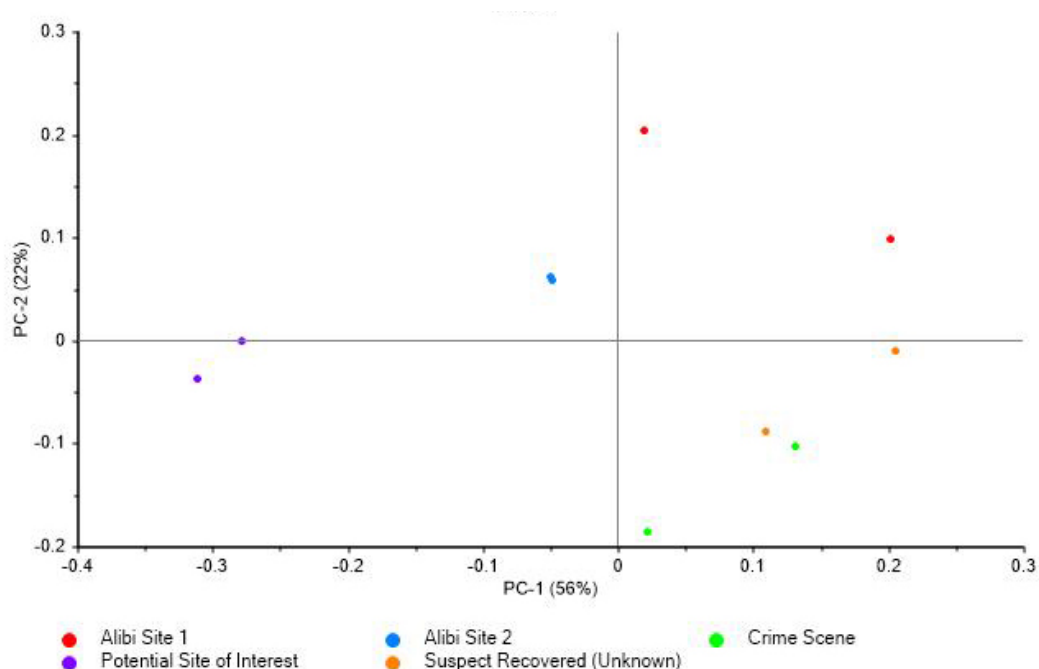
Appendix 6.17 ChemCentre XRD patterns obtained from the crime scene soil samples, illustrating the variation in patterns from duplicate samples.



Appendix 6.18 Scree plot depicting the cumulative variance retained by each PC in the blinded case simulation CSIRO XRD dataset.

Appendix 6.19 Number of correct vs incorrect classifications for samples in the ChemCentre XRD blinded case simulation calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Classified	% Correct
Alibi Site 1	2	0	-	100
Crime Scene	2	0	-	100
Alibi Site 2	2	0	-	100
Potential Site of Interest	2	0	-	100
% Total Correct				100



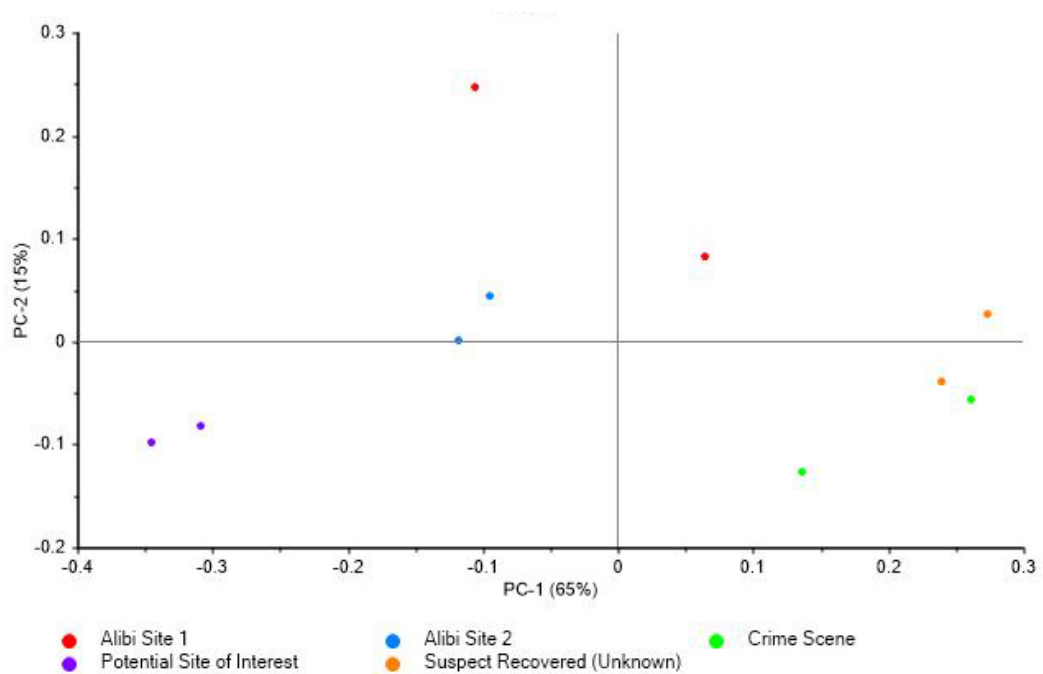
Appendix 6.20 2-dimensional PCA scores plots generated using the first two PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding ChemCentre XRD patterns.

Appendix 6.21 Discriminant values of the suspect recovered soil's replicates from the ChemCentre XRD blinded case simulation validation dataset (rounded to three decimal places), with predictions shaded green. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

Suspect Recovered Sample	Discriminant Values				Predicted Location	Closest Value Within (%)
	Alibi Site 1	Crime Scene	Alibi Site 2	Potential Site of Interest		
Sample A	-12.094	-2.664	-5.567	-16.060	Crime Scene	109
Sample B	-5.646	-8.283	-6.233	-25.655	Alibi Site 1	10

Appendix 6.22 Number of correct vs incorrect classifications for samples in the CSIRO XRD blinded case simulation calibration set using a 2-PC LDA model (percentages rounded to nearest whole number).

Location	Correct	Incorrect	Classified	% Correct
Alibi Site 1	2	0	-	100
Crime Scene	2	0	-	100
Alibi Site 2	2	0	-	100
Potential Site of Interest	2	0	-	100
% Total Correct				100



Appendix 6.23 2-dimensional PCA scores plots generated using the first two PCs, showing the variability of soil samples from different locations within the blinded case simulation, based on their corresponding CSIRO XRD patterns.

Appendix 6.24 Discriminant values of the suspect recovered soil's replicates from the CSIRO XRD blinded case simulation validation dataset (rounded to three decimal places), with predictions shaded green. The last column demonstrates how far away the next closest prediction was, as a percentage of the lowest discriminant value obtained.

Sample Replicate	Discriminant Values				Predicted Location	Closest Value Within (%)
	Alibi Site 1	Crime Scene	Alibi Site 2	Potential Site of Interest		
Sample A	-7.915	-4.979	-17.977	-57.079	Crime Scene	59
Sample B	-7.660	-2.167	-12.526	-44.927	Crime Scene	254

Appendix 7 Attribution Statements:

To whom it may concern,

I, **Talia Newland**, contributed conceptualisation, investigation, methodology, visualization, writing-original draft, writing-review & editing to the following paper:

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics for the forensic analysis of Western Australian sandy soils." *Forensic Chemistry*, 2022. 28: 100412.

I, as a Co-Author, endorse that this level of contribution by the candidate indicated above is appropriate.

Kari Pitts

Simon W. Lewis

To whom it may concern,

I, **Talia Newland**, contributed conceptualisation, investigation, methodology, visualization, writing-original draft, writing-review & editing to the following paper:

K. Pitts, S. Lewis, and **T. Newland**. "Geochemistry | Soil and Mineralogical Analysis." In *Encyclopedia of Analytical Science (Third Edition)*, P. Worsfold, C. Poole, A. Townshend, and M. Miró, 2019. p. 292-301. Oxford: Academic Press.

I, as a Co-Author, endorse that this level of contribution by the candidate indicated above is appropriate.

Kari Pitts

Simon W. Lewis

To whom it may concern,

I, **Talia Newland**, contributed conceptualisation, investigation, methodology, visualization, writing-original draft, writing-review & editing to the following paper:

T. G. Newland, K. Pitts, and S. W. Lewis. "Multimodal spectroscopy with chemometrics: Application to simulated forensic soil casework." *Forensic Chemistry*, 2023. 33: 100481.

I, as a Co-Author, endorse that this level of contribution by the candidate indicated above is appropriate.

Kari Pitts

Simon W. Lewis

To whom it may concern,

I, **Talia Newland**, contributed conceptualisation, investigation, methodology, visualization, writing-original draft, writing-review & editing to the following paper:

T. G. Newland, K. Pitts, and S. W. Lewis. "Negative result: Application of Raman spectroscopy to the forensic analysis of an arid, sandy, soil." *Forensic Science International: Reports*, 2022. Submitted for publication.

I, as a Co-Author, endorse that this level of contribution by the candidate indicated above is appropriate.

Kari Pitts

Simon W. Lewis