Multimodal spectroscopy with chemometrics for the forensic analysis of Western Australian sandy soils

Talia G. Newland[1], Kari Pitts[2], Simon W. Lewis[1*]

[1]School of Molecular and Life Sciences, Curtin University, GPO Box U1987, Perth, Western Australia, 6845.

[2]Physical Evidence Section, Forensic Science Laboratory, ChemCentre, Resources and Chemistry Precinct, Cnr Manning Road and Townsing Drive, Bentley, Western Australia, 6102, Australia.

*corresponding author: s.lewis@curtin.edu.au

Abstract

This study demonstrates a multi-modal analytical sequence suited to the characterisation of sandy soils, which remain an underutilised form of forensic trace evidence. Within the Swan Coastal Plain in Perth, Western Australia, most soils are heavily leached with only small deviations in their mineral compositions. Their lack of clay and organic matter has led to inorganic methods of analysis predominating, in addition to experimentation with modified techniques. One example is utilisation of the quartz-recovered fine fraction, which is suitable for dry, quartz-dominated sandy soils. In this study, preliminary investigations used the spectroscopic techniques microspectrophotometry, infrared spectroscopy, and x-ray diffraction, to develop a multi-faceted approach for the forensic analysis of the quartz fine fraction of soils. These data were then combined with principal component analysis to demonstrate how chemometrics can assist with objective characterisation and differentiation of sandy soil samples for forensic purposes. Chemometric analysis has not previously been attempted with data obtained from the quartz fine fraction. This methodology is transferable to other jurisdictions where dry, sandy soils predominate.

Keywords: Forensic science, Soils, Inorganic, Spectroscopy, Chemometrics.

1.0   Introduction

Soil can be of major value as trace evidence due to its complexity, heterogeneity, and transferability (1-5). There has been a reported increase in prevalence of soil evidence in recent international criminal investigations, leading to increased demand for more reliable techniques for objective forensic examination of soils (6-9). Making more extensive use of soil evidence requires improved knowledge of the soils within specific regions, such as the Swan Coastal Plain within Perth, Western Australia. Understanding the frequency and nature of a particular trace is central to being able to evaluate its source (10). This is particularly true for soils, as their properties can fluctuate greatly across different locations.

The Swan Coastal Plain in Western Australia encompasses most of the Perth metropolitan region (11-13). It is formed of four main systems; the Quindalup, Spearwood, and Bassendean dunes, and the Pinjarra Plain (11-14). A more detailed description of the Perth dune systems can be found in Pitts et al. (15). Within these dunes, the majority of soils are heavily leached quartz sands, with only small deviations in their mineral compositions (13, 14). Analytical approaches focusing on the inorganic content of these soils predominate, due to the lack of clay and organic matter. While there is a current demand for increased use of soil evidence to assist in solving crime, there is still an absence of any in-depth investigations in published literature into sandy soils, such as those found on the Swan Coastal Plain, and their role as geological evidence in a forensic context.

Within forensic soil science, techniques for the analysis of the inorganic content have been widely established across many disciplines (2, 6, 9, 16-19). Most of the methods that make use of bulk soil samples for analysis (20) are impractical for soils that have minimal variation in their bulk chemistry, such as quartz sands (21, 22). These soils require more detailed examinations to be able to distinguish between different sources. In many approaches to forensic soil analysis, several techniques are undertaken on separated fractions of soil, allowing further differentiation compared to the bulk chemistry (22-25). The comparison of the quartz-recovered fine fraction (primary and secondary minerals < 20 µm found as clay and fine silt coatings on quartz grains) is a method recently developed by Pitts and Clarke, which allows for differentiation of minute and fractionated forensic samples (21).

A further challenge to forensic soil analysis is the interpretation of results. This is an issue that has been the topic of significant discussion within the international forensic community over recent decades (10, 26, 27). One approach that will assist in a more objective and transparent interpretation of soil evidence is to utilise multivariate statistical methods, or chemometrics, to evaluate the relationships between samples (18, 24, 28). Data can be analysed using unsupervised techniques to establish patterns or correlations, such as principal component analysis (PCA), or supervised techniques that require pre-established groupings, such as linear discriminative analysis (LDA) (18, 24, 29). 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 (8, 9, 24, 28, 29). This can prove useful when undertaking common 'questioned versus known' comparisons. In addition, examination of loadings plots can provide information on the chemical factors that allow differentiation between samples (30).

Whilst the use of chemometrics has been extensively applied to organic forensic soil analysis (2, 4, 6, 9, 18, 29, 31-34), there has been limited reports of use with inorganic

data (3, 22, 35, 36). In addition, 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 (2, 18, 24, 29, 35, 36). Soil analysis, like other areas of forensic analysis, is typically carried out in a sequence, utilizing the least destructive methods first and gathering further information as the sequence progresses. It would therefore be more realistic to look at a suite of instrumental methods and experiment with combining them in a sequence. 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. Chemometric analysis has also never been demonstrated on data obtained from methods analysing the quartz 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.

In this study, preliminary proof-of-concept investigations explored the use of well-established spectroscopic techniques to further develop a multi-faceted approach for forensic analysis of the quartz fine fraction of soils, that can provide better evidence of association in a courtroom context. These data were then used in combination with PCA to demonstrate how chemometrics can be utilised for the objective characterisation and differentiation of sandy soil samples for forensic purposes.

2.0   Materials and Methods

2.1   Collection of soil samples

Soil samples were collected from a large range of areas across the Perth (Western Australia) metropolitan region, with the locations specifically chosen to allow for a selection of samples from differing dune and plant systems. This was done by first gently clearing the area of large organic material such as leaf litter or mulch, placing a wooden frame (30 cm x 30 cm) 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. This process was repeated in the other three corners as well as the centre, producing a total of 5 soil samples per location. The GPS coordinates of these locations were recorded, along with information on the accompanying geological and botanical influences surrounding the sites. All samples were initially stored in a -20 °C freezer and then freeze-dried to minimise degradation. Nine (9) initial locations were selected for preliminary investigations, with 2 soil samples from each location analysed for this study (these were randomly selected from the total of 5 collected).

2.2   Preparation of quartz fine fraction

Preparation of the quartz fine fraction was based on the method developed by Pitts and Clarke (21). 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, and transferred to liquid scintillation vials. Deionised water was added to each vial to a height of approximately 1.5 cm, and samples were then ultrasonicated for 10 minutes. One by one, samples were agitated and then settled for 30 seconds before the liquid was decanted into new vials, to isolate particles of approximately < 20 µm. These were then centrifuged, and the supernatant was removed without disturbing the semi-solid layer concentrated below. The semi-solid material was then mixed with the remaining liquid before the suspensions were pipetted onto sample holders and dried as a thin film.

2.3   Microspectrophotometry analysis

The prepared quartz fine fractions were homogenised and mounted on glass microscope slides 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.

2.4   Attenuated total reflectance infrared analysis

The prepared quartz 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. 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 soil sample. 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.

2.5   X-ray diffraction analysis

X-ray diffraction 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 CoKα 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.

2.6   Order of analysis

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. Analysis was initiated by conducting a visual examination of the soil samples as the first step, followed by analysis of the colour instrumentally using MSP, then ATR-FTIR spectroscopy, and lastly XRD. This is the order in which a forensic examiner would generally already carry out the forensic analysis of soils (8, 37). While XRD is presented as the final step in our methodology, the sample preparation was actually 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 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 analysis.

2.7   Statistical analysis of spectroscopic data

Chemometric analysis was conducted using The Unscrambler® X 10.5 (CAMO Software AS, Oslo, Norway). The regions associated with interference from the ATR diamond crystal (2350 – 1950 cm^{-1}) in IR trials and the low background plates (narrow regions centred on approximately 18.9, 38.5, and 59.3 degrees 2θ) in XRD trials were excluded from chemometric analysis to prevent them from influencing the model. In chemometric analysis of XRD patterns, generally peak picking or profile fitting occurs first, and the data generated from this is then utilised for multivariate analysis [21]. This study however made use of the whole XRD pattern for chemometric analysis (minus the peaks associated with the low background plates) in the same way that MSP and ATR-FTIR spectra are used in their entirety. Data were baseline offset corrected and normalised (range normalised for all XRD trials, unit vector normalised for all other trials) prior to PCA analysis. Data were then mean-centred and subjected to principal component analysis (PCA) using the non-linear iterative partial least squares (NIPALS) algorithm. 3-dimensional scores plots were generated using the scores from the relevant principal components (PCs) to visualise the sample distribution.

3.0   Results and Discussion

3.1   Preliminary considerations

A total of 18 samples from 9 locations were chosen for analysis (full details outlined in Table 1). Photographs taken of the soil samples (Fig. 1) prior to sieving and quartz recovery were used to determine their overall visual colour. 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. It is evident that several of the soil samples are immediately distinguishable from one another based on visual inspection alone, however locations 1, 3, 8, and 9 have little distinctive differences, requiring more advanced techniques to objectively tell them apart.

Figure 1. Soil samples obtained from the 9 different locations used throughout this study, illustrating the visual colour differences.

Table 1. The soil samples used in preliminary trials throughout this study, with their associated locations and visual descriptions (*sample 2a was only included in XRD trials). 2 samples were analysed from each location. Map showing locations and GPS coordinates is included in ESI (Fig. A1 & Table A1).

Soil Sample	Location No.	Location (Perth Suburb)	Visual Appearance	Dune System
1a	1	Kings Park	Grey sand	Spearwood
1b				
2a*	2	Coolbinia	Dark brown sand/mulch	Spearwood
2b				
3a	3	Bayswater	Grey sand	Bassendean
3b				
4a	4	Balga	Yellow sand	Spearwood
4b				
5a	5	Edgewater	Yellow/brown sand	Spearwood
5b				
6a	6	Murdoch University	Brown sand	Spearwood
6b				
7a	7	Champion Lakes	Red sand	Pinjarra Plain (mixed)
7b				
8a	8	Yangebup	Grey sand	Spearwood
8b				
9a	9	Banjup	Grey sand	Bassendean
9b				

3.2   Microspectrophotometry

MSP spectra (Fig. 2) were collected from the extracted quartz fine fraction of 17 soil samples originating from the 9 different locations detailed in Table 1. Whilst some samples could be separated from the population based on visual examination of the spectra alone, e.g. sample 7a, others were too alike to confidently distinguish. Chemometric methods were employed to identify and enhance any differences between them.
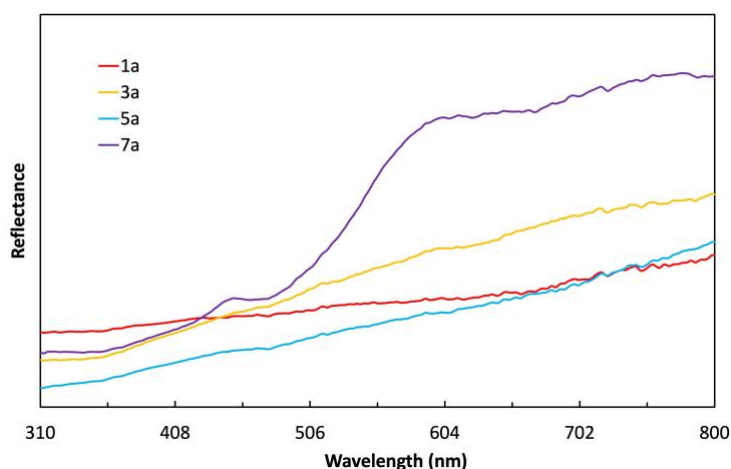

Figure 2. MSP reflectance spectra showing the variability in composition of 4 soil samples collected from differing locations within the Swan Coastal Plain.

PCA performed on these spectra revealed that 97.9% of the total variance in the dataset could be described by the first three PCs (Fig. A2). Three-dimensional score plots generated using these PCs (Fig. 3) 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.
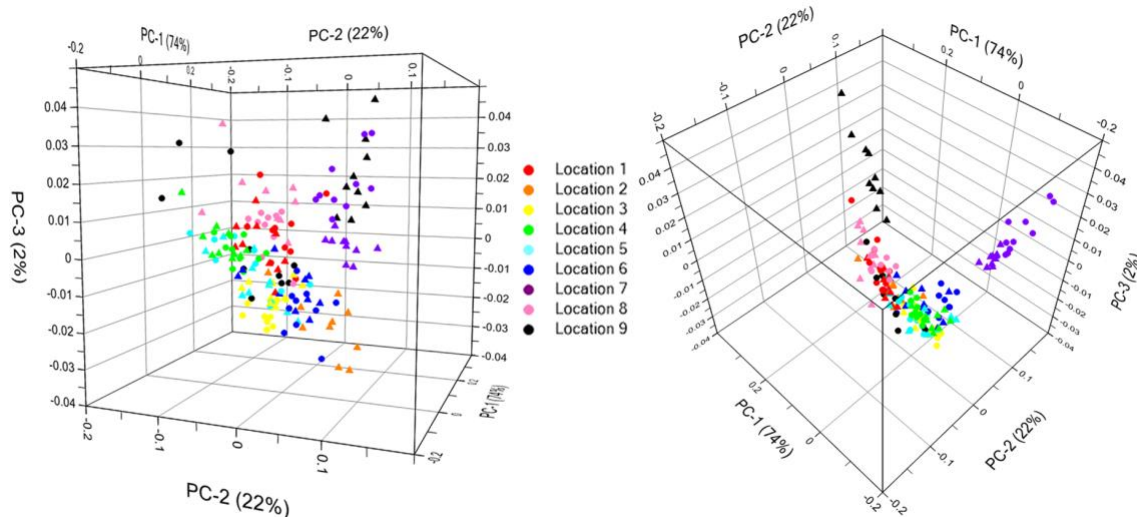

Figure 3. 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from different locations based on their corresponding MSP spectra, with individual 'a' samples represented by dots and 'b' samples represented by triangles.

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 stronger 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 9 soils contained the most intra-sample variability in their composition, with the samples spread out considerably across all PCs and overlapping with almost all the other location clusters on at least one PC. In this instance, using the full MSP data for chemometric analysis highlights the large degree of intra-variability within soil samples, due to their naturally occurring environments.

The factor loadings for the first three PCs (Fig. 4) can be studied to determine the regions in the MSP spectra that are associated with (and likely responsible for) the discrimination of samples along each component.
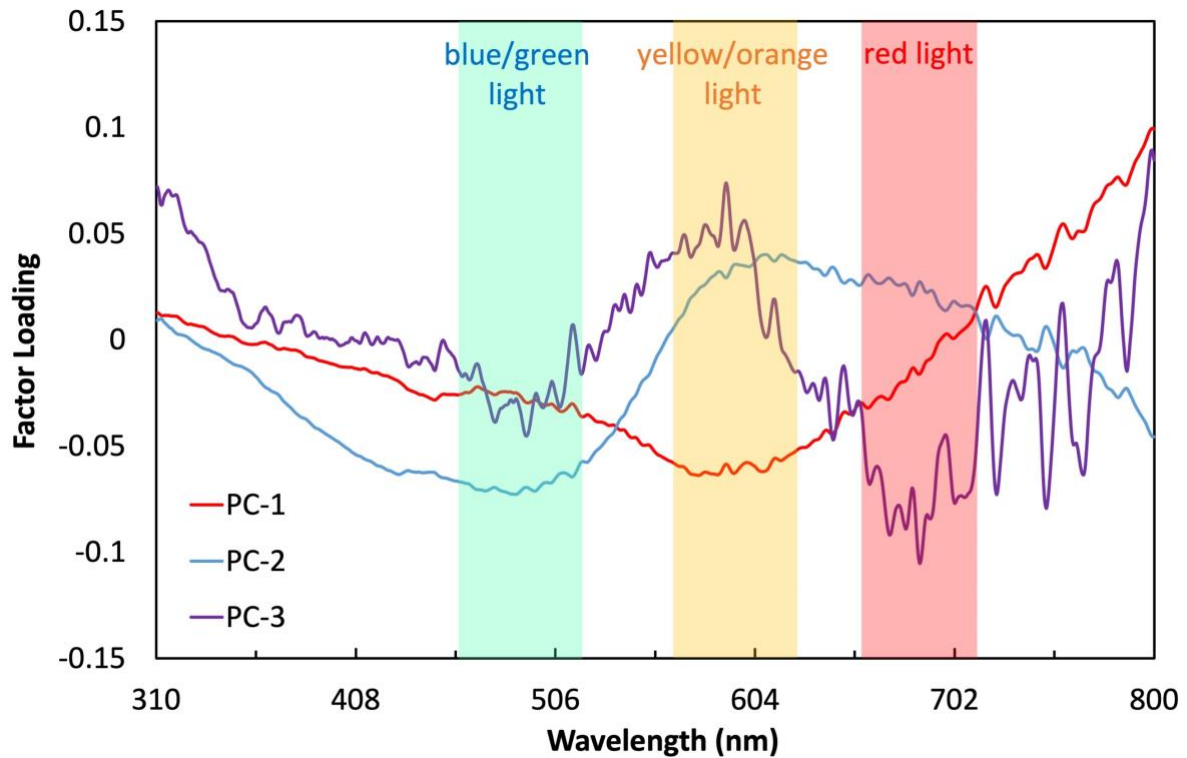
Figure 4. Factor loading plot of PCs 1-3 for PCA of the soil MSP reflectance dataset, with the main areas of interest highlighted.

The loadings plot for PC-1 has a global minima ca. 580 nm, associated with the reflectance of the yellow/orange region of the visible light spectrum. Samples separated along this PC may therefore contain different proportions of the yellow/orange components present in their soil. For example, all the yellow and red sands (locations 4, 5, and 7) are skewed along PC-1 in a negative direction. This is due to their high proportion of yellow/orange components combined with the negative association of PC-1 with this spectral region. The loading plots for PC-2 and PC-3 are very similar, revealing a negative correlation with ca. 490 nm, associated with the reflectance of blue/green light, and positive correlation ca. 600 – 620 nm, associated with the reflectance of orange light. The red soils from location 7 are best separated from the rest of the population along PC-2, while all the similarly coloured brown sands are situated nearest the red sands, also in a positive direction along PC-2. The loading plot for PC-3 also displays one global minima ca. 690 nm, associated with the reflectance of red light. However, these loadings exhibited a large degree of noise. It is questionable as to whether PC-3 (responsible for just 2% of the total variance) represents any systematic variation between samples, or simply noise in the dataset. When the scores plot was replotted utilising only PC-1 and PC-2 (Fig. 5), comparable separation was achieved to that from the 3D model, indicating that PC-3 does not significantly contribute to the separation of these samples.
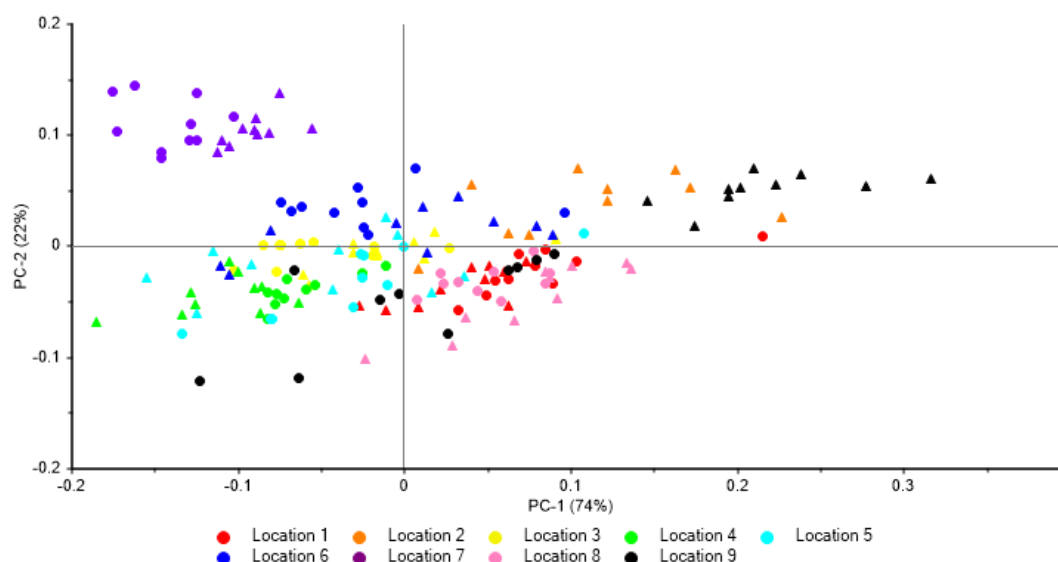
Figure 5. 2-dimensional PCA scores plot showing the variability of soil samples from different locations based on their corresponding MSP spectra, with individual 'a' samples represented by dots and 'b' samples represented by triangles.

Colour coding of the soil samples in the PCA model based on visually assigned groupings (outlined in Table 1) rather than sampling location, resulted in clustering based on their overall colour as expected (Fig. 6).
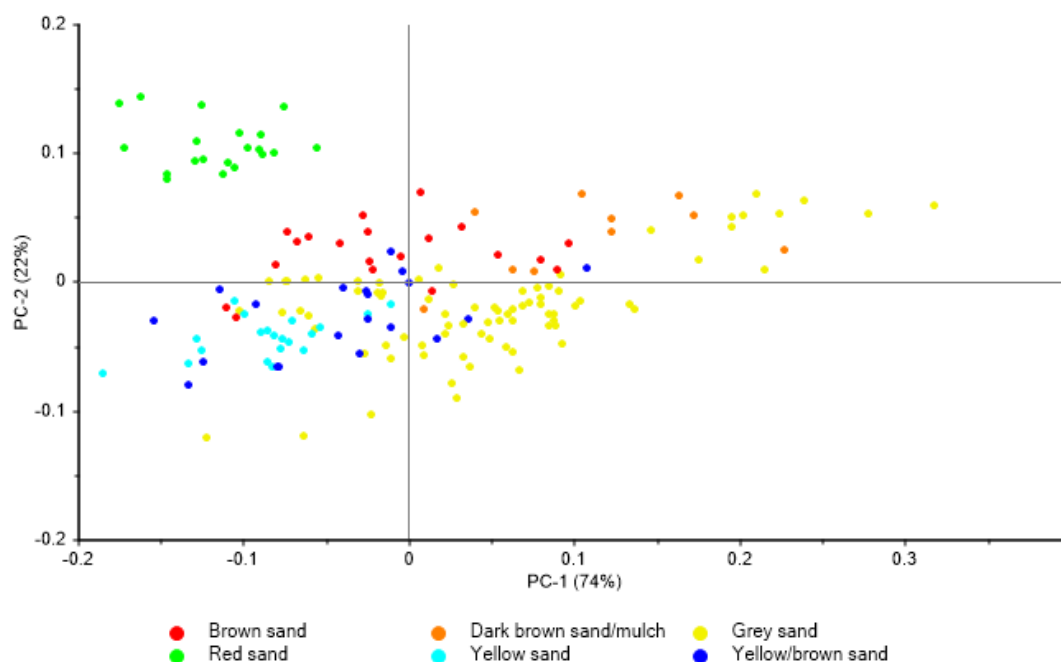

Figure 6. 2-dimensional PCA scores plot showing the variability in the visual appearance of soil samples based on their corresponding MSP spectra.

The red sands were again discriminated from the rest of the population, however, there was a great deal of overlap between the yellow and yellow/brown sands, yellow/brown and brown sands, and the brown and dark brown sands respectively. The grey sands were once again dispersed throughout the plot, overlapping with all clusters except the red sands. While the MSP spectra are very closely correlated with the visual appearance of the soils, they do 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. 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 (9). Unfortunately, even after isolation of the quartz fine fraction (< 20 μm), this micro-heterogeneity still poses difficulties with MSP, indicating that the only solution may be to use a macro-spectrophotometric approach instead.

3.3   Infrared spectroscopy

Use of IR spectroscopy has been hesitantly applied to naturally occurring heterogenous materials due to the large degree of variability within samples (4, 9). While studies have been conducted that make use of IR spectroscopy to analyse soil for forensic purposes, these have involved very different soil types to those found in the dry, sandy climate of Western Australia (9, 29, 38).

ATR-FTIR absorbance spectra (Fig. 7) were collected from the extracted quartz fine fraction of the same 17 soil samples outlined above.
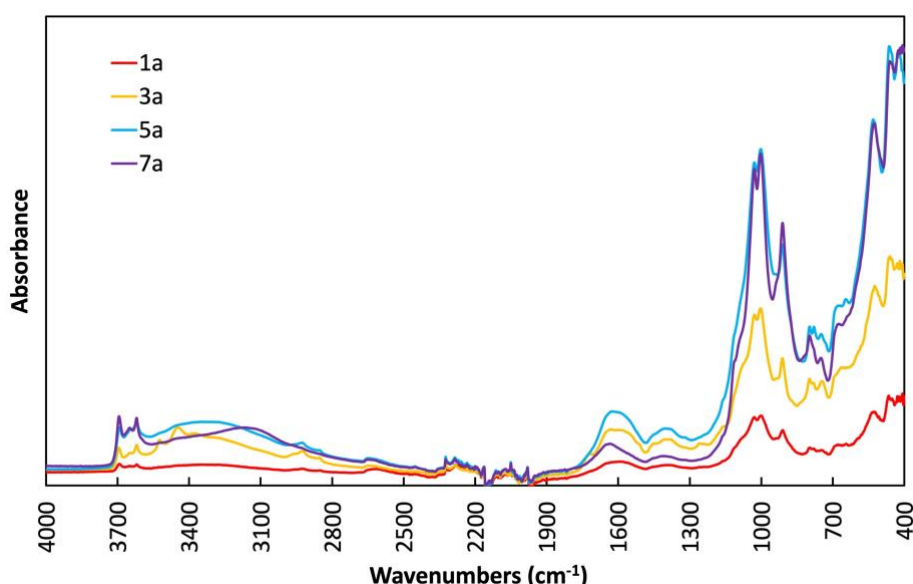

Figure 7. ATR-FTIR absorbance spectra showing the variability in composition of 4 soil samples collected from different locations within the Swan Coastal Plain. The region associated with the ATR-IR diamond (2350 – 1950 cm^{-1}), though present in these spectra, was removed prior to chemometric analysis.

Whilst some obvious differences can be seen between sample spectra, chemometric methods were applied to maximise the differentiation between samples. A PCA model was built using these spectra to highlight the variance in samples obtained from different locations. PCA revealed that 93.2% of the total variance in the dataset could be described by the first three PCs (Fig. A3). Three-dimensional score plots generated using these PCs (Fig. 8) resulted in most soil samples forming distinct clusters based on the location from which they were collected, with only a few locations displaying slight intra-site sample separation.
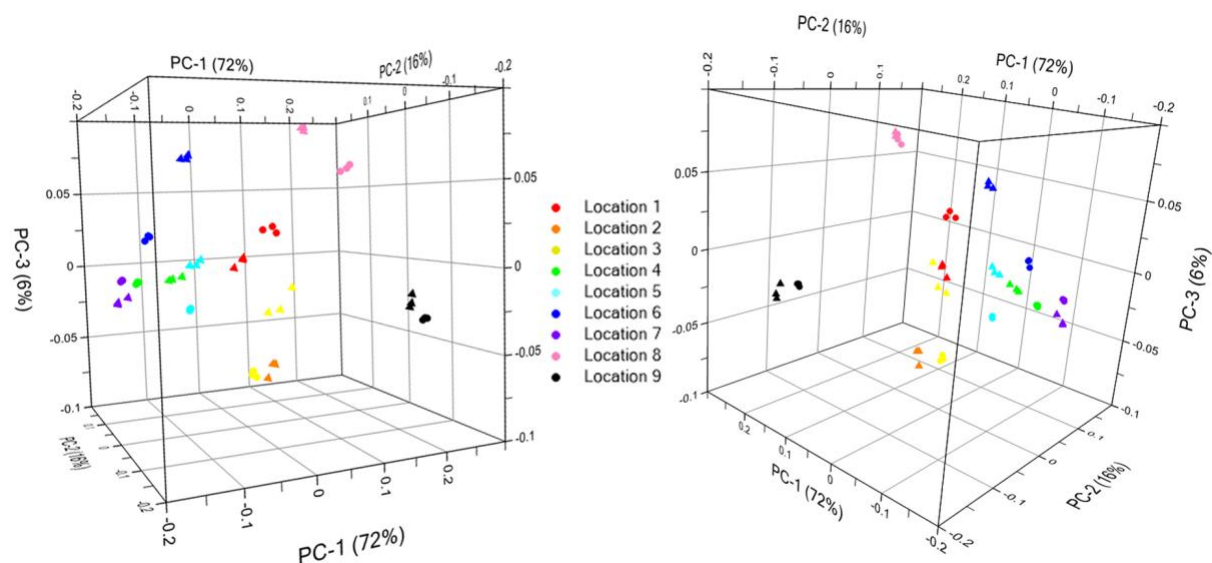
Figure 8. 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from different locations based on their corresponding ATR-FTIR spectra, with individual 'a' samples represented by dots and 'b' samples represented by triangles.

Soils from locations 3 and 6 exhibited the largest degree of separation between intra-site samples, forming two separate clusters based on individual samples rather than the overall location from which they originated. Both locations were accessible by the public and highly managed, with one being a public park and one a garden at a university campus.

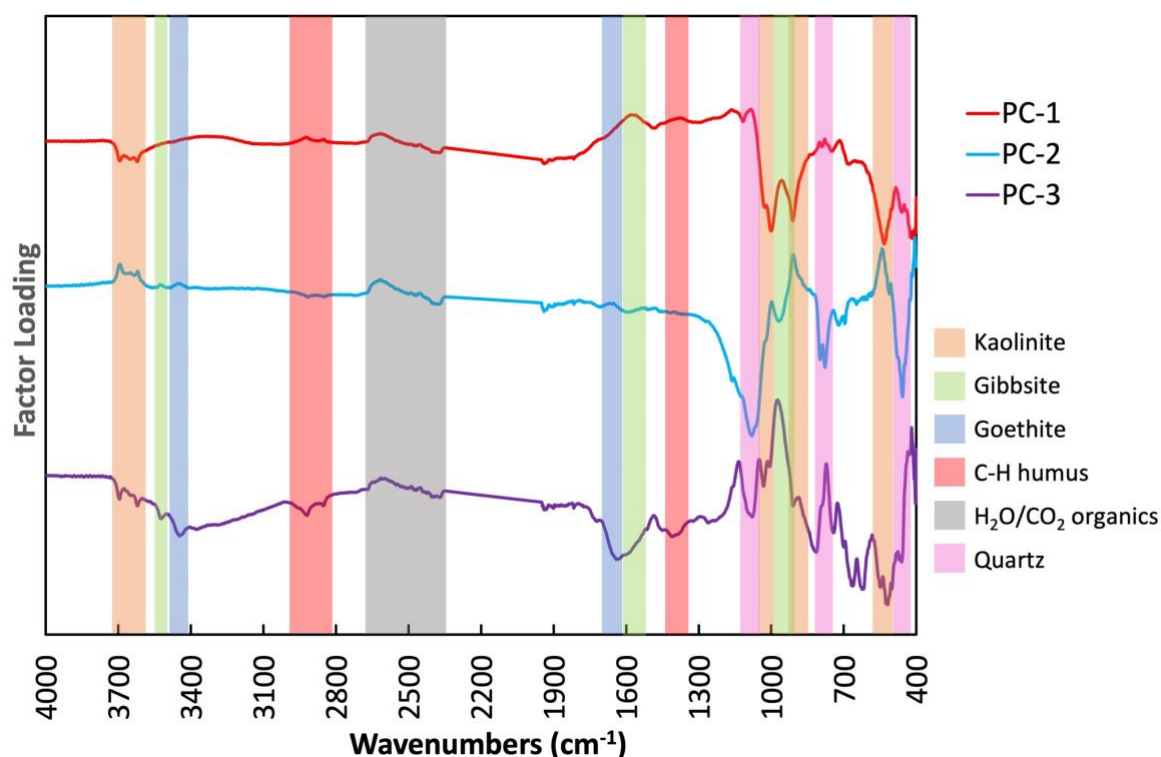The factor loadings for the first three PCs are shown in Fig. 9.

Figure 9. Factor loading plot of PCs 1-3 for PCA of the soil ATR-FTIR absorbance dataset, with the main peaks of interest highlighted and annotated with their contributing compounds.

Many of the peaks within the loadings are distorted due to a high degree of interference and overlap between positive and negative correlations, making mineral identification challenging. Minor variations in mineral content would create differences in the overall spectra, but it is difficult to completely identify the cause from the loadings alone. Identification was attempted based on the known IR spectra of common minerals and compounds found within Swan Coastal Plain soils, found in Table 2 below.

Table 2. Common minerals found in soil, 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)

[1]Thermo Fisher Scientific for Nicolet FTIR, "HR comprehensive Forensic FT-IR"
[2]Thermo Fisher Scientific, 2008, "HR Inorganics"
[3]SensIR technologies, 2001, "Common chemicals by Diamond ATR"

ATR-FTIR spectra obtained from representative samples situated at the extremes of each PC (Figures A4-A8) were used to inform the mineral associations made using the loadings plot. Variation along PC-1 is attributed to negative correlations with kaolinite, with positive contributions from quartz and possibly gibbsite/organics. Samples from location 7 are best separated along this PC in a negative direction, closely followed by locations 4, 6, and 5, alluding to greater relative levels of kaolinite in these soils. Samples from locations 8 and 9 are best separated in a positive direction on PC-1, due to having relatively higher concentrations of quartz. PC-2 suggests negative correlation with quartz, with possible smaller contributions from gibbsite and carbon-based organics. It also shows a positive correlation with kaolinite, with some variation seen in the spectra around the 1030 cm⁻¹, 1006 cm⁻¹, and 940 cm⁻¹ regions. Samples from location 9 are again the best separated from the rest of the population due to negative scores along PC-2, suggesting that they once again contain higher relative levels of quartz, and lower amounts of kaolinite in their quartz-recovered fine fraction. PC-3 has a positive correlation with quartz and possibly gibbsite, and negative correlations with possible iron oxides/hydroxides, C-H from organics (humus), and kaolinite. Samples from location 8 were skewed positively along PC-3 due to lacking kaolinite and containing relatively higher concentrations of quartz/gibbsite. Samples from location 6 were close behind, however, this may be due to sample heterogeneity as the other location 6 sample has a less positive score along PC-3. Samples from locations 2 and 3 (the latter again shows sample variability) were situated at the most negative point along PC-3, indicating that they may contain the highest levels of iron oxides/hydroxides, C-H from organics (humus), and kaolinite in their soil.

While tentative assessments can 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 2 are being detected through ATR-FTIR analysis. Using another analysis technique in sequence with ATR-FTIR, such as XRD, would allow for cross-referencing of the minerals being detected to confirm the above associations.

While the previous PCA model was able to distinguish samples based on their original locations, improved separation was investigated by attempting to reveal greater structure in the ATR-FTIR data based on other attributes, such as their visual appearance. Colour coding of the soil samples in the PCA model based on their visual appearance (outlined in Table 1) resulted in most of the distinct colour groups clustering similarly, except for the grey sands, which were dispersed across a larger area within the plot (Fig. 10).
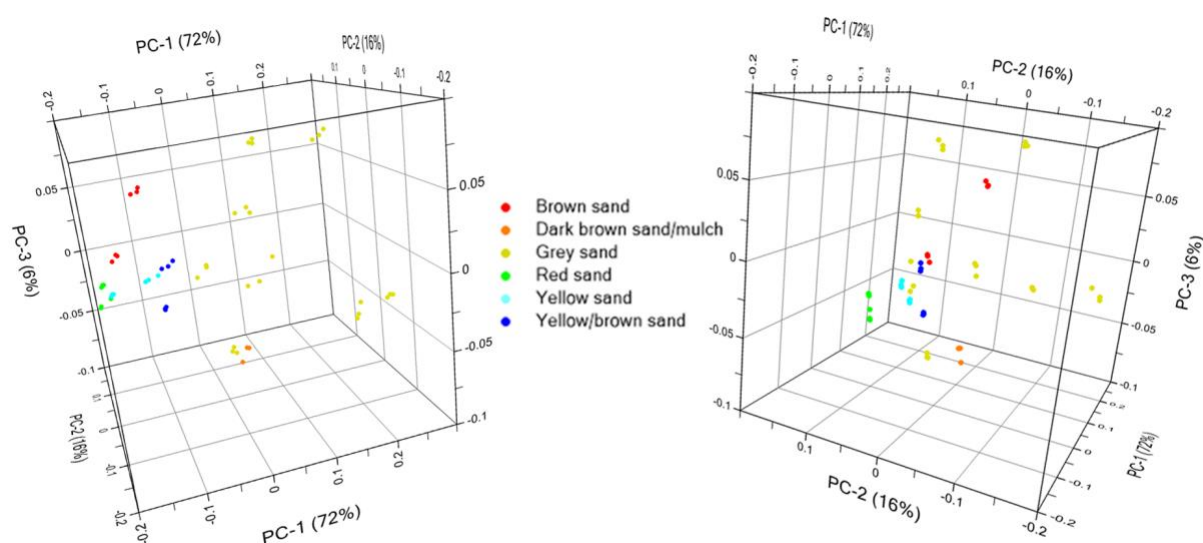

Figure 10. 3-dimensional PCA scores plot (shown from two perspectives) showing the variability in the visual appearance of soil samples based on their corresponding ATR-FTIR spectra.

Slight overlap was observed in clusters of similar colours, for example, yellow, yellow/brown, and red sands on PCs 1-3, as most of the separation along these PCs is potentially due to minerals that are responsible for the red and yellow hues of the soils, such as hematite, goethite, and kaolinite (often stained red brown from felsic). There was also a degree of intra-site variability, mainly within the brown sands. The yellow/brown sands were positioned in the model between the grey sands and the yellow sands on PC-1, with a small degree of overlap, highlighting their similarity with both groups. It is evident 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. A more precise determination of the colour of these soils could provide further information that would enhance separation of visually comparable soils.

MSP was therefore investigated as a grouping method for the PCA model generated from the ATR-FTIR spectra. Munsell colour values are routinely used by soil scientists for classifying soil (9, 17); 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 values consisting of letters and numbers

(39). This method was employed to allocate each soil sample a Munsell value generated from its MSP spectra. However, a difficulty arose in this case when attempting to group soil colours for interpretation of the PCA model, as any minor change in colour will alter the associated Munsell value and result in numerous individual qualitative classifiers that are challenging to categorise into groupings.

As classifiers for PCA must be input as a category to allow several samples to belong to each grouping, MSP spectra for each replicate were converted into 1931 L*a*b* values. The L*a*b* colour system is another method of measuring colour within three dimensional space that produces 3 numerical values that are better 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 (40). As most of the soil samples being analysed were made up of similar colours, with the most distinguishably coloured samples already individualised, the a* and b* values generated from their MSP spectra were relatively uniform, with the L* value changing the most significantly between different samples. For this reason, the average L* value for each sample was obtained from its 10 replicate spectra, and this was used to classify each sample according to the information in Table 3.

Table 3. The soil samples analysed with MSP and their associated L* values and assigned categories.

Soil sample	L* values from MSP	Average L* value category	Visual Appearance
1a	56.14	54-58	Grey sand
1b	56.60	54-58	
2b	57.79	54-58	Dark brown sand/mulch
3a	62.94	62-66	Grey sand
3b	57.03	54-58	
4a	59.70	58-62	Yellow sand
4b	60.47	58-62	
5a	55.78	54-58	Yellow/brown sand
5b	63.24	62-66	
6a	68.38	66-70	Brown sand
6b	68.45	66-70	
7a	71.99	>70	Red sand
7b	71.71	>70	
8a	52.49	50-54	Grey sand
8b	45.89	<50	
9a	53.15	50-54	Grey sand
9b	52.18	50-54	

Colour coding of the samples within the ATR-FTIR PCA model was then carried out based on their L* values obtained from the MSP spectra (Fig. 11).
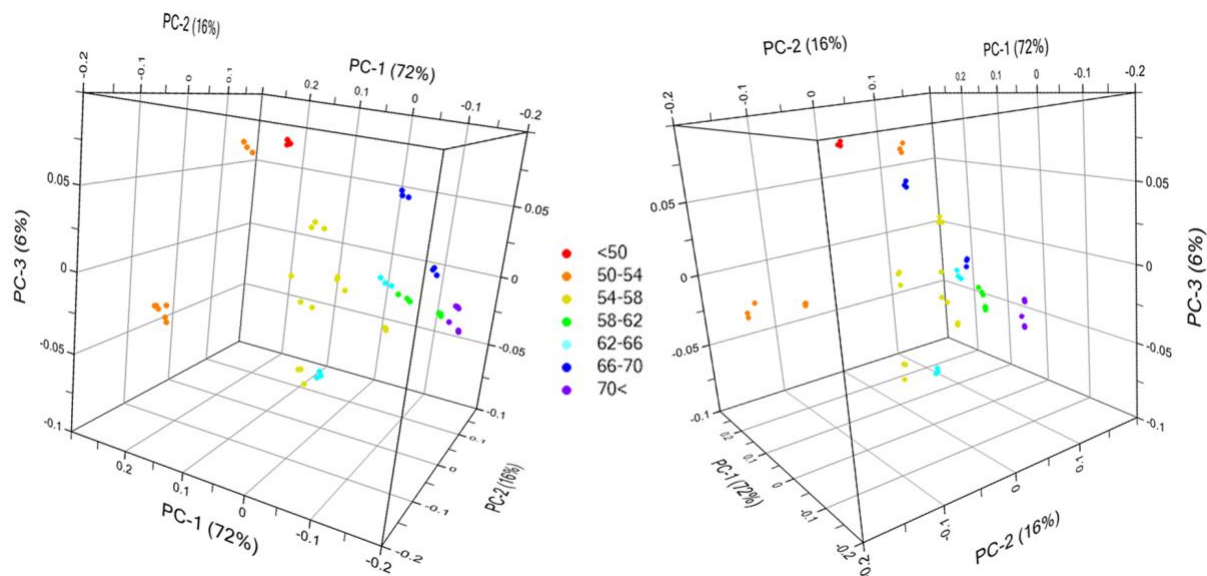
Figure 11. 3-dimensional PCA scores plot (shown from two perspectives) based on the ATR-FTIR spectra of soil samples, showing the variability in the L* values obtained from corresponding MSP spectra from the same soils.

This resulted in most of the samples with similar L* values clustering near each other with limited overlap. In most instances however, several groups were formed within one L* value category, instead of clustering based on sample replicates. These intra-class samples exhibited slight separation from each other, notably within the 50-54 group which is distanced considerably across PCs 2 and 3. This is most likely due to an outlier originating from location 8, as the samples are clustered close to their location counterparts in the <50 category, despite the same site generating two different L* values. The 62-66 samples were also substantially separated from each other across PCs 2 and 3, due to locations 3 and 5 having been assigned two different L* value categories for their intra-site samples, highlighting the degree of colour heterogeneity between soils located just centimetres apart. This result was, at best, equal to that obtained through the grouping of the ATR-FTIR dataset using visual appearances of the soil, indicating that L* values are not necessarily able to give any further information on the colour of soils that is not already distinguishable with the naked eye. However, it is likely that information is potentially being lost when transforming these complex MSP spectra into single L* values, as spectra are generalised into patterns when being converted into numeric values and as a consequence, two slightly different spectra may be classified as the same colour (41). The data provided within MSP spectra may still be useful for conducting pairwise sample comparisons, by allowing for differentiation of samples that cannot otherwise be distinguished by eye or by colourimetric values such as L* a* b*.

3.4   X-ray diffraction

X-ray diffraction (XRD) was applied as a non-destructive method that can provide both qualitative and semi-quantitative information on the inorganic material present within a small sample (42). XRD patterns (Fig. 12) were collected from the extracted quartz fine fraction of the full 18 soil samples outlined in Table 1.
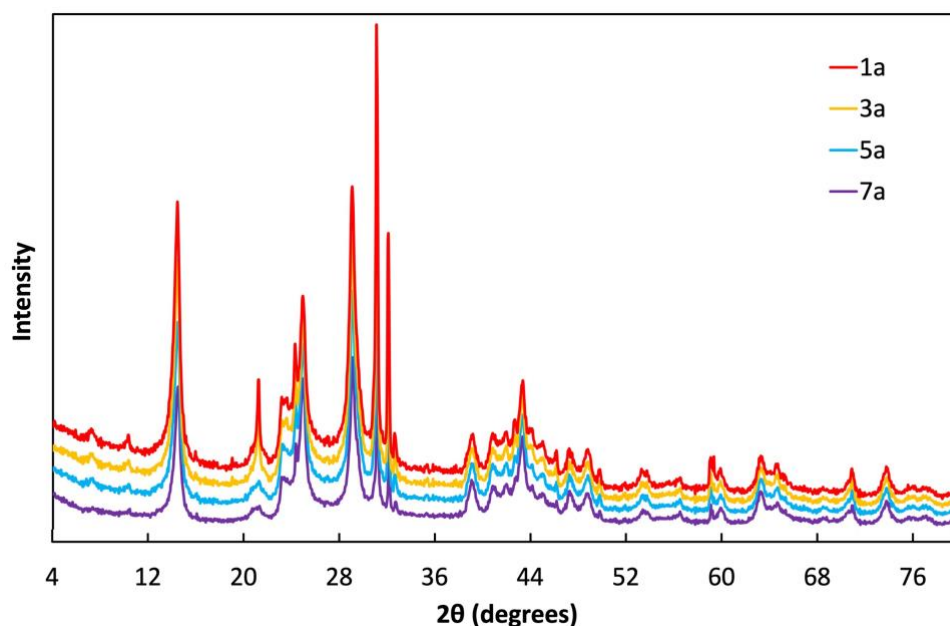
Figure 12. XRD patterns showing the variability in composition of 4 soil samples collected from different locations within the Swan Coastal Plain (stacked for increased visibility). Low background plate peaks (narrow regions centred on approximately 18.9, 38.5, and 59.3 degrees 2θ), though present in these spectra, were removed prior to chemometric analysis.

These patterns emphasise the high level of similarity between samples, and with it, the advantage of utilising chemometric analysis to identify subtle differences within a sample set that cannot be recognised visually by the examiner. This data was used to build a PCA model; PCA revealed that 93.6% of the total variance in the dataset could be described by the first four PCs, however upon visualisation of the influence of PC-4 on the scores plots, it did not improve the discrimination of soils, and hence the first three PCs (accounting for 89.5% of the variance) were utilised for the PCA model (Fig. A9). Three-dimensional score plots generated using the first three PCs (Fig. 13) resulted in most of the soils clustering based on the locations from which they were collected, however, locations 1, 2 and 4 showed some intra-site variability, with samples from the same location separating within the model.
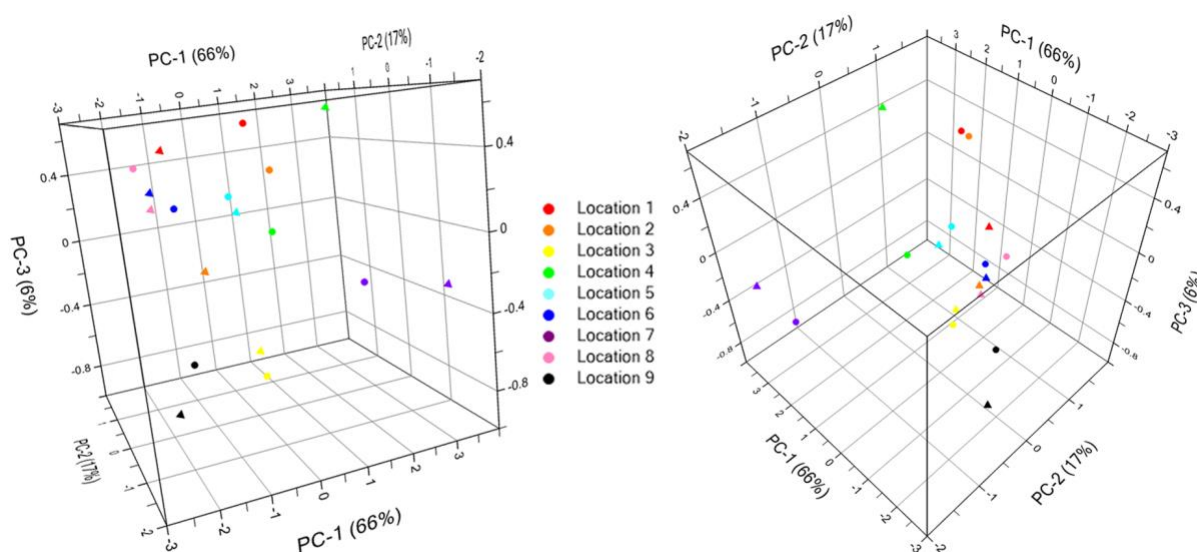
Figure 13. 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from different locations based on their corresponding XRD patterns, with individual 'a' samples represented by dots and 'b' samples represented by triangles.

These three sites were both easily accessible and subject to a high degree of human interference, e.g. a public park or roadside verge, with two of the locations highly managed. It is therefore logical to draw the conclusion that human interference is contributing to the variation shown across a significant number of these sites. Whilst the intra-site disparity and the resulting degree of overlap with other locations indicated a high variability in the composition of these soils, the separation between associated samples was not substantial. For example, the two samples from location 1 were separated from each other along PC-1, and samples from locations 2 and 4 were separated from their associated site samples along both PC-1 and PC-3. In contrast to the ATR-FTIR model, XRD achieved improved discrimination of the samples from locations 4, 5, 6 and 7. Modelling based on ATR-FTIR spectra had these clusters situated very close together, whereas using the XRD data allowed for enhanced separation of nearly all these groups, with all intra-site samples still clustered closely, except for location 4. The XRD model also managed to isolate location 3 soils from location 2 soils, which overlapped in the ATR-FTIR model. 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, especially when utilised in sequence with other techniques. This is an advancement from the work of Pitts and Clarke (21), who utilised percentage intensities for selected minerals.

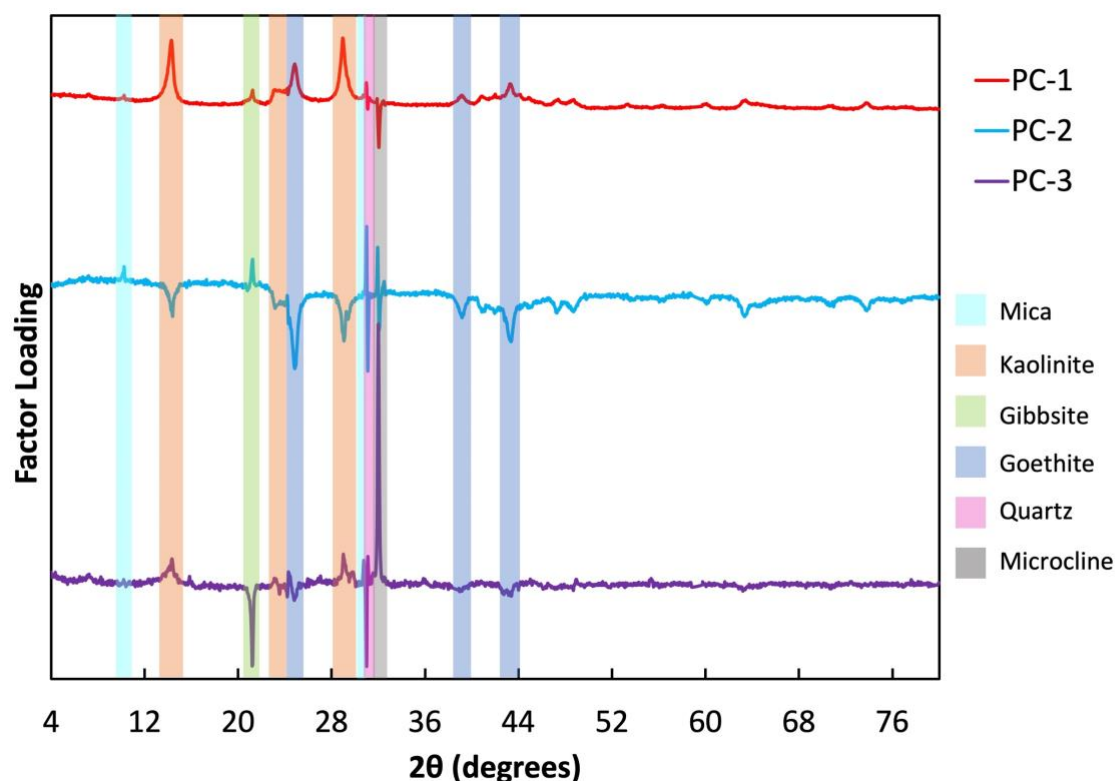The factor loadings for the first three PCs are shown in Fig. 14.

Figure 14. Factor loading plot of PCs 1-3 for PCA of the soil XRD dataset, with the main peaks of interest highlighted and annotated with their contributing minerals.

There is some uncertainty stemming from identification based on the XRD pattern alone, as kaolinite and halloysite are difficult to resolve and the presence of halloysite is questionable in sandy soils. Hence using the full pattern into PCA to capture all the information present is valuable, so that complete identification is not essential. The loading plot for PC-1 exhibits an overall positive correlation with kaolinite, with minor contributions from gibbsite and goethite, and a negative correlation with microcline feldspar. Samples from location 7 are best separated along PC-1 in a positive direction, and therefore may be assumed to have higher concentrations of kaolinite in their soil. The loading plot for PC-2 appears to be positively influenced by mica with possible contributions from quartz, and negatively influenced mostly by goethite and kaolinite (quartz and microcline feldspar are correlated with both positive and negative loadings across PC-2, which could potentially be a normalisation issue, negating their influence on the data). Location 7 soils are once again the most skewed by PC-2, this time in a negative direction, indicating higher amounts of goethite and kaolinite in these samples, consistent with PC-1. The loading plot for PC-3 is significantly positively correlated with microcline feldspar, with minor contributions from kaolinite (positive) and quartz (negative). Samples from locations 3 and 9 are best separated along this PC in a negative direction, and therefore may be assumed to have higher concentrations of quartz. Soils from location 7 are positioned centrally on PC-3, while all other locations again exhibit a positive trend, indicating higher concentrations of microcline feldspar within their soils.

The inclusion of samples from additional sites could help to increase separation. For example, aluminium substitution for iron within goethite will shift the positioning of goethite peaks, potentially allowing for greater differentiation between soils containing variably substituted goethite. 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. 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.

As XRD specifically identifies the minerology of the soil samples, the PCA model was re-visualised (Fig. 15) to present the samples grouped according to the dune system within the Swan Coastal Plain from which they originated (outlined in Table 1).
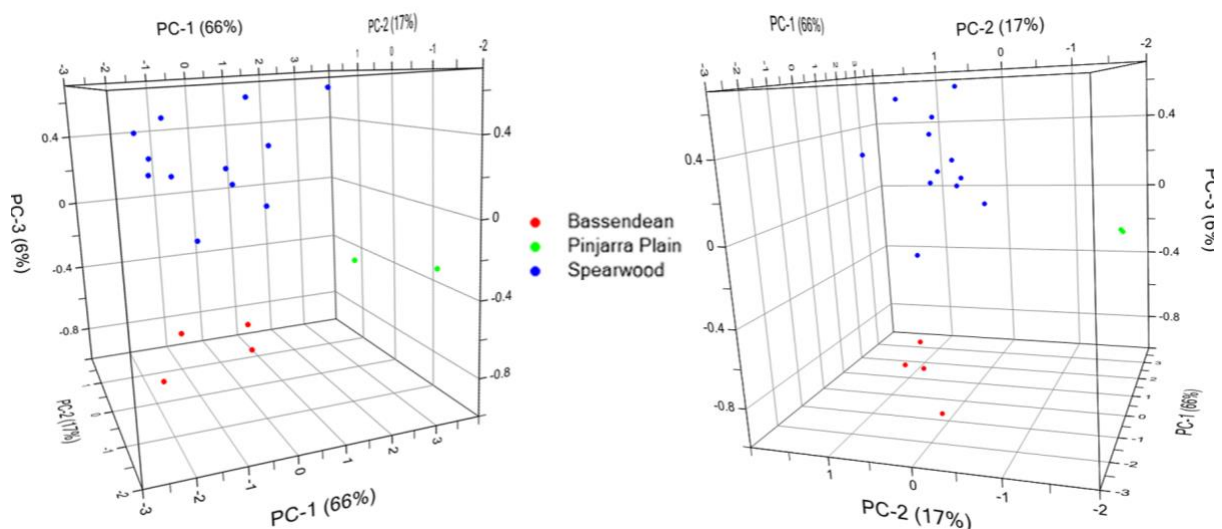

Figure 15. 3-dimensional PCA scores plot (shown from two perspectives) showing the variability of soil samples from differing dune systems based on their corresponding XRD patterns.

Full visual separation between the Spearwood, Bassendean, and Pinjarra Plain 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 dune systems, were shown to contain higher relative levels of kaolinite in comparison to other samples. The Pinjarra Plain is situated atop a bed of kaolinite clay, and hence these soils can contain relatively high concentrations of this mineral. Bassendean soils are situated between the Pinjarra Plain and Spearwood dunes, and were indicated to contain higher relative concentrations of quartz, a trend that has been observed in recent studies (21). Samples collected from the Spearwood dunes are also consistent with this previous study, containing higher levels of microcline feldspar and mica than their easterly counterparts (21).

Taking the same approach as with the ATR-FTIR data, further grouping of samples in the PCA model based on their visual appearance (outlined in Tables 1 & 3) resulted in a few samples clustering close by their equally coloured counterparts, notably the yellow/brown sands and brown sands, as well as full separation of the red sands (Fig. 16).
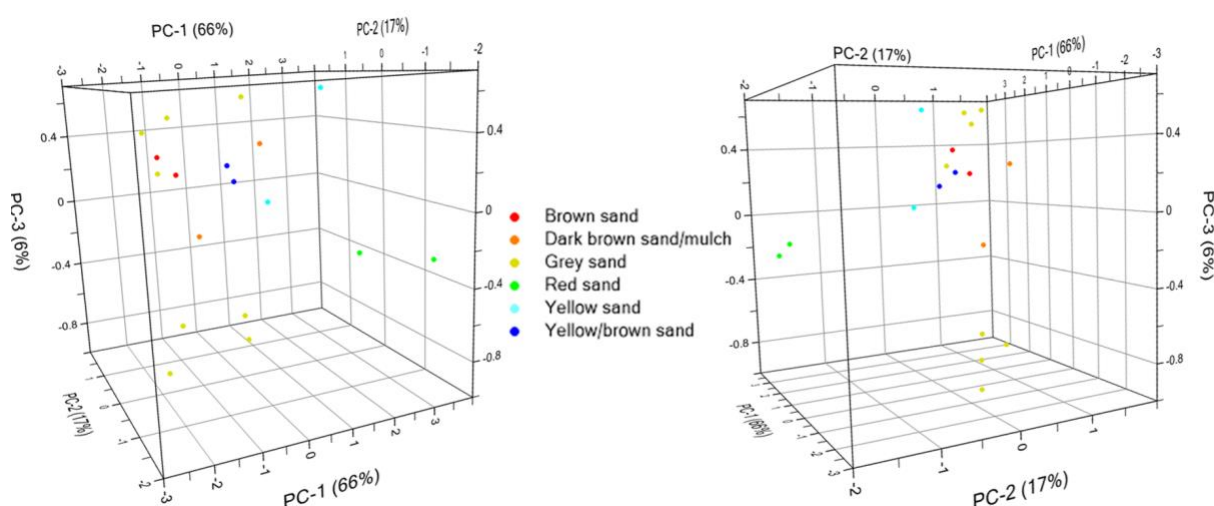

Figure 16. 3-dimensional PCA scores plot (shown from two perspectives) showing the variability in the visual appearance of soil samples based on their corresponding XRD patterns.

The presence of some distinctively coloured minerals such as kaolinite and goethite 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, as 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 Fe 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. For this reason, classifying the data based on their associated MSP L* values (outlined in Table 3) also failed to improve the discrimination of samples using our method (Fig. A10). The use of MSP calculated L* values for differentiation of soils could still be valuable for future studies that utilise larger sample sizes or different soil types.

3.5   Analytical Sequence

As indicated in the methods, when focusing on the inorganic content of forensic soils, using several non-destructive techniques in combination allowed for more accurate identification of minerals and separation of soils from similar locations. These methods each provided information that builds on knowledge obtained from a previous technique; taken in isolation, each analysis method provided only part of the entire picture.

PCA using the MSP spectra highlighted the large degree of variability within natural samples and, whilst achieving some clusters based on soil locations and visual appearance, did not allow for effective separation between locations. As expected, the MSP spectra were closely correlated with the visual appearance of the soils, essentially providing a more objective determination of the colour of the sample. However, when converting to L* a* b* colorimetric values this 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 for PCA in this instance. Nevertheless, MSP analysis is still valuable for forensic pairwise comparisons and may show value with highly coloured soils.

Using the ATR-FTIR spectra resulted in tighter clustering of intra-site soils than any other method, with better separation between locations achieved. When samples in the model were grouped based on visual appearances, clusters were formed, albeit with a small amount of overlap between classes of similar colours. Separation also occurred within the colour classes based on individual locations. The compounds responsible for most of the variance within the dataset correlated both with the original location of the sample, as well as the overall visual appearance of the sample.

PCA using the XRD patterns allowed for separation of most soils; the factor loadings indicated that the minerals most influencing separation were correlated with the dune system that the sample originated from and, though wide-spread, distinct clustering and full class separation were achieved using this method. The XRD model was also able to discriminate between samples that were left overlapping within the ATR-FTIR model, demonstrating how these methods can be used in combination with each other to provide additional information and increased separation.
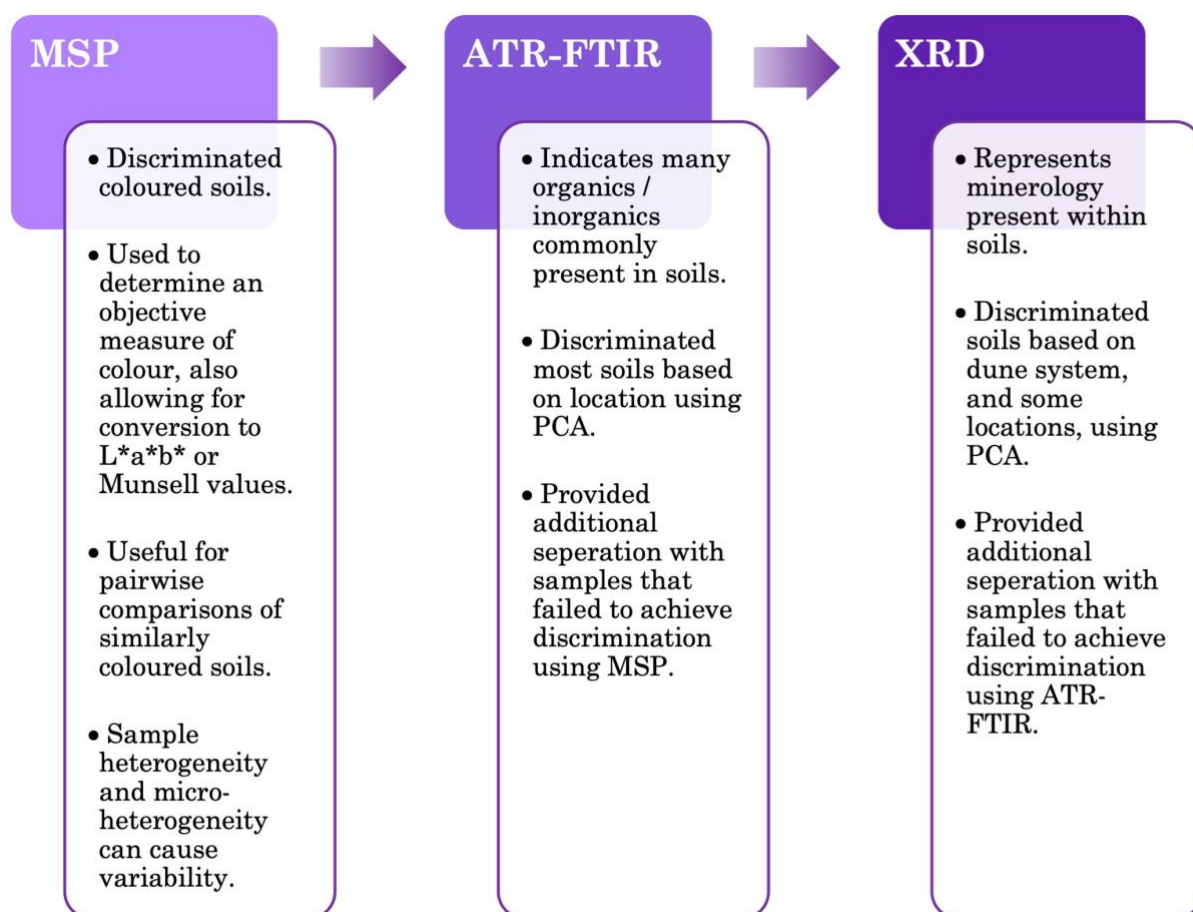
Figure 17. Analytical sequence for examination of the quartz fine fraction of sandy forensic soil samples from the Swan Coastal Plain in Perth, Western Australia, with the information obtained through analysis at each stage.

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. Sandy soils such as those found in the Swan Coastal Plain contain minimal organics, which reduces the breadth of instrumentation available for their analysis.

4.0   Conclusions and Future Work

Examination of the quartz-recovered fine fraction of Perth sandy soils has demonstrated that chemometrics can be used successfully in combination with spectroscopic techniques to discriminate between sandy soils. Analysis of these soils was especially challenging due to their very low levels of clay and organic matter, and the data generated through inorganic methods were difficult to objectively differentiate based on visual examination of the spectra, requiring more advanced statistical analyses.

This approach provides good evidence that several analysis methods utilised in combination, with chemometrics performed alongside each stage of the sequence, can maximise the differentiation obtained by forensic practitioners when examining inorganic forensic soil samples. The method explored in this study not only provides a more objective interpretation of the examination of forensic soils, of which the significance has been highlighted many times within the forensic community (26, 27), but also offers a

statistical basis to determine the limits of performance. Additionally, by utilising a particularly challenging set of samples, detailed data is being 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 (10).

While this study is transferable to other jurisdictions with dry, sandy soils, this research should be explored further by utilising a larger sample population and conducting more detailed statistical analysis of the data by performing other post-PCA methods, such as linear discriminant analysis. As previously shown with other forms of forensic evidence, e.g. fibres, this chemometric approach is also suitable for use with other statistical interpretation methods, such as a Bayesian interpretation or traditional 'question vs known' examinations (28, 43, 44).

Author Contributions

Talia Newland: Conceptualisation, Methodology, Investigation, Formal Analysis, Visualization, Writing – Original Draft
Kari Pitts: Conceptualisation, Supervision, Resources, Writing – Review & Editing
Simon W. Lewis: Conceptualisation, Project administration, Supervision, Visualization, Writing – Review & Editing

Acknowledgments

References

1. 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;26:21-9.
2. McCulloch G, Dawson LA, Brewer MJ, Morgan RM. The identification of markers for Geoforensic HPLC profiling at close proximity sites. Forensic Science International. 2017;272:127-41.
3. Testoni SA, Melo VF, Dawson LA, da Silva Salvador FA, Kunii PA. Validation of a Standard Operating Procedure (SOP) for Forensic Soils Investigation in Brazil. Rev Bras Cienc Solo. 2019;43(e0190010).
4. Prandel LV, Melo VF, Testoni SA, Brinatti AM, Saab SdC, Dawson LA. Spectroscopic techniques applied to discriminate soils for forensic purposes. Soil Research. 2020;58:151–60.
5. 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):350-3.
6. 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):403-15.
7. Woods B, Paul Kirkbride K, Lennard C, Robertson J. Soil examination for a forensic trace evidence laboratory - Part 2: Elemental analysis. Forensic Science International. 2014;245:195-201.
8. Ritz K, Dawson L, Miller D, editors. Criminal and Environmental Soil Forensics. Netherlands: Springer Science + Business Media; 2009.

9. Woods B, Lennard C, Kirkbride KP, Robertson J. Soil examination for a forensic trace evidence laboratory - Part 1: Spectroscopic techniques. Forensic Science International. 2014;245:187-94.

10. National Institute of Forensic Science Australia New Zealand. Research and Innovation Roadmap. Australia & New Zealand: Australia New Zealand Policing Advisory Agency (ANZPAA); 2020.

11. Western Australian Planning Commission. Perth Coastal Planning Strategy. Perth: Department of Planning; 2008.

12. Bastian LV. Residual soil mineralogy and dune subdivision, Swan Coastal Plain, Western Australia. Australian Journal of Earth Sciences. 1996;43(1):31-44.

13. Middle G. Geomorphology of Swan Coastal Plain: Garry Middle (VisionEnvironment); 2014 [Available from: http://www.garrymiddle.net/geomorphology-of-swan-coastal-plain/.

14. Earth Science Western Australia. Swan Coastal Plain. 2016.

15. Pitts K, Lewis S, Newland T. Forensic Analysis of Soil Evidence. In: Worsfold P, Townshend A, Poole C, Miró M, editors. Encyclopedia of Analytical Science. 3 ed: Elsevier; 2018.

16. 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):24-30.

17. Petraco N, Kubic TA, Petraco ND. Case studies in forensic soil examinations. Forensic Science International. 2008;178(2-3):e23-7.

18. 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;96:33-42.

19. 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;270:153-8.

20. 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.

21. Pitts KM, Clarke RM. The forensic discrimination of quartz sands from the Swan Coastal Plain, Western Australia. Forensic Science International: Reports. 2020;2(100130).

22. 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).

23. 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):178-92.

24. 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):37-44.

25. 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):52-63.

26. 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 E. Support for Forensic Science Research: Improving the Scientific Role of the National Institute of Justice. Washington, DC; 2015.

27. Science and Technology Select Committee. Forensic science and the criminal justice system: a blueprint for change. UK: House of Lords; 2019.

28. 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;2:15-21.

29. Chauhan R, Kumar R, Sharma V. Soil forensics: A spectroscopic examination of trace evidence. Microchemical Journal. 2018;139:74-84.

30. Sauzier G, van Bronswijk W, Lewis SW. Chemometrics in forensic science: approaches and applications. Analyst. 2021;146(8):2415-48.

31. 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 and Justice. 2011;51:77-82.

32. 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;288:310-26.

33. 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;251:22-31.

34. 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).

35. Kammrath BW, Koutrakos A, Castillo J, Langley C, Huck-Jones D. Morphologically-directed Raman spectroscopy for forensic soil analysis. Forensic Science International. 2018;285:e25-e33.

36. Jantzi SC, Almirall JR. Characterization and forensic analysis of soil samples using laser-induced breakdown spectroscopy (LIBS). Analytical and Bioanalytical Chemistry. 2011;400:3341–51.

37. Fitzpatrick RW. Soil: forensic analysis. In: Jamieson A, Moenssens A, editors. Wiley Encyclopaedia of Forensic Science. Chichester: John Wiley & Sons; 2009. p. 2377–88.

38. Cox RJ, Peterson HL, Young J, Cusik C, Espinoza EO. The forensic analysis of soil organic by FTIR. Forensic Science International. 2000;108:107-16.

39. Islam K, McBratney AB, Singh B. Estimation of soil colour from visible reflectance spectra. SuperSoil: 3rd Australian New Zealand Soils Conference; Sydney, Australia. www.regional.org.au/au/asssi/: University of Sydney; 2004.

40. 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.

41. Investigation FBO. Standard Guide for Microspectrophotometry and Color Measurement in Forensic Paint Analysis. USA: Forensic Science Communications; 2007.

42. 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):13-23.

43. Garbolino P, Taroni F. Evaluation of scientific evidence using Bayesian networks. Forensic Science International. 2002;125:149-55.

44. Norgaard A, Rasmusson B. The likelihood ratio as value of evidence-more than a question of numbers. Law, Probability, and Risk. 2012;11:303-15.
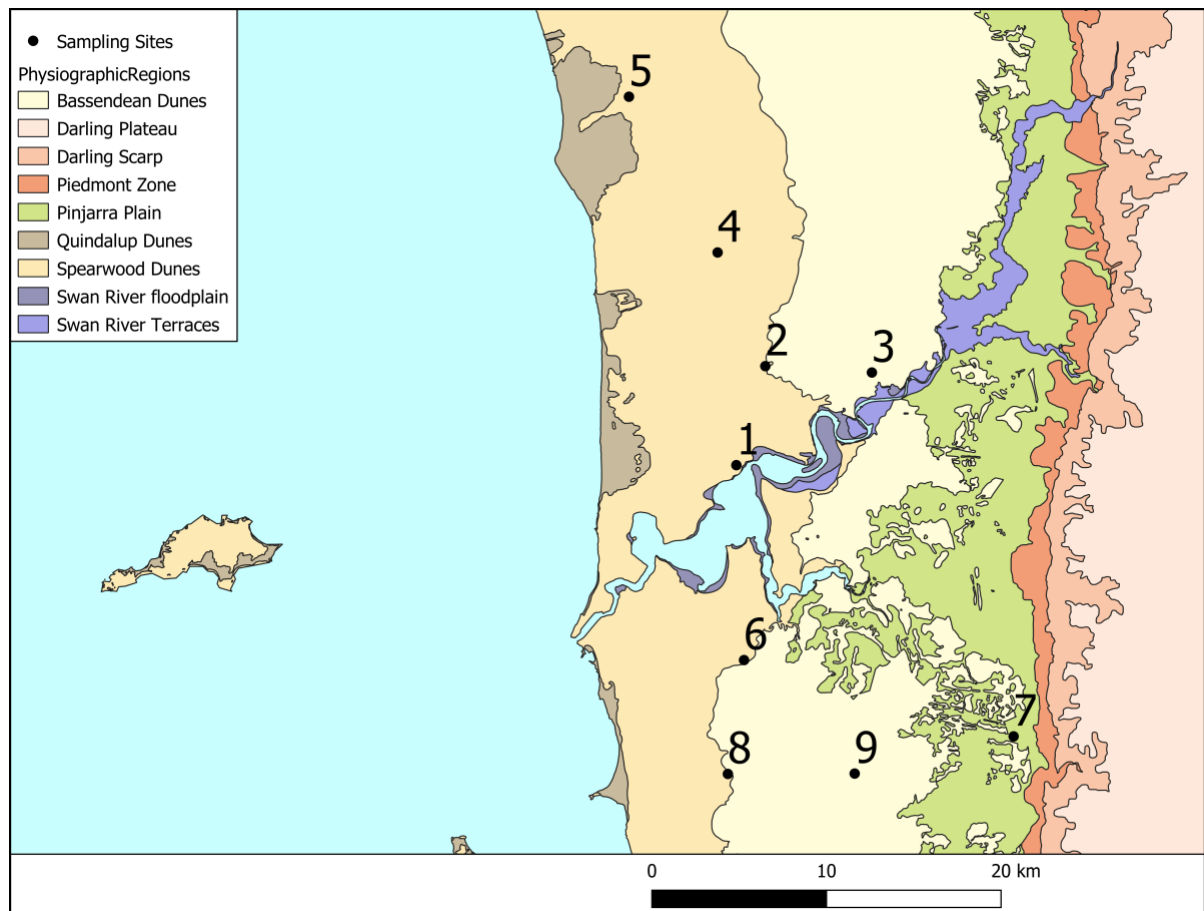
Figure A1. Map of the Swan Coastal Plain in Perth, Western Australia, illustrating the 9 locations that the soil samples were collected from for use within this study. GPS coordinates of sample locations were overlaid onto 'Physiographic regions' from Gozzard JR. Sea to scarp — geology, landscape, and land use planning in the southern Swan Coastal Plain. Geological Survey of Western Australia; 2010. p. 72. ISBN 978-74168-251-9. Image produced within QGIS 3.18.1-Zürich.

Table A1. GPS coordinates of the 9 locations that the soil samples were collected from for use within this study.

Location	Latitude	Longitude
1	-31.9646960°	115.8357950°
2	-31.9138106°	115.8539786°
3	-31.9176633°	115.9184918°
4	-31.8549414°	115.8258240°
5	-31.7741340°	115.7732363°
6	-32.0652384°	115.8391782°
7	-32.1059759°	116.0022922°
8	-32.1238569°	115.8286284°
9	-32.1243506°	115.9056052°
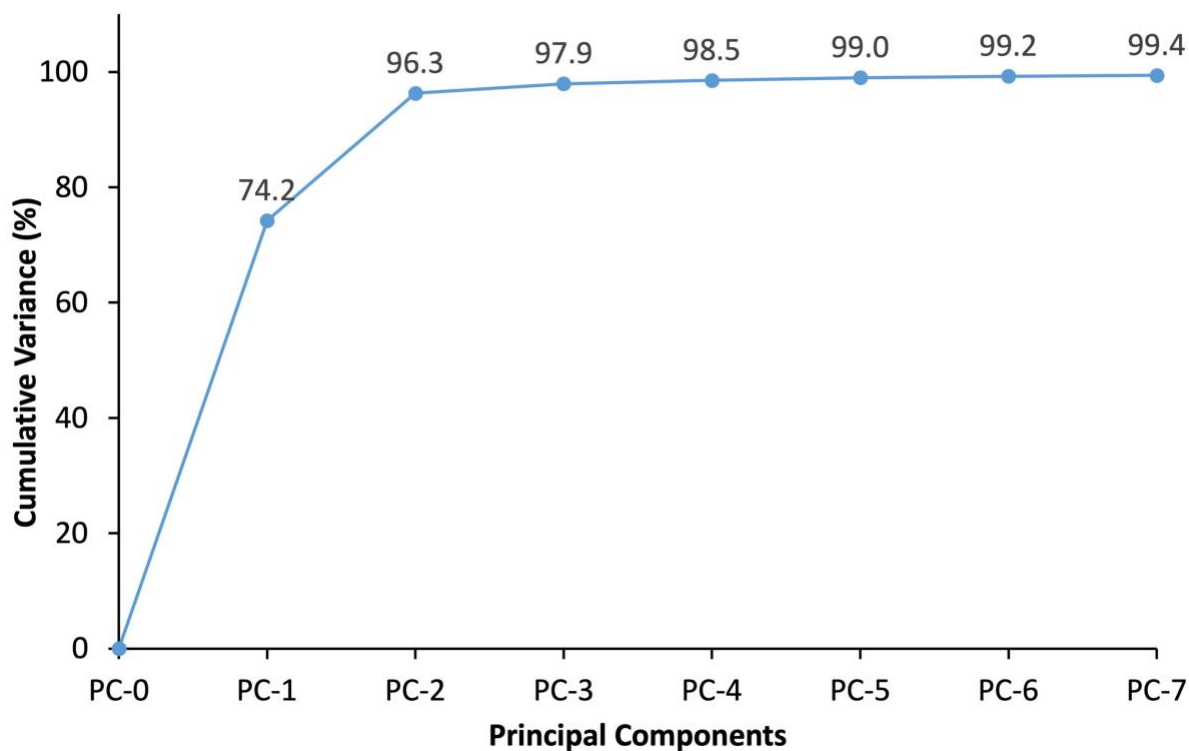
Figure A2. Scree plot depicting the cumulative variance in the MSP dataset retained by each PC.
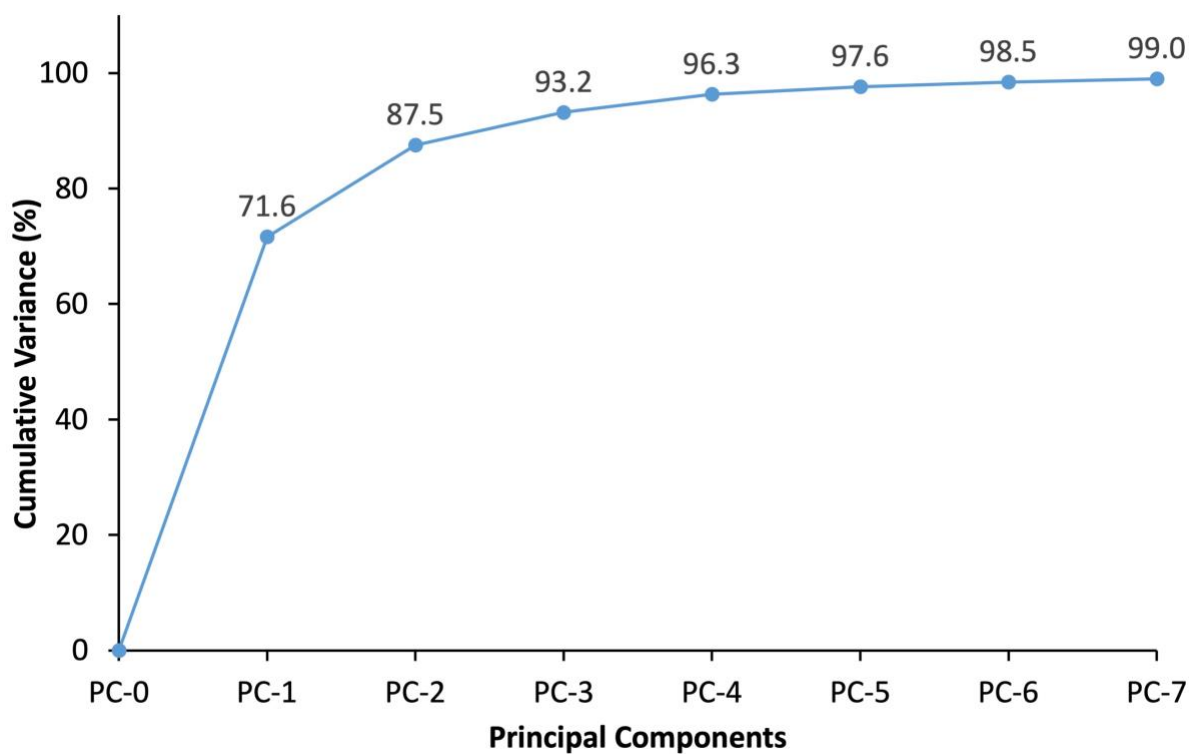
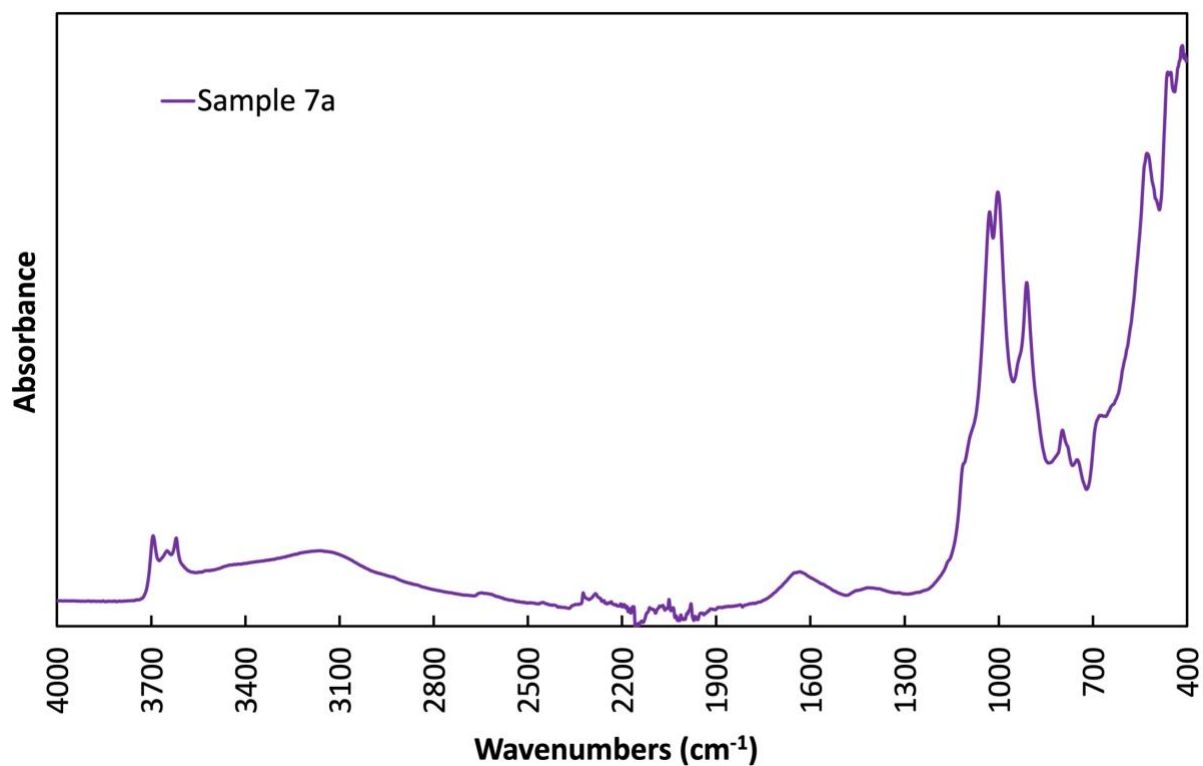**Figure A3.** Scree plot depicting the cumulative variance in the ATR-IR dataset retained by each PC.

Figure A4. Representative raw ATR-FTIR spectra of sample 7a, which is situated at the most negative value of PC-1 within the ATR-FTIR PCA scores plots. The region associated with the ATR-IR diamond ($2350 - 1950$ cm^{-1}), though present in these spectra, was removed prior to chemometric analysis.
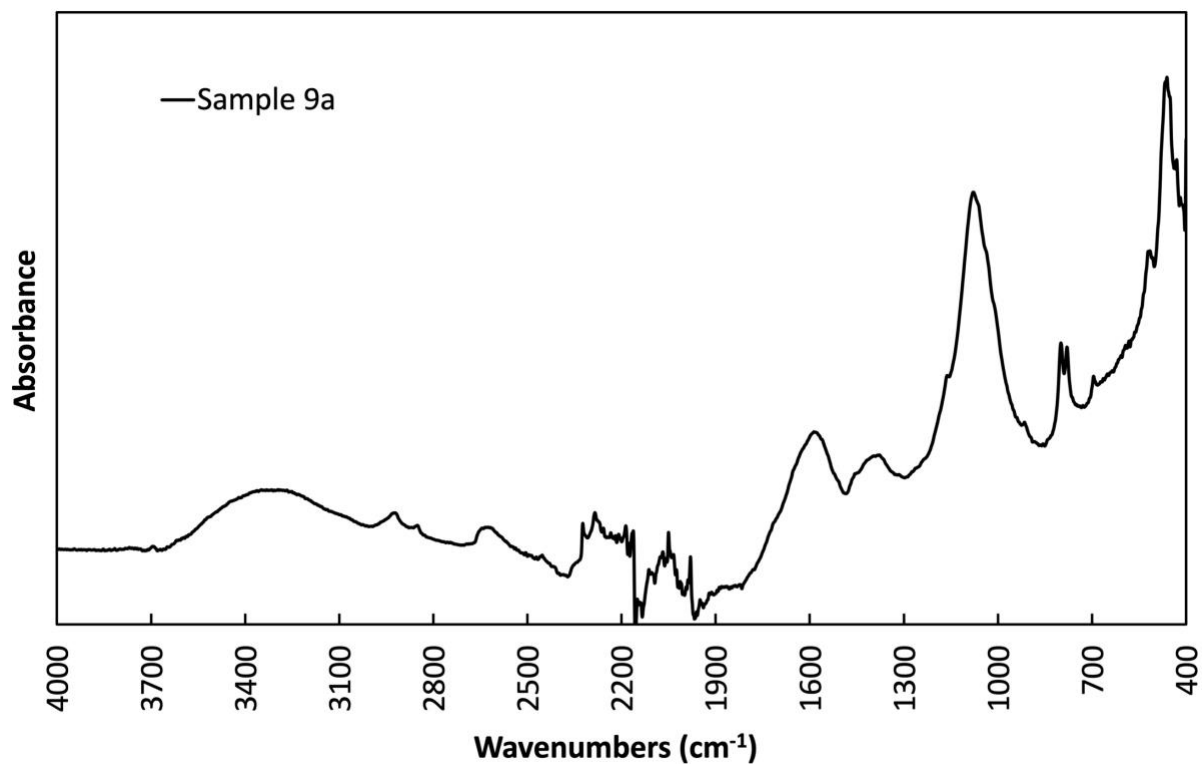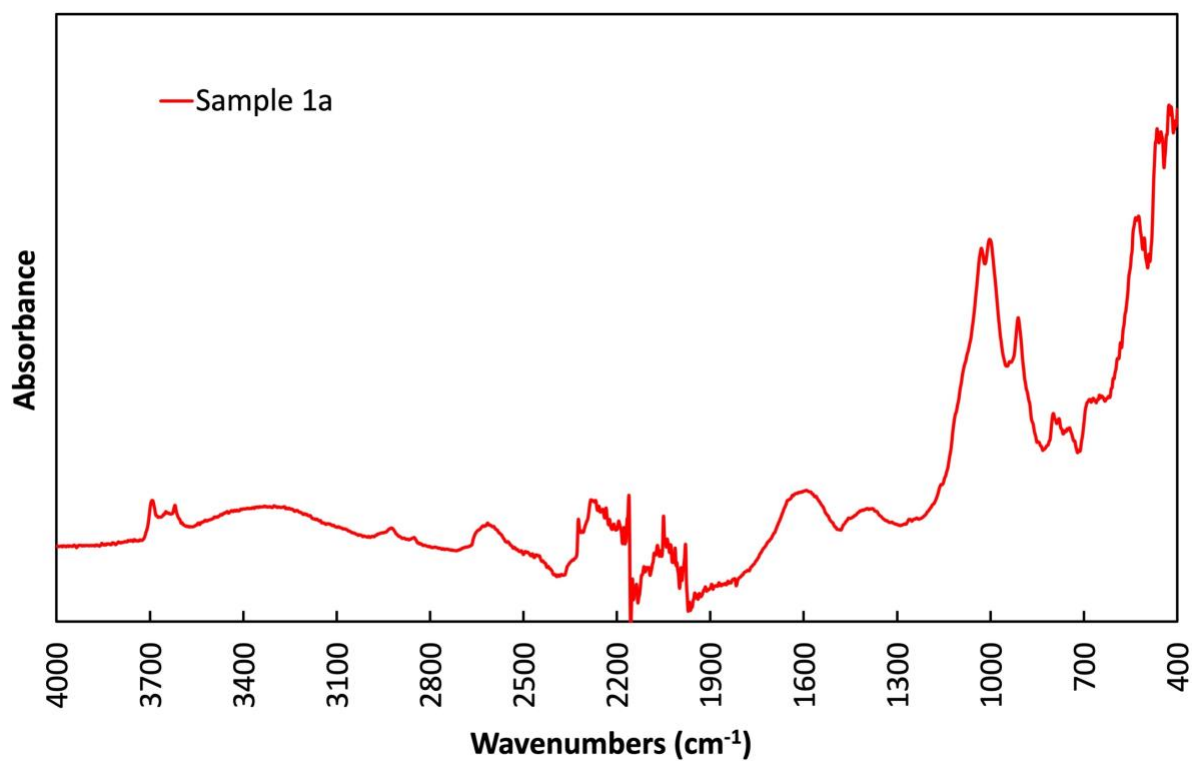
Figure A5. Representative raw ATR-FTIR spectra of sample 9a, which is situated at the most positive value of PC-1 and the most negative value of PC-2 within the ATR-FTIR PCA scores plots. The region associated with the ATR-IR diamond (2350 – 1950 cm^{-1}), though present in these spectra, was removed prior to chemometric analysis.

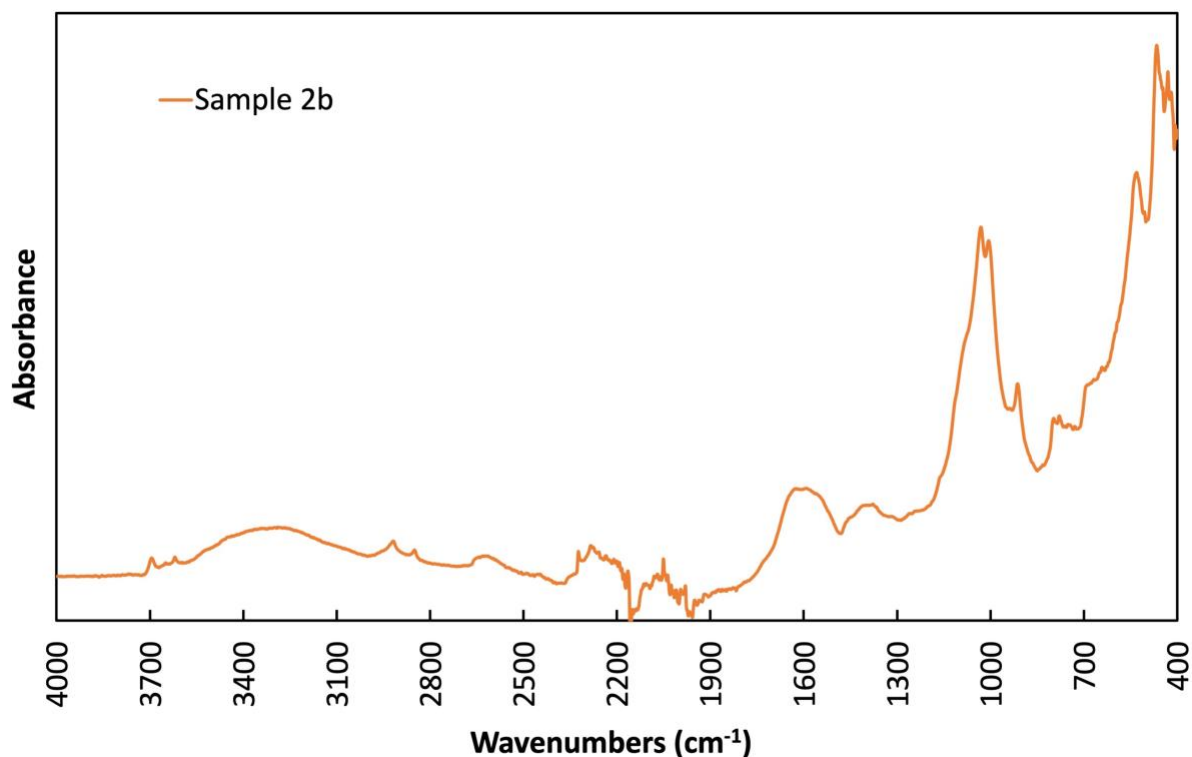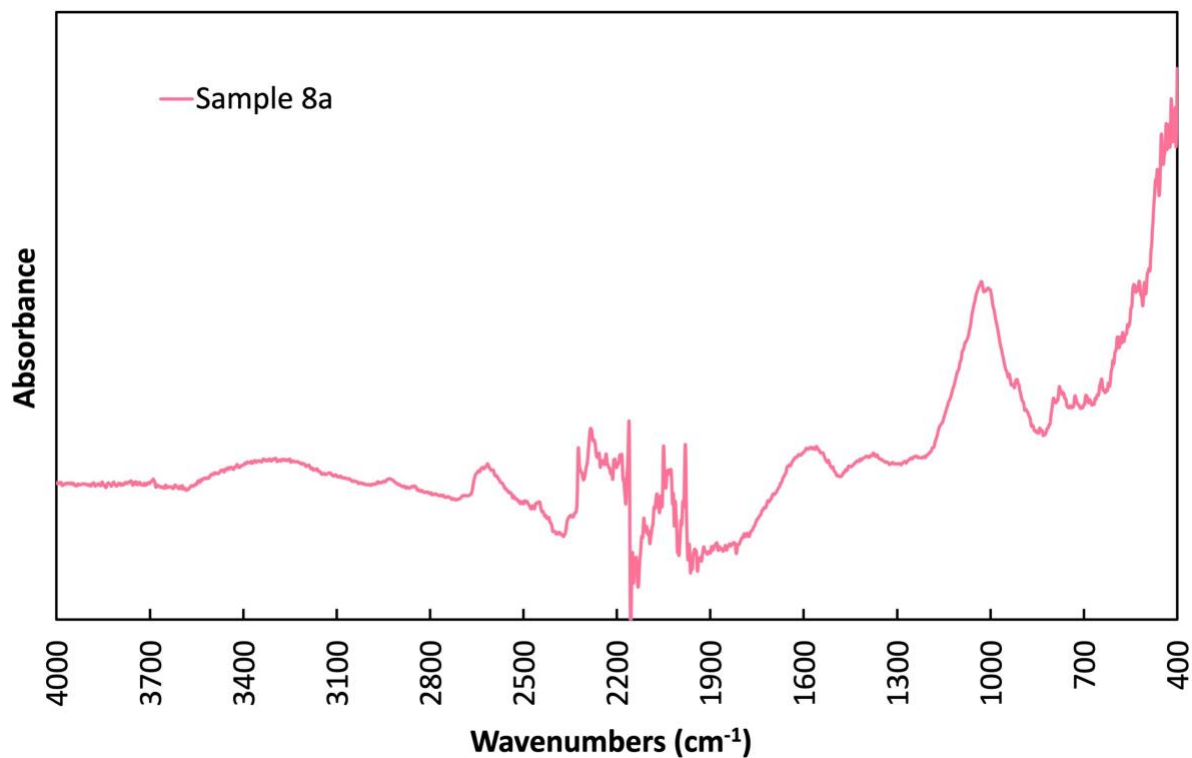**Figure A6.** Representative raw ATR-FTIR spectra of sample 1a, which is situated at the most positive value of PC-2 within the ATR-FTIR PCA scores plots. The region associated with the ATR-IR diamond (2350 – 1950 cm^{-1}), though present in these spectra, was removed prior to chemometric analysis.

Figure A7. Representative raw ATR-FTIR spectra of sample 2b, which is situated at the most negative value of PC-3 within the ATR-FTIR PCA scores plots. The region associated with the ATR-IR diamond (2350 – 1950 cm^{-1}), though present in these spectra, was removed prior to chemometric analysis.

Figure A8. Representative raw ATR-FTIR spectra of sample 8a, which is situated at the most positive value of PC-3 within the ATR-FTIR PCA scores plots. The region associated with the ATR-IR diamond (2350 – 1950 cm^{-1}), though present in these spectra, was removed prior to chemometric analysis.
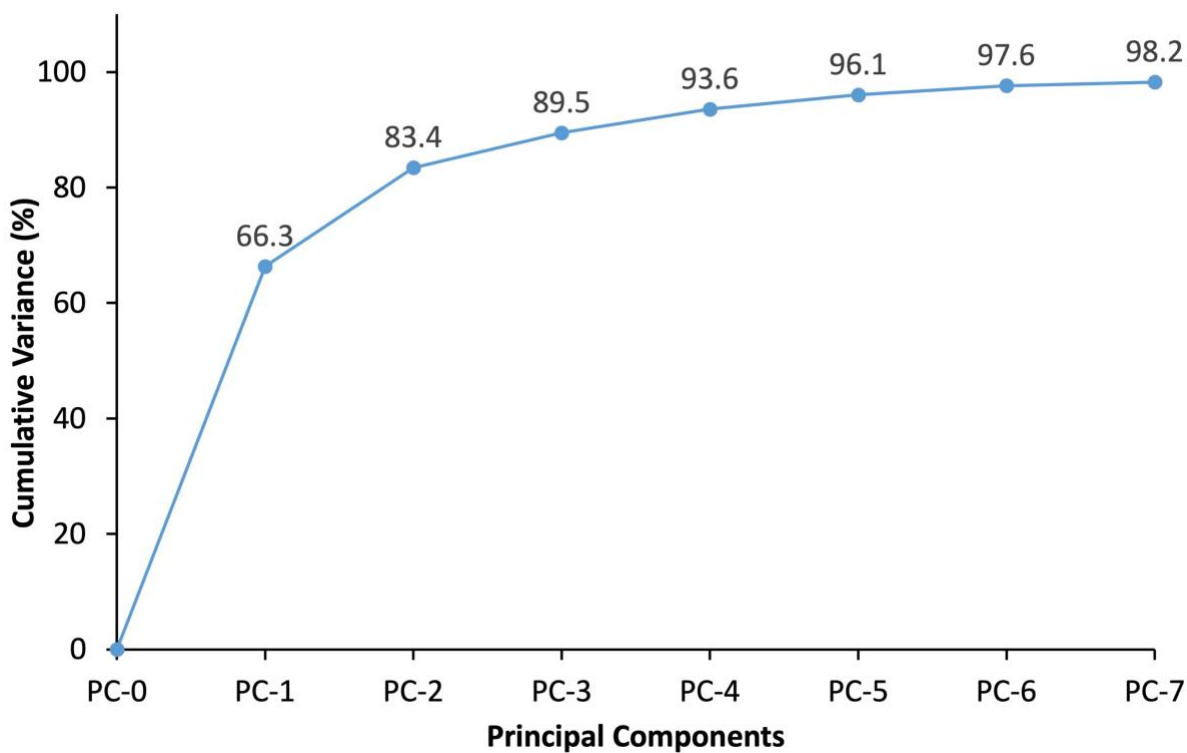

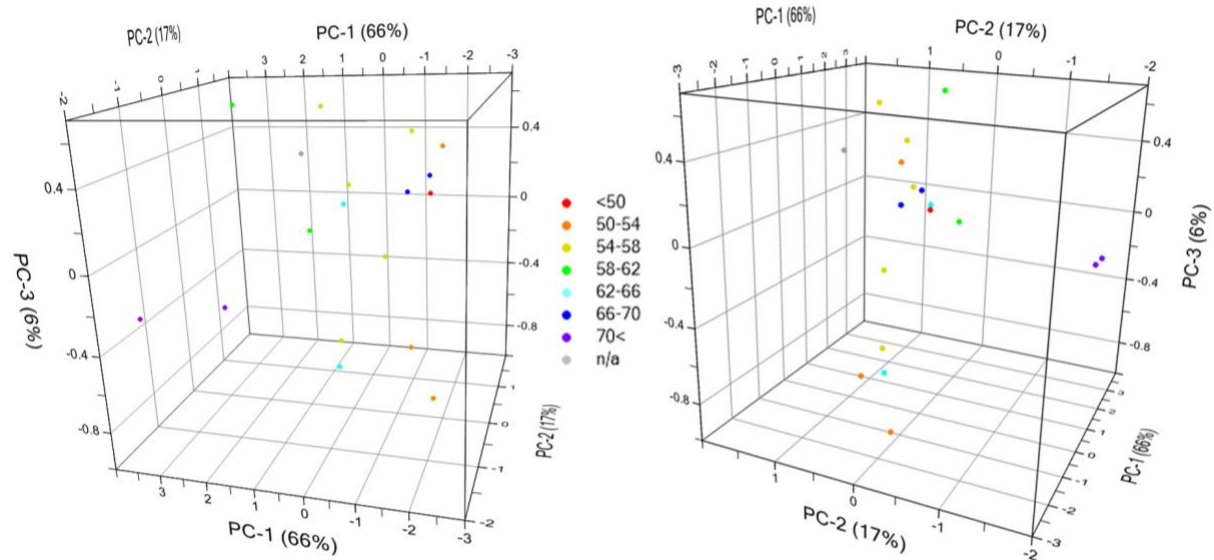**Figure A9.** Scree plot depicting the cumulative variance in the XRD dataset retained by each PC.

Figure A10. 3-dimensional PCA scores plot based on the XRD patterns of soil samples, showing the variability in the L* values obtained from corresponding MSP spectra from the same soils.