Development of a spectroscopic method to identify archaeological remains and soils using reflectance spectra in the visible to near infrared region

Dissertation
zur Erlangung des Grades
„Doktor der Naturwissenschaften“
im Promotionsfach Geographie

am Fachbereich Chemie, Pharmazie und Geowissenschaften
der Johannes Gutenberg-Universität in Mainz

Yoon Jung Choi

geb. in Seoul

Mainz, 2018

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 28.08.2018

Abstract

Colour of buried archaeological remains tends to be different from the adjacent soils due to anthropogenic activities influencing the chemical properties of the soil. Such distinctive colour difference indicates that buried remain might be easily identified by their colour, but the colour recognised by the human eye (characterised by the contributions of light in the red, green and blue spectral regions) is not always sufficient to clearly distinguish between natural soils and archaeological artefacts. This thesis attempts to use the full information content of reflectance spectra in the visible and near infrared spectral range to identify archaeological remains. Since such reflectance spectra are attenuated in a complex way by scattering and absorption processes, the spectral characteristics of archaeological remains are investigated using a modified principal component analysis (PCA) method. The PCA method is extended in a way which allows to quantify the differences between a spectrum of interest and a group of selected natural soils by a distance value 'D'. Large D values indicate that the spectrum represents a non-natural soil, e.g. an archaeological material. Archaeological sites investigated in this thesis have provided positive results (large D values) for pits, ditches and archaeological features influenced by fire activity. The developed method works best if reference spectra of local soils are used, but even with a global database of soil spectra, archaeological materials can still be identified. These results indicate that the method can be applied in a universal way to any archaeological sites. However, this hypothesis has to be further proven by additional investigations. It should also be noted that the developed method might not only be used for archaeological applications but also to distinguish in a general sense whether the soil colour difference is due to natural process or anthropogenic influence. Another important result of this study is that spectral features of archaeological remains can still be identified with much lower spectral resolution than provided by the spectrometer used (3 – 10 nm). These findings demonstrate a promising approach to use instruments with coarser spectral resolution, but largely improved temporal resolution, which might even allow continuous 2D imaging applications.

Abstract

Zusammenfassung

Durch lokale Variationen in der chemischen Zusammensetzung können archäologische Überreste die Farbe des Erdbodens verändern. Dies kann zur visuellen Identifikation (rote, blaue und grüne Farbkanäle) archäologischer Fundstellen genutzt werden. Diese Form der Auswertung ist jedoch auf die subjektive Wahrnehmung des menschlichen Auges beschränkt und daher weniger empfindlich für Farbunterschiede zwischen natürlichem Boden und archäologischen Fundstellen. In dieser Arbeit wird eine Methode vorgestellt, welche den vollen Informationsgehalt spektral aufgelöster Reflektanzen im sichtbaren und nah-infraroten Spektralbereich verwendet, um archäologischen Fundorte zu identifizieren. Aufgrund des Einflusses komplexer Streu- und Absorpttionsprozesse auf das gemessene Reflektanzspektrum wird für die Bestimmung der spektralen Eigenschaften eine modifizierte Form der Hauptkomponentenanalyse (PCA) verwendet: Die Methode wird um ein Abstandsmaß erweitert, welche es ermöglicht, die Unterschiede zwischen natürlichem Erdboden und archäologischen Fundstätten zu quantifizieren.

Hohe Werte für dieses Abstandsmaßes kennzeichnen Bodentypen, deren spektrale Charakteristika stark von natürlichem Boden abweichen und somit auf vergrabene Fundstellen hindeuten: In dieser Arbeit konnten auf diese Weise Gräben, andere Vertiefungen und verbrannte Fundstücke identifiziert werden. Die beschriebene Methode funktioniert am besten, wenn natürliche Erde aus der Umgebung des Messorts als Referenz verwendet wird. Allerdings kann man bereits mit Referenzspektren einer globalen Datenbank gute Ergebnisse erzielen. Ob dies eine universelle Anwendung der Methode ermöglicht, muss jedoch in weiteren Studien untersucht werden. Die hier vorgestellte Methode zeigt natürliche und anthropogen verursachte Unterschiede in der Beschaffenheit des Erdbodens auf und ist nicht nur auf archäologische Fundstätten beschränkt. Es konnte außerdem gezeigt werden, dass eine gröbere spektrale Auflösung als die vorhandenen 3 bis 10 nm ausreichend ist, um Unterschiede in der Bodenbeschaffenheit zu kategorisieren. Diese Tatsache ermöglicht die Verwendung anderer Instrumente mit geringerer spektraler, aber höherer zeitlicher Auflösung und 2D bildgebender Verfahren.

Contents

Chapter 1 Introduction

Buried archaeological remains are occasionally visible due to the differences in physical or chemical properties of the remains from surrounding natural soils (Oonk et al., 2009a; Rowlands and Sarris, 2007). Such differences can be identified by non-destructive archaeological prospection methods (Alexakis et al., 2009; Ceraudo, 2013; Cox, 1992; Doneus et al., 2014; Kvamme, 2005; Lasaponara and Masini, 2007; Themistocleous et al., 2015; Wilson, 1975), such as remote sensing techniques (Agapiou et al., 2010, 2013; Aqdus et al., 2008; Lasaponara and Masini, 2008, 2011; Themistocleous et al., 2015; Winterbottom and Dawson, 2005). However, the identification of buried remains utilising changes in soil properties is not common in remote sensing or spectroscopic application due to the complexity of soils and remains beyond the visible range (Cavalli et al., 2009; Themistocleous et al., 2015; Wetterlind et al., 2013).

The difference in soil colour have been traditionally discovered through a simple camera or bare eye approach depending on archaeologists' personal experience and knowledge based on the shape and location of the feature (Wilson, 1975). However, such bare eye prospection methods are subject to limitations since they are restricted to only three spectral bands (red, green and blue). In addition, these methods are also extremely subjective as well as time consuming (Trier et al., 2009). Therefore, a quantitative and reproducible method using visible to near infrared spectroscopic information is needed to overcome these restrictions. This thesis focuses on the development of a methodology to prospect buried archaeological remains using spectral information of archaeological soils in comparison to those of natural soils.

In order to achieve the aim of this thesis, soil spectra from known archaeological sites were gathered by a portable hand-held spectrometer (visible to near infrared). First, an optimal spectral range is determined to identify archaeological features efficiently. Using this spectral range, this study applies a modified principal component analysis (PCA) to statistically identify archaeological signatures. This method is developed to quantify the dissimilarity between archaeological and natural soil spectra from any unknown archaeological site.

The thesis describes the individual steps for the development of the new methodology and applies the method to several archaeological sites. The results indicate that the method can well be used to discriminate archaeological material from natural soils. But further investigations are needed to test the method at various archaeological sites and diverse environmental conditions.

1.1 Outline

General background information of archaeological prospection is introduced in Chapter 2 including discussions of problems and limitations of the individual methods. The chapter also includes an explanation of a specific archaeological terminology, soil mark features, and various methods developed to identify such features beyond the eye. The chapter describes soil chemical studies to understand the link between archaeological and natural soils and the results of spectroscopy.

Chapter 3 gives an overview of the instrument and measurement settings for data collection and outlines the methodologies used and developed within this thesis. Various pre-treatment methods of the spectra (to improve the quality of the measurements) before the main analysis are explained. The chapter also describes a specific spectral feature in the low wavelength range which may be an instrumental artefact. Different ways to numerically represent spectral features of archaeological remains are introduced together with the general concept of PCA. The main methodology (modified PCA) developed in this thesis is explained step by step.

The new modified PCA method is applied to two case study areas, Italy (Chapter 4) and Hungary (Chapter 5), which have the contrasting environment and are therefore expected to yield different results. All the methods introduced in the methodology chapter (Chapter 3) are applied to spectra recorded in the first case study region, Italy, and the effectiveness of each method is analysed. The methods yielding the best results are also applied to the Hungarian site.

The overall summary is presented in Chapter 6, and also the limitations and problems of the new method are discussed. Based on the results obtained in this thesis recommendations for additional applications further improvements are also provided.

Chapter 2 Background of the study

2.1 Archaeological prospection

Archaeological prospection is a field of archaeology which utilises various scientific instruments to identify buried remains (Batayneh, 2011). The traditional prospection methods, such as field walking surveys, also known as pedestrian survey (Kvamme, 2005; Schiffer et al., 1978), are still prevalent in archaeology (Bevan and Conolly, 2004; Foard, 1977), scientific prospection methods are highly needed not only to identify buried remains, but also to reduce the time of prospection for rescue archaeology.

The scientific prospecting methods include geochemical surveys (Gaffney and Gaffney, 2000; Gerlach et al., 2012; Holliday and Gartner, 2007; Jones and Sarris, 2000; Kattenberg, 2008; Kvamme, 2003, 2006; Persson, 2005; Schlezinger and Howes, 2000; Shackley, 2011; Simpson et al., 1998; Terry et al., 2004; Theocaris et al., 1996; Wilson et al., 2008; Wynn, 1990) and remote sensing methods (e.g. from airborne or satellite observations, Aqdus et al., 2008; Beck et al., 2007; Lasaponara and Masini, 2005, 2017; Rowlands and Sarris, 2007; Traviglia, 2006; Verhoeven and Sevara, 2016; Winterbottom and Dawson, 2005). Furthermore, scientific prospection methods are less prone to arbitrary and subjective interpretation.

Among many archaeological prospection methods, aerial archaeology is a powerful tool widely used to identify buried remains through aerial images (Alexakis et al., 2009; Ceraudo, 2013; Cox, 1992; Doneus et al., 2014; Kvamme, 2005; Lasaponara and Masini, 2007; Themistocleous et al., 2015; Wilson, 1975). From the 19th century, buried archaeological remains are recognised by trained archaeologists through photographs taken during flights (Batayneh, 2011; Ceraudo, 2013; Wilson, 1975). Figure 2.1.1 locates various archaeological remains identified by aerial archaeology around the Mainz region in Germany. Current studies extend the use of aerial images from multi- to hyperspectral images with a large coverage and higher spatial resolution (ranging from 4 m to 1 m, Doneus et al., 2014) obtained from the visible to near infrared spectral range and sometimes even in the thermal infrared region (Berlin et al., 1977; Rowlands and Sarris, 2007). Also with technological improvements, such

high spatial and spectral resolution are also becoming available on spaceborne platforms (Drzewiecki and Rączkowski, 2008; Kruse et al., 2003; Lasaponara and Masini, 2011, 2014; Lasaponara et al., 2016).

In aerial images, the buried remains are identified by 1) the different colour and/or brightness compared to the surrounding natural soil and 2) the characteristic shape which cannot be formed naturally. Such features are called 'soil marks' since they are visible on bare soil surfaces over buried remains (Bowen, 1975; Cantoro et al., 2017; Evans and Jones, 1977; Jones and Evans, 1975). Similarly, 'crop mark' features (or sometimes referred to as 'vegetation mark') can also be identified by a changed crop growth (height, colour, stress level and other factors) over buried remains (Agapiou and Hadjimitsis, 2011; Agapiou et al., 2011, 2012; Doneus et al., 2014; Evans and Jones, 1977; Lasaponara and Masini, 2007). Figure 2.1.2 shows various soil mark features identified from aerial photographs ranging from the early aerial photographs taken in black and white (visible due to the tonal difference between the soils) to recent images of buried remains identified from the plane and by Google Earth.

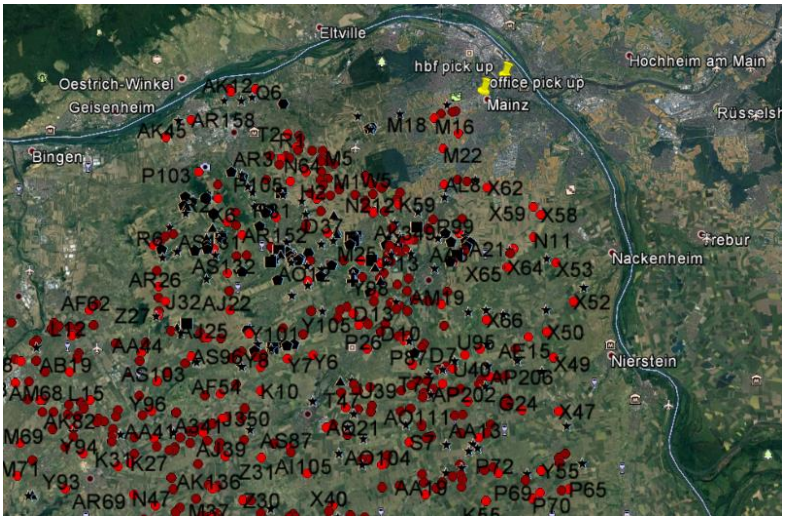


Figure 2.1.1 Buried archaeological remains identified through aerial photographs. Remains identified by Dr. P. H. Positioned in Google Earth by Dr. P. B. from Johannes Gutenberg University Mainz.

2.2 Soil marks

Soil mark features are important signatures of buried remains, mostly recognised from a distance as long, thin, circular and dark soil features (Bowen, 1975; Gerlach et al., 2012; Wilson, 1975). These features are mainly observed on arable land without vegetation cover by the colour, tonal and texture differences compared to the surrounding natural soil (Ceraudo, 2013). Such differences in soils are due to the ancient materials spread on the topsoil which originated from buried remains. In archaeological surveys, soil marks proved to be important indicators of archaeological sites (Taylor, 1979; Verhoeven and Sevara, 2016).

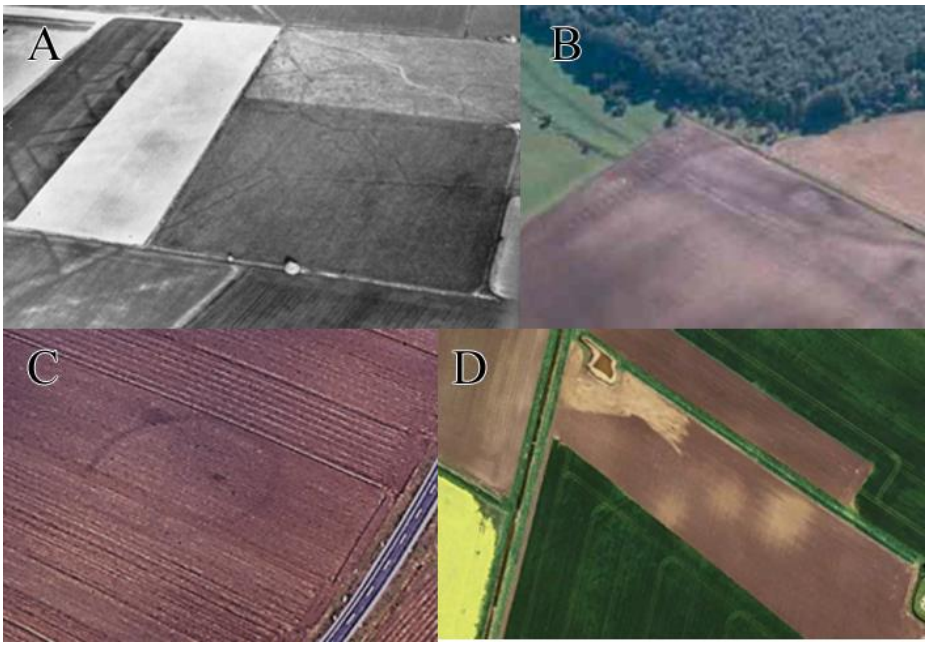


Figure 2.1.2 Images of soil mark features.
A: Romano-British agriculture in the Fens (UK). Crop and soil marks visible from east-south-east view (Wilson, 1975). Photo: University of Cambridge, Crown Copyright reserved.
B: Medieval road near Coughton (UK). Photo: The Oxford Archaeological Unit.
C: Prehistoric ring ditch near Stadecken Hayer (Germany). Photo taken by Dr. P. H.
D: Ring barrows in Fenlands, England. Image taken from Google Earth.

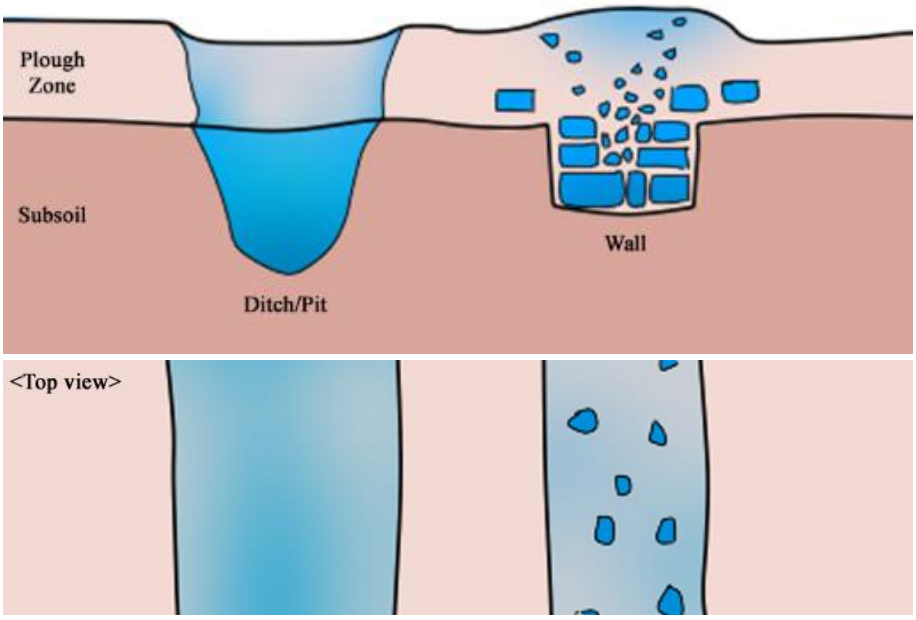


Figure 2.2.1 Schematic description of the soil mark formation. After ploughing, the top part of the buried remains is transported to the surface. This will result in colour and texture difference compared to the surrounding natural soil. Modified from Beck, 2010.

Such soil marks are created when the buried remains are transported to the surface of the ground by human (ploughing etc.) or biochemical (earthworms etc.) activities (Clark, 2003; Taylorab, 1979). Among these activities, ploughing has revealed numerous buried remains on the soil surface in Europe. One interesting feature is that soil marks are mainly concentrated in Europe and not much discovered in other regions. These regional differences are not yet well understood but are assumed to be the consequence of different farming traditions (Lee, 2011).

Figure 2.2.1 shows how the soil marks are formed in arable lands after ploughing. The mechanism is simple. When the terrain with archaeological remains buried inside is ploughed, the plough scraps off the top part of the buried remain and transfers this to the surface of the soil. As a result, buried remains such as ceramic pieces, building rock fractures, archaeological pits are scattered on to the ground surface. Such transported remains are sometimes different in their texture, object fragments and colour to the surrounding natural soil (Jones and Evans, 1975). For examples, pit features tend to be highly organic and therefore generate fairly darker patches (Lauer et al., 2014) whereas ditch features tend to hold more moisture than the surrounding natural soils and thus appear as darker lines Taylor (1979). Also, if such transported buried materials are observed from a distance, they might show characteristic shapes (as shown in Figure 2.1.2), that help identifying them as archaeology. In summary, buried remains are observed by the contrast between archaeological features and the natural background soils based on the assumption that the buried structures alter the natural properties (Gallo et al., 2009; Kruse et al., 1993).

Studies have attempted to identify soil mark features in a systematic and reproducible way to reduce time-consuming manual inspections. Such methods are also applied in particular to satellite images to identify archaeological remains on a global scale (Ceraudo, 2013). However, most of these studies using on soil marks are still in an early explorative state (e.g. Crawford, 1939; Donoghue and Shennan, 1988; Evans and Jones, 1977; Taylorab, 1979). Recent studies focus on analysing not soil mark features but crop mark features (Agapiou et al., 2013; Bassani et al., 2009; Doneus et al., 2014; Jones et al., 1968) since soil marks are more difficult to detect (Themistocleous et al., 2015).

2.3 Limitations in soil mark prospection

2.3.1 Colour (limited spectral range)

Soil colour is an important indicator not only for soil characteristics (Webster and Butler, 1976; Viscarra Rossel et al., 2006c) but also for ancient anthropogenic activities (Aston et al., 1998; Canti and Linford, 2000; James, 1999; Jones and MacGregor, 2002). For example, Eckmeier and Gerlach (2012) used differences in soil colour to estimate the age of the archaeological context. However, identification of soil mark features to prospect buried remains is extremely subjective to the trained archaeologists since the interpretation of the images is mainly based on personal experiences and limited to the visible electromagnetic spectrum between 400 and 700 nm (Doneus et al., 2014). Therefore, soil marks are sometimes difficult

to detect since these marks may not be visible in the images if observed over different periods or at different spatial or spectral resolution (Themistocleous et al., 2015).

Studies have used hyperspectral remote sensing methods (vis-NIR region) to identify soil marks. Such data provide information on moisture content (Liu et al., 2003), texture and chemical minerals of soils (Ben-Dor et al., 2002; Chabrillat et al., 2002; Leone and Escadafal, 2001; Metternicht and Zinck, 2003). These studies attempted to observe soil mark features not only in the visible spectral region but also in the near infrared and short wave infrared spectral regions since identification of soil, vegetation, rock and mineral materials is promising in these spectral ranges (Clark et al., 1990; Gao and Goetz, 1990; Stoner and Baumgardner, 1981). For example, Donoghue and Shennan (1988) used edge enhancement to show highly enhanced soil marks in red wavebands (605 – 690 nm) since soils reflect red light strongly. Infrared images can be effectively used for identifying crop mark features since the near infrared region (760 – 900 nm) is sensitive to vegetation (Fowler, 2002). However, depending on the time of the survey, the presence of vegetation obscures the pure soil signature (Miller, 1977). Even in semi-arid regions, the ability to distinguish soil reflectance features from vegetation features becomes difficult for vegetation cover above 50 to 60 % (Murphy and Wadge, 1994). Cavalli et al. (2009) used hyperspectral images (400 to 1280 nm) to identify the optimal spectral range for archaeological prospection using various analysis methods such as spectral unmixing (Small, 2001). The results showed that hyperspectral sensors covering the 400 – 800 nm spectral range with a spectral resolution of 20 – 50 nm is most effective for identifing archaeological features when the land cover abundances are known.

Different soil colours of soil mark features are often observed by airborne and satellite images with sufficient spatial ranges (Fowler, 1996; Fowler, 2002; Menze and Ur, 2007), but not many studies have been published identifying soil mark features beyond the visible spectral range since archaeological remains do not have unique spectral characteristics (Bassani et al., 2009).

2.3.2 Shape

Archaeological remains typically have unique shapes. Circles, squares, rectangles and straight lines are generally of human origin and occur less frequently as products of nature (Kvamme, 2005). However, such shapes can be similar to traces of geological features or recent human activities and, therefore, may confuse archaeologists. Figure 2.3.1 shows images of features resembling soil marks but are not traces of ancient settlements. When scientists attempted to develop methods to identify soil marks by their characteristic shapes, geological features were often problematic.

Hu and Li (2017) revealed 70 new archaeological ruins in north western China which have geometric features (square or polygon) using GeoEye satellite data. Similarly, Trier et al. (2009) introduced an automatic detection method of soil and crop marks based on pattern recognition of panchromatic bands and high resolution satellite images. These soil marks were determined by their circular shapes through local contrast enhancement and template matching (ring features identified by archaeologists in prior) and were successfully applied to

ring-shaped remains in southeast Norway. However, some false rings (non-archaeological features) were also detected.

Classifying archaeological features among various geologies or landscapes based on their shapes can reveal unidentified remains, but difficulties also occur since it can also extract geological features since the origin of the shapes are not clearly understood.

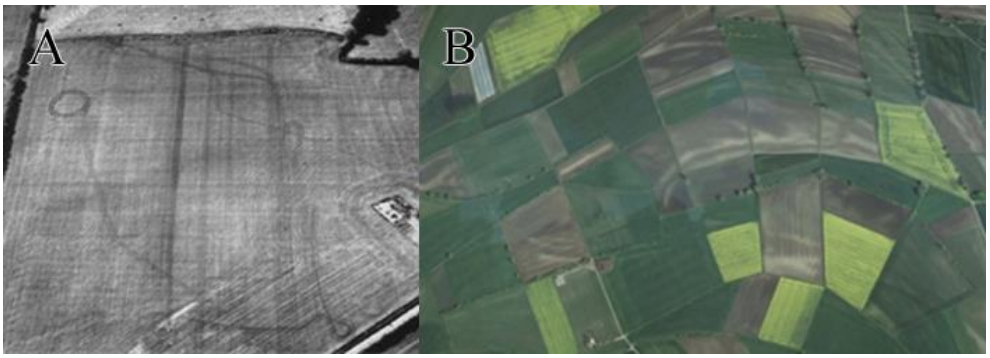


Figure 2.3.1 Arial photos showing examples of geological and recent human activities which look similar to soil or crop marks.
A: Aerial photo taken in Southwest of Dagnall (Photo: University of Cambridge, Crown Copyright reserved). The ring-ditch like feature on the top left is actually created where a tractor had turned in a wide circle.
B: Image of traces of meanders and other geological features which look similar to soil mark patterns in Hessen, Germany (Google Earth, 2018).

2.4 Archaeological analysis methods of the soil chemical composition

The above section (Chapter 2.3) showed that the detection of soil mark features is restricted to their colour and shape due to the complexity of soils (Cavalli et al., 2009). Soil marks themselves already consist of complex soil structures as buried remains are transported up to the surface soil. Therefore a more fundamental approach is needed. In order to achieve this, studies on how scientists have analysed buried remain (not soil marks but remains and archaeological soils themselves) are investigated in this chapter.
Anthropogenic activities may affect the chemical composition (Gianfreda et al., 2005; Ulrich, 1986) of the archaeological soils by either enrichment or depletion of specific chemical elements and thus the formation of archaeological soils (Oonk et al., 2009a). Also, buried remains themselves are likely to alter the physical and chemical characteristics of the soil compared with those of the surroundings due to variations in soil depth and drainage (Rowlands and Sarris, 2007). Therefore soil analysis is a useful tool to understand human activity reflected in the development of soils (Pastor et al., 2016). Walkington (2010) described how soil studies can reveal human influence. However, geochemistry as an aid to archaeological prospection is not as popular as the geophysical methods since the understanding of the links between soil chemistry and archaeology is limited (Oonk et al., 2009a).

Element analyses of soils are widely used to characterise archaeological soils (Pastor et al., 2016; Wilson et al., 2009) and, thus, Table 2.1 illustrates the chemical elements observed in

specific archaeological sites. Among various chemical analysis methods, X-ray fluorescence is a powerful tool to determine the elemental compositions of soil and thus have frequently been used for archaeological applications (Shackley, 2011). For details see section 3.1.4. XRF is a moderately specific, but fast and relatively cheap device to detect elements for geochemical soil analysis (Bernick et al., 1995; Bousquet et al., 2007; Kalnicky and Singhvi, 2001; Martin et al., 2007).

As shown in Table 2.1, during the past decade, archaeological soils have been characterised by anomalous element levels (P and heavy metals such as Ca, K and Mg). Ca, Cu, Mg, K, Na, P and Zn are commonly found in archaeological soils (Cook and Heizer, 1965; Eidt, 1984; Haslam and Tibbett, 2004; Middleton and Price, 1996; Ottaway and Matthews, 1988) since these elements are often present in occupation waste (Greweling, 1962; Hao and Chang, 2003; Maly et al., 1999). Ca, P and K can be used to understand specific ancient anthropogenic activities, such as metal production and craft work (Oonk et al., 2009a). Copper is one of the most promising anthropogenic indicators because it is relatively stable in soils (Fontes and Gomes, 2003). Among these chemical elements, phosphorous (P) is a powerful element to trace anthropogenic activities (Fernández et al., 2002; Holliday and Gartner, 2007; McCawley and Mckenell, 1971; Sánchez et al., 1996; Terry et al., 2000) due to its abundance in plant/animal tissue, bone and ashes (Bethell and Máté,1989; Proudfoot, 1976). But it is also a rather unreliable indicator of human occupation in some instances (Entwistle et al., 1998, 2000a,b).

However, the sources of these elements sometimes may be related not only to ancient human activities but also to modern anthropogenic inputs which can have profound chemical effects on soils (Bølviken et al., 2004; Golia et al., 2008; Marcos et al., 1998; Qafoku et al., 1999; Sánchez-Martín et al., 2007). In addition, many of these elements may not be found uniformly in different archaeological soils (Oonk et al., 2009a). Also, the determination of a geochemical baseline comprising the natural soil composition in order to recognise the anthropogenic impact may often be problematic (Matschullat et al., 2000).

Geochemical analyses of the elemental composition of archaeological soils can provide useful information, but often a definitive proof remains unconfirmed (Oonk et al., 2009a).

Table 2.1 Examples of archaeological sites and features and associated element enrichments in their soils (modified from Oonk et al., 2009a).

Archaeological site/feature	Elements	References	Methods
Burials/graves	P, Cu, Mn, Ca	Bethell and Smith,1989; Cook and Heizer, 1965; Keeley, 1981; Parson, 1962.	Disodium, dihydrogen ethylenediamenetetraacetate (EDTA), P analysis.
Hearths	P, K, Mg	Barba et al., 1996; Knudson et al., 2004.	Ring chromatography, phosphor-vanado-molybdate, plasma-atomic emission spectrometer (ICP-AES).
Middens	P, K	Chaya, 1996; Fernández et al., 2002; Parnell et al., 2001; Wells et al., 2000.	Molybdovanado-phosphate colorimetric method, Mehlich II solution (Terry et al., 2000), ammonium acetate (Normandin et al., 1998), Mehlich II extraction solution (Mehlich, 1978), Hach reagents (Hach, Co., Loveland, CO).
(Farm)houses	P, Ca, Mg, Fe, K, Th, Rb, Cs, Pb, Zn, Sr, Ba	Chaya, 1996; Entwistle et al., 2000a,b; Fernández et al., 2000; Manzanilla, 1996; Parnell et al., 2001; Wells et al., 2000; Wilson et al., 2005,2006.	Molybdovanado-phosphate colorimetric method, hydrochloric acid, ammonium acetate (Normandin et al., 1998), Mehlich II solution (Terry et al., 2000), Mehlich II extraction solution (Mehlich, 1978), inductively coupled plasma optical emission spectrometry (ICP-OES).
Painted buildings	Heavy metals	Wells et al., 2000.	Mehlich II solution (Terry et al., 2000), DTPA (diethylenetriaminepentaacetic acid), Thermo Jarrell Ash ICP spectrometer (Linderholm & Lundberg, 1994).
Mining, metals melting and production sites	Cu, Pb, Mn	Hong et al., 1994; Jenkins, 1989; Maskall and Thornton, 1998; Monna et al., 2004; Pyatt et al., 2002.	X-ray diffraction analysis, inductively coupled plasma optical emission spectrometry (ICP-OES), graphite furnace atomic absorption spectrometry (GFRAAS), lost on ignition (LOI), inductively coupled plasma–mass spectrometer (ICP-MS).
General archaeological sites	B, Cu, Mg, Mn, Ni, P, Se, Zn, K, Ba, Ca, Na	Bethell and Smith, 1989; Cook and Heizer, 1965; Ottaway and Matthews, 1988.	Atomic absorption spectrophotometry, Perkin Elmer 1100 AA spectrophotometer, P analysis.

2.5 Soil spectroscopy analysis methods

Soil minerals absorb light from the ultraviolet to the mid infrared spectral region and, therefore, allow to obtain information about the chemical composition of a mineral (Viscarra Rossel et al., 2006a,c). Thus, the spectral reflectance of soils in the visible to near, infrared (vis–NIR) region can be used to characterise the soil constituents (Ben-Dor and Banin, 1995a,b; Ben-Dor et al., 1997; Brown et al., 2006; Escadafal, 1993; Stenberg, 2010; Stoner et al., 1979; Torrent and Barron, 1993; Viscarra Rossel et al., 2006b, 2008b). Soil minerals are determined from their absorption bands (at specific wavelengths range) found in the soil reflectance spectra (Bishop et al., 1994; Clark, 1999; Clark et al., 1990; Hunt, 1977; Hunt and Salisbury, 1970; Morris et al., 1985; Sherman and Waite, 1985; Viscarra Rossel et al., 2006a,b). Figure 2.5.1 illustrates absorption features of different soil minerals.

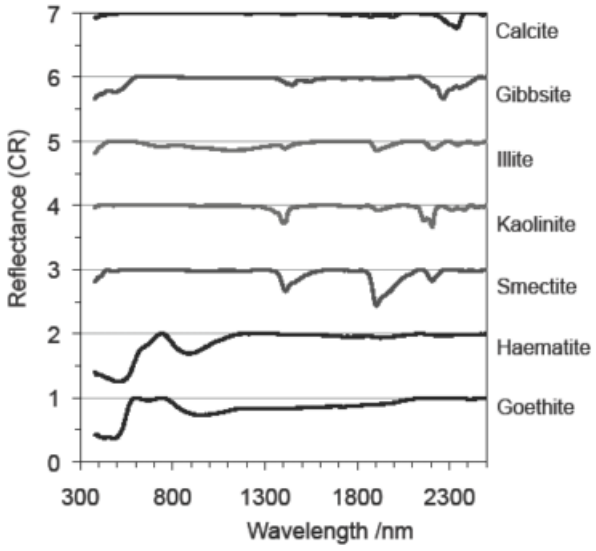


Figure 2.5.1 Continuum removed reflectance spectra of common soil minerals (modified from Stenberg et al., 2010).

Ben-Dor and Banin (1990) used the NIR region (1750 – 2550 nm) to estimate the carbonate concentration in soils and Ben-Dor and Banin (1994) used the 400 – 1100 nm spectral range to estimate the contents of $CaCO_3$, Fe_2O_3, Al_2O_3, SiO_2, free Fe oxides, and K_2O. Further studies have also used the vis-NIR spectral range to estimate the abundance of chemical constituents such as C, N, $CaCO_3$, Ca, K, Fe and etc. (Ben-Dor et al., 2008; Chang et al., 2001; Terhoeven-Urselmans et al., 2006; Viscarra Rossel et al., 2011), carbonate content (Summers et al., 2011), moisture content (Ben-Dor et al., 2008; Liu et al., 2003), and soil texture (Stenberg et al., 2010; Summers et al., 2011). Wetterlind et al. (2013) also identified clay minerals, organic matter and soil moisture content using continuum removal techniques.

Viscarra Rossel et al. (2006a) investigated the first ten principal components (PCs) of soil spectra. Here, the most characteristic absorption features of the mineral and organic spectra are highlighted by the first derivative spectra, and the first three PCs account for approximately 75% of the variation in the data. Viscarra Rossel et al. (2009) used soil spectroscopy to estimate the soil colour and mineral composition from the spectra and make predictions of the

clay content using PCA. PCA detects common patterns within soil samples as well as putative outliers (Lopo et al., 2017).

Using such soil spectroscopy methods also studies on identifying the mineral content of archaeological soils were made. Matneya et al. (2014) showed *in situ* subsurface reflectance spectroscopy of archaeological sites as tomographic maps. Here, they plotted tomographic maps of absorbance in the blue spectral range and visually correlated the spectroscopy results to tomographic maps of several chemical constituents (P, organic matter, Ca and Mg).

However, these spectroscopy analysis methods focused on the mineral concentration of the soil spectra. As already discussed in Chapter 2.4, the differences between the mineral composition of archaeological soils and natural soils are not yet well understood (Oonk et al., 2009a). Also, so far not many archaeological soil spectroscopy studies were carried out. In archaeological prospection, field soil spectroscopy is mainly used as ground truth data for satellite and airborne images (Agapiou et al., 2010; Milton et al., 2009). Thus more studies on spectroscopic approaches of archaeological remains are needed. Here it is important to point out that vis-NIR spectroscopy is a fast, non-destructive technique which has the potential to analyse some of the essential constituents of soils (Wetterlind et al., 2013). Therefore, this thesis applies and extends the PCA to archaeological soils since PCA is already a powerful tool in soil spectroscopy for revealing the information in the NIR (Martens and Naes, 1989; Stenberg et al., 1995; Wold et al., 1987).

2.6 Summary

The main aim of this thesis is to find a spectroscopic method which can differentiate ‘archaeological features’ from natural background soils. In the field of archaeological prospection, much research has been performed using airborne remote sensing images to detect buried structures (mainly in the visible range) by their shape, but not many studies have been done to understand the spectral features of the soils which are influenced by buried remains. A better understanding of the spectral features of archaeological remains is needed in order to establish a universal spectral procedure which can be used to identify archaeological features at any site.

Normally, archaeological features on the surface or within a soil profile are identified by archaeologists or soil scientists by bare eye. This is possible because most of the archaeological materials have different soil compositions than the surrounding natural soils and, therefore, show different soil colours. However, this is not always the case, and sometimes it is difficult for geoarchaeologists to distinguish archaeological strata within the soil profile, especially when there is not enough historical information about the site (i.e. when geoarchaeologists do not know what type of archaeological material they are expecting. For example, building materials are brighter in colour, burned materials are reddish in colour and pits are darker in colour, etc.). Therefore, although bare eye observation is useful and can already identify the distinctive features of archaeological material, this observation method can miss some useful archaeological information since the human eye only has receptors for three colours (red,

green and blue), while the measured spectra could much better resolve the spectral information, even outside the visible range. Also, bare eye observation is a very person-dependent analysis method and may differ from one to another. Due to different personal assessments, uncertainties and confusions can arise during soil survey (Penleton and Nickerson, 1951). Therefore, in this thesis, a more quantitative approach in detecting archaeological features among soils is developed.

The method developed in this thesis analyses soil spectroscopy is applied to reflectance spectra from the walls of excavated soil pits, which can also be used for archaeological soil prospection in aerial archaeology. It is known that archaeological remains do not always have unique shapes or spectral characteristics (Cavalli et al., 2009), but some archaeological artefacts may have spectral signatures which are distinguishable from the normal background materials (Buck et al., 2003). So far the use of field spectroscopy for archaeological applications is an open research area (Agapiou et al., 2010). Therefore, the PCA based method developed in this thesis can probably also be applied to the aerial prospection of archaeological materials.

Chapter 3 Methods

This chapter explains how the spectral measurements were taken and how they were treated before being used for spectral analysis. Then it is followed by the application and further development of a modified PCA method to the measured spectra. Also, colour index methods and the x-ray fluorescence (XRF) method are described.

3.1 Instrument and measurement procedure

3.1.1 Soil profile

This section illustrates how the soil profiles were chosen and obtained from each site. In addition, a detailed explanation of soil profiles and their horizons are given.

For comparison of the dataset, sites with extreme contrasting environments and soil compositions were chosen from Italy and Hungary. To avoid large variability within the different sites from each country, the study sites were chosen within the same region (two archaeoloiogcal sites in Calabria for Itlay and three sites in Sárvíz Valley for Hungary). Various numbers of pits were excavated for each site, where at least one pit contains buried archaeological remains and several contain natural soils. For the sites in Calabria, Italy, seven natrual soil pits were excavated for site 1 and five natrual soil pits were excavated for site 2 to cover the large variation in soil properties within the same site (a detailed description of the sites can be found in Chapter 4). For the Hungarian site, the number of pits was limited (refer to the reasoning explained in Chapter 5.1). Therefore, in these sites, pits were made at the boundaries of archaeological features where the profiles of the pits contain a natural soil profile as well as an archaeological profile (with an archaeological stratum). An overview of these sites and pits given in Figure 3.1.1.

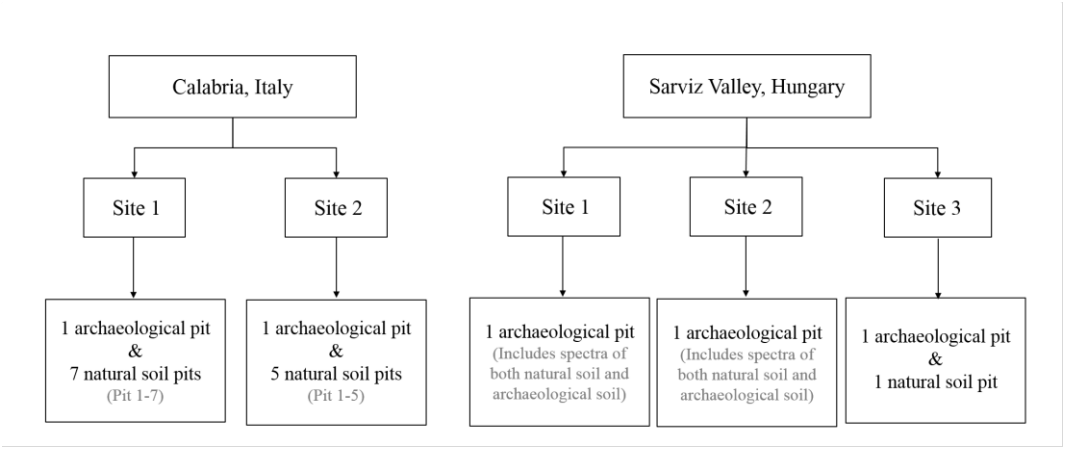


Figure 3.1.1 Chart of the different sites and pits, from which spectra are analysed in this thesis. Each site contains one pit with buried archaeological remains.

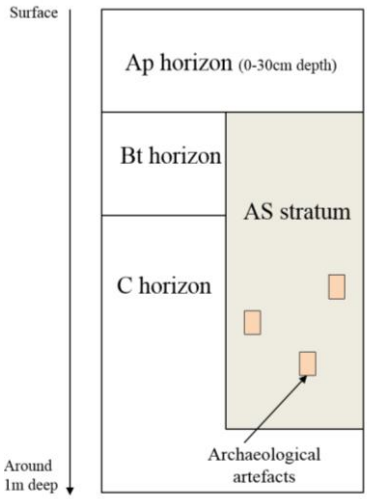


Figure 3.1.2 A Schematic image of a soil profile. Naming and division of each horizon are explained in Chapter 3.1.1.
Horizon depths vary for different pits, but the Ap horizon is fairly homogeneous for pits from the same sites since the depth of ploughing is similar. The depth of the pits is approximately around 1 m, which is the depth where the parent soil is observed. AS stratum is a layer of archaeological soils (AS) interrupting the soil profile. This stratum also sometimes contains archaeological artefacts such as ceramics and burned materials.

Soil spectra were measured on the side of the pit extending from the surface to the bottom of the pit. Such measurements were made to collect fresh *in situ* soil and archaeological spectra (Viscarra Rossel et al., 2009). Figure 3.1.2 shows a schematic image of a soil profile with different soil horizons annotated. A soil horizon is a layer of soil which has a distinct characteristic related to the soil forming processes. According to the World Reference Base for Soil Resources (WRB), these soil horizons can be divided into O, A, B and C horizons.

The sites chosen for this thesis are located on arable lands and, therefore, do not contain the O horizon, which is the horizon containing organic matter (plants etc.). As shown in Figure 3.1.2, the soil profiles in this thesis typically consist of only the A (Ap), B (Bt) and C horizons.

The A horizon is the surface soil (topsoil) which represents the Ap horizon in this thesis. Ap stands for the topsoil (A) where ploughing is constantly repeated and therefore, the thickness of the Ap horizon depends on the size of the plough (normally it is around 10 to 30 cm thick). This is the soil horizon where visible ‘soil marks’ (refer to Chapter 2.2) are observed since ploughing transports the buried archaeological remains to the surface.

Often a B horizon is formed below the A horizon. This is the subsoil horizon which has undergone a chemical or physical alteration of the parent material. In this thesis, B horizons will be referred to as Bt horizons that only differ from the former in their high content of clay. In the Hungarian site, no Bt horizon is visible.

The C horizon is formed between the B horizon (or the A horizon if the B horizon does not exist) and the bedrock. This horizon contains parent material and is little affected by soil forming processes. Therefore, it is the unconsolidated material which may or may not be the same as the parent material which formed the A or B horizons.

The three horizons explained above (Ap, Bt and C) are the typical soil horizons found in natural soil profiles in this thesis. However, for archaeological profiles, also an archaeological stratum might be present within the soil horizons as shown in Figure 3.1.2 (AS stratum). The abbreviation 'AS' is used to represent 'archaeological soil' in this thesis. This AS stratum consists of archaeological artefacts or just soils influenced by ancient anthropogenic processes. Buried archaeological artefacts such as pottery (pieces of ceramics) or burned materials will be referred to as ARCH in this thesis. Depending on the buried depth of the archaeological remains, the location of the AS stratum can vary between the Bt and C horizon.

Within these soil horizons, around three to four spectral measurement points were made per horizon. For every measurement point, the spectral measurement was repeated 3 to 4 times.

3.1.2 Instrumental setting

For spectral measurement of soils, the ASD (Analytical Spectral Devices, Inc.) FieldSpec Pro FR spectrometer was used. This spectrometer covers a wavelength range 350 – 2500 nm (2201 spectral bands), and it is separated into three spectrometers. The typical scan time is less than three seconds where the integration time of individual scans is 34 ms. The spectral sampling interval (or spectral bandwidth), which is obtained by dividing the spectral range by the number of elements in the detector array, is 1nm. Each spectrum was acquired as an average of 35 measurements. Figure 3.1.3 shows the schematic image of the ASD spectrometer and Table 3.1 provides the technical details and spectral resolution of each individual spectrometer within the ASD. For a more detailed description of the spectrometer, refer to FieldSpec Pro User's Guide (Malvern Panalytical, 2002).

The ASD spectrometer can either use sunlight or artificial light as its light source. First, the spectra were planned to be measured with sunlight, but during the field measurements, this became problematic because many measurements were affected by shadowing effect (depending on the sun direction and the depth in the pits). To avoid such problems, in this thesis all spectral measurements were made with a halogen light source. The so-called ASD Contact Probe (designed for contact measurements of solid raw materials), which includes a calibrated halogen light source and a fibre connected to the spectrometer, was used. Figure 3.1.4 shows a photo of the Contact Probe and how measurements were taken on the soil profile.

When sunlight is used as a light source, there are specific wave bands which have to be

removed in order to get meaningful reflectance spectra. These wave bands are called 'atmospheric windows', where the atmosphere essentially absorbs all solar photons. Figure 3.1.5 shows an example of a soil reflectance spectrum obtained with sunlight where the so-called 'atmospheric window correction' is made.

Table 3.1 Instrumental settings of the three spectrometers within the ASD FieldSpec Pro FR spectrometer. More detailed information about the spectrometer can be found at Malvern Panalytical (2002).

	Spectral range	Sampling interval	Spectral resolution (FWHM)	Detector
VNIR spectrometer	350 – 1050 nm	1.4 nm	3 nm@700 nm	512 element VNIR silicon photodiode array. Fixed grating, array detector, based spectrometer.
SWIR spectrometer1	900 – 1850 nm	2 nm	10 nm@1400 nm	TE cooled, graded index SWIR InGaAs (Indium gallium arsenide) photodiodes array.
SWIR spectrometer2	1700 – 2500 nm	2 nm	12 nm@2100 nm	

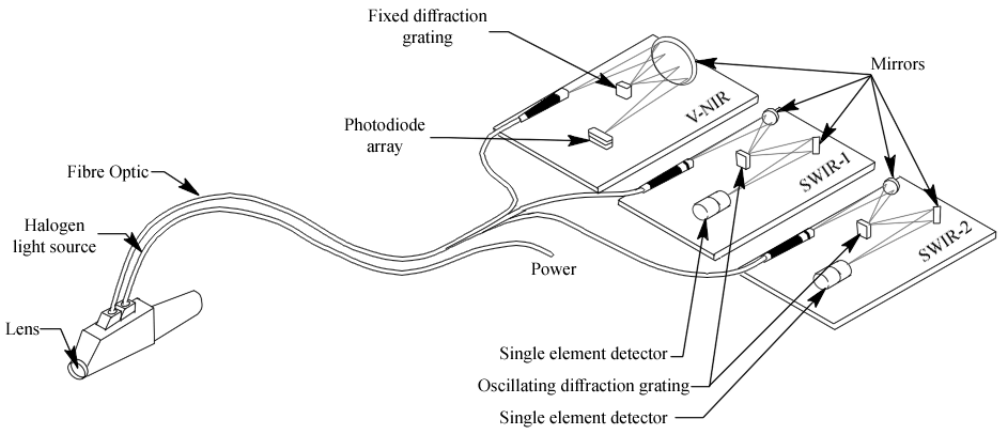


Figure 3.1.3 Schematic image of the optical paths of the ASD FieldSpec Pro spectroradiometer (modified from Mac Arthur et al., 2012). The ASD consists of three spectrometers, a VNIR and two SWIR spectrometers.

Figure 3.1.4 Image of the contact probe (including a halogen light source together with fibre) and its application in the field.

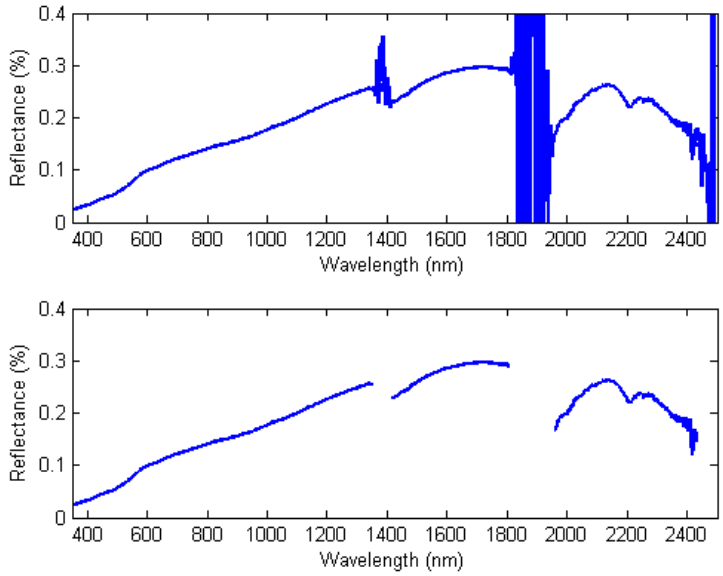


Figure 3.1.5 Spectrum of surface soil from Italy where the atmospheric windows are included (top) and removed (bottom). In the top figure, unrealistic features are obtained in the atmospheric windows due to the low intensity of the sunlight. In the bottom figure, the atmospheric windows are removed at bands around 1355 – 1420 nm, 1810 – 1955 nm and 2400 – 2500 nm. Such a correction does not have to be applied when an artificial light source is used.

To reduce atmospheric variability as much as possible and to obtain a uniform intensity of the light source, a so-called white reference spectrum should be measured. These white reference spectra measurements were repeated approximately every 10 minutes. For this white reference measurement, a white Spectralon reference panel, which has around 99% reflectance (Malvern Panalytical, 2002), from the ASD was used (Figure 3.1.6). Spectralon reflects more than 99% over a range of 400 – 1500 nm and more than 95% in the 250 – 2500 nm range. It has the highest diffuse reflectance of any known material or coating over the ultraviolet, visible, and near-infrared regions of the spectrum. More details on the use of white reference measurements are given in Chapter 3.2.1.

Figure 3.1.6 Image of the ASD Spectralon used during the fieldwork.

3.1.3 Soil samples

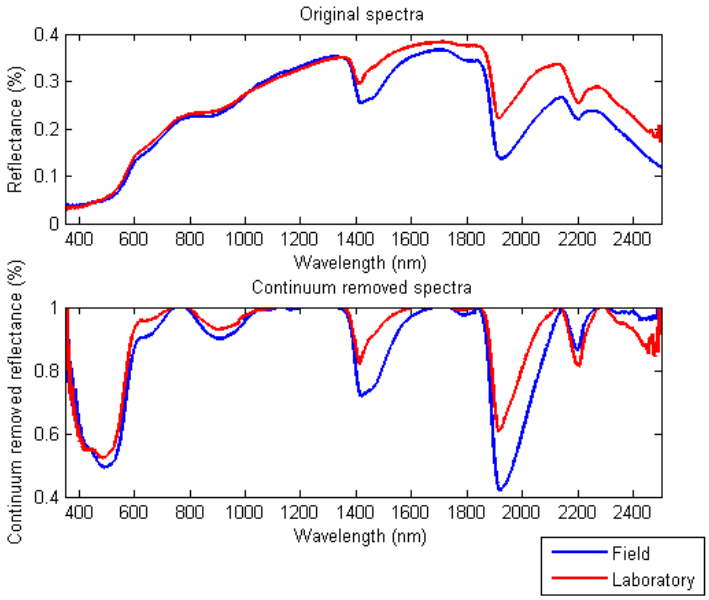


Figure 3.1.7 Field and laboratory spectral measurements of soil from Italy. The top figure shows the original spectra and the bottom figure shows the continuum removed spectra (refer to Chapter 3.2.2 for a detailed explanation of this correction).

The soil samples were collected from the ground surface to the deeper horizons of the soil pit. This was done to understand and obtain various spectral measurements of natural soils of the site. After spectral measurements were taken (with a halogen lighting ASD contact probe), around 300 g of soil from the measurement point was collected and taken to the laboratory. The collected samples were instantly double packed with plastic zip bags to avoid any loss of moisture and decomposition of organic material.

In the lab, the soil samples were air-dried. Drying the samples in the oven results in loss of some organic material and change in the chemical properties of the soil, this might also affect the spectral properties of the archaeological and natural soils. After the samples were dried, they were sieved below 2mm. By sieving <2mm, organic materials (grass etc.) and rock

fragments were excluded from the sample. Soil samples were gathered for XRF measurements (Chapter 3.4.2) and also reflectance spectra of these soils were measured to obtain laboratory measurements for the same soils as measured in the field.

Figure 3.1.7 illustrates the spectral difference between field and laboratory soil. While similar spectral features are found in both spectra, systematic differences are also observed. These differences are especially evident in the continuum removed spectra (for the details of the continuum removal see Chapter 3.2.2). The depth of the spectral features, particularly at longer wavelengths is systematically larger for the filed spectra compared to the laboratory measurements. The differences are probably related to the change in soil moisture and the different soil structure since the soil particles are finer and closer packed in the soil samples which lead to enhanced scattering and stronger reflection (Johnson et al., 1998; Salisbury and Wald, 1992). However change in soil moisture content does not alter soil composition (Stoner et al., 1980, 1982) and, therefore, field and laboratory measurements are comparable to each other. Since the difference between field and laboratory spectral measurements have minor effect in soil matter, this thesis will focus on mainly using field spectral measurements.

3.2 Pre-treatment of spectra

With the spectra obtained from the soil profiles, some pre-treatment should be performed before using these spectra for any spectral analyses. The methods illustrated in Figure 3.2.1 will be explained in more detail in this section.

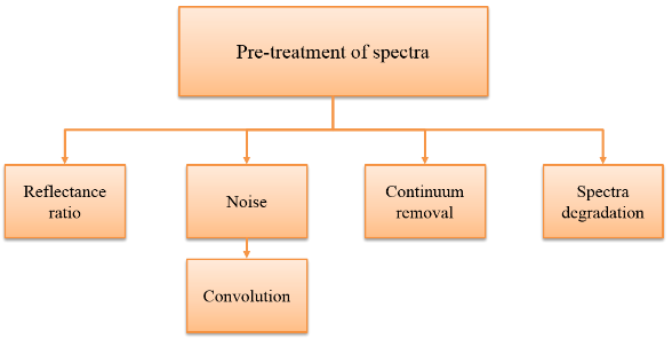


Figure 3.2.1 Flow chart of the pre-treatments of spectra that will be explained in Chapter 3.2.

3.2.1 Reflectance

In the end of Chapter 3.1.2, the white reference measurement was introduced. To correct the varying intensity, the reflectance spectrum measured (S_m) should be corrected using this white reference measurement (S_{wr}). To obtain the corrected spectrum (S_r), S_m is divided by S_{wr} as shown in the Equation [1]. Figure 3.2.1 shows how the original spectrum (S_m) change after white reference correction is applied.

$$S_r = \frac{S_m}{S_{wr}} \quad [1]$$

The main reason for doing such a correction is to have a uniform reflectance of the spectra gathered. The intensity of the halogen light source is not stable and varies according to the condition of the battery status. By diving the S_{wr}, one can correct also for these effects and thus calibrate the spectra.

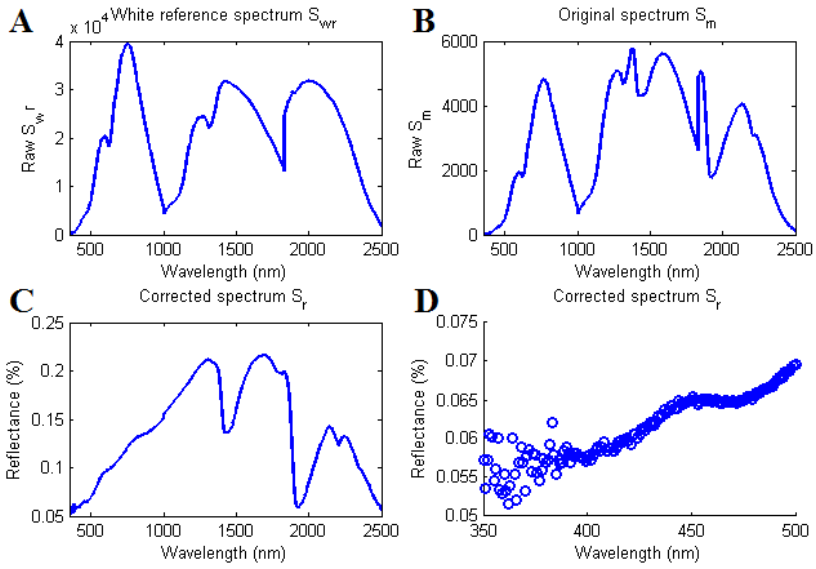


Figure 3.2.2 Example of a soil spectrum (B) and a white reference spectrum (A) measured with the halogen lamp. The corrected spectrum S_r is also shown (C). At short wavelengths the reflectance spectra are strongly affected by noise due to the low intensity in the 350 – 400 nm regions (D).

Before making the correction, there is one point to be noticed about the S_m spectra. Figure 3.2.2 shows spectra from a top soil measured at site 1 in Italy. The intensity (or in other words, the number of photons collected) in the lower wavelength region around 350 nm is very small. Thus no useful spectral information can be gathered at short wavelengths because the spectra mainly consist of noise (also in the corrected spectra). Figure 3.2.2.C represents the spectrum corrected according to the Equation [1]. Since low intensities are also found in the 2500 nm region, the first 50 nm last 100 nm were removed and only the wavelength 400 – 2400 nm range was used in this thesis.

3.2.2 Continuum removal

Following the white reference correction and the selection of a useful wavelength region, the continuum removal (CR) correction can be applied to the spectra, which is an important step of the spectral analysis for soil spectral measurements. This concept was introduced by Clark and Roush (1984) and is used to isolate particular absorption features in reflectance spectra. It

is a mathematical function that can be used to better represent the absorption features of interest. The continuum removal is a widely used in soil spectroscopy to isolate specific absorption features (Clark et al., 1984), for band depth analysis (Mielke et al., 2015; Noomen, et al., 2006; Viscarra Rossel et al., 2009; Zhang, 2010) and for correlating specific spectral absorption features to clay, silt and sand content (Curcio et al., 2013; Lagacherie et al., 2008). A study from Loum et al. (2016) showed how the effect of continuum removal process improves the analysis of soil properties.

The continuum can be calculated using different functions such as straight-line segments, Gaussian functions, polynomials or splines (Clark and Roush, 1984). Among these methods, in this thesis, continuum lines are calculated with straight line segments as shown in Figure 3.2.3 will be used.

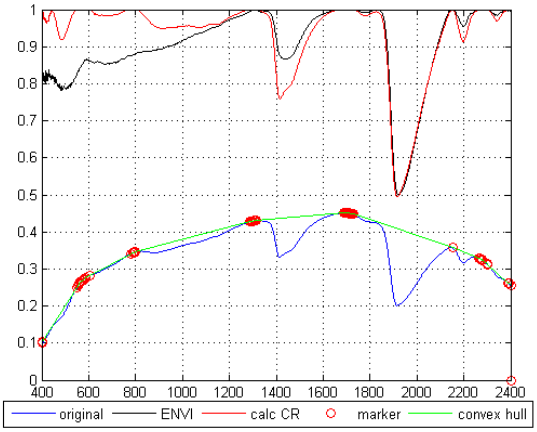


Figure 3.2.3 Spectrum of a burned material from Italy. The original spectrum is shown in blue line and the continuum removed spectrum is represented in red (MATLAB calculation) and black (ENVI calculation) lines. The red circles are the markers found in the original spectrum through the convex hull method and the green line represents the straight-line segments (continuum line) connecting these markers.

To perform the straight-line segment continuum removal, one should first construct a line called 'continuum line' between the peaks (which are referred as the marker points) of spectra as shown in Figure 3.2.3. Then the continuum removed value of a specific wavelength is the division between reflectance values at that wavelength over the value of the continuum line at that wavelength, which is mathematically represented in the Equation [2].

$$continuum\,removed\,spectrum(\lambda) = \frac{reflectacne\,value\,of\,the\,spectrum(\lambda)}{continuum\,line(\lambda)} \qquad [2]$$

Studies often calculate the continuum removal function using the ENVI (Environment for Visualizing Images) software (Huang et al., 2004; Loum et al., 2016; Richter et al., 2009; Shepherd and Walsh, 2002). However, the continuum removal through ENVI does not perfectly draw continuum lines (especially at the lower wavelength region) across all absorption features, because they have limited the number of continuum line processes to ensure that the spectrum does not emphasise poorly defined or noisy features (Harris Geospatial Solutions, 2004).). The black plot in Figure 3.2.3 shows the limitation of the continuum removal by ENVI. To overcome this problem, the continuum removal using the convex hull ('convhul' command) method through MATLAB is used in this thesis.

As shown in Figure 3.2.3, markers (red circle) are plotted for every convex hull creating straight lines (continuum lines, shown in green) between every marker points. Then the continuum removed spectrum is calculated according to Equation [2]. Here, the continuum removed spectra using MATLAB (red line) is compared to the spectrum created by the ENVI software (black line). Notice that the continuum removal using the convex hull method in MATLAB enhances the absorption features more than the ENVI program.

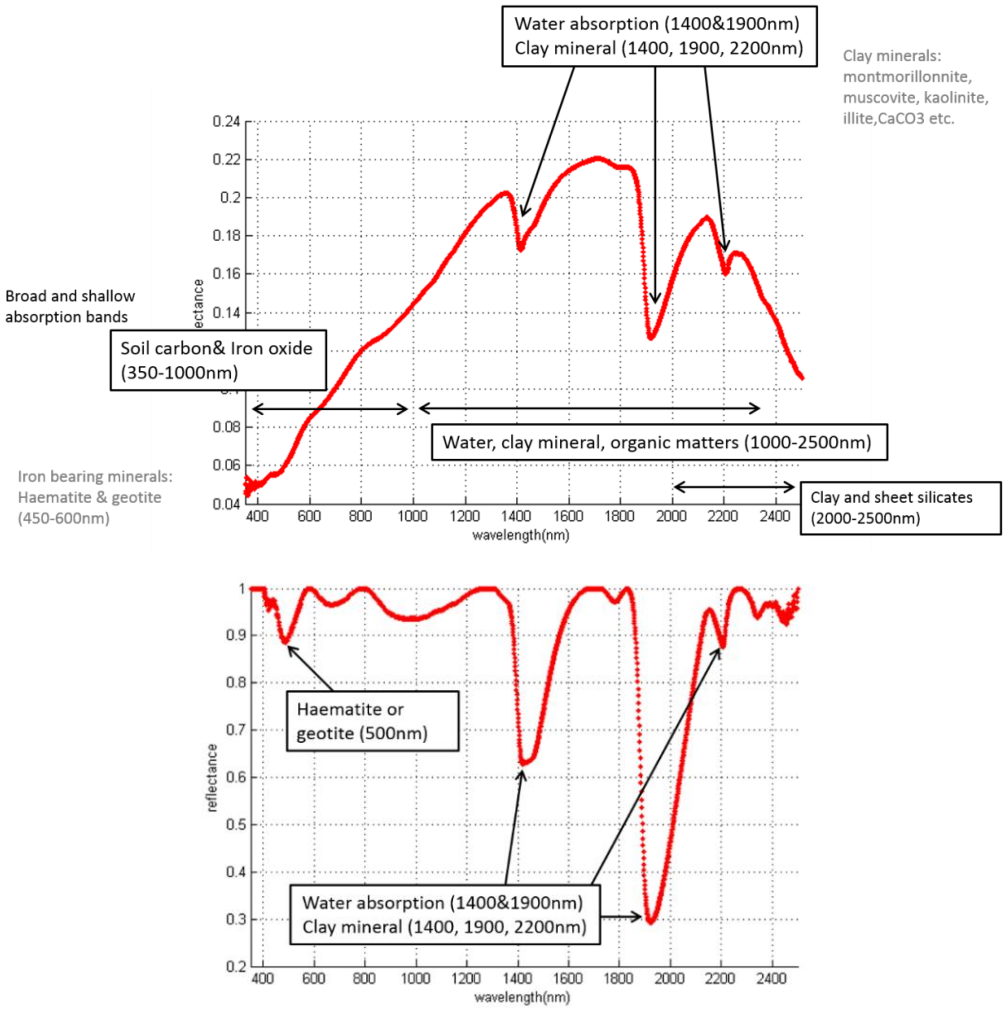


Figure 3.2.4 Spectrum of soil from the Ap horizon in Italy. The figure shows how absorption features of specific minerals in soil (or other absorption features) which were not clearly visible in the original spectrum (top figure) are enhanced by the continuum removal process.

By doing the continuum removal on the soil spectra, absorption features (e.g. from minerals) of the spectra are enhanced. This allows soil scientist to observe soil properties in more detail and use band-depth identification. Figure 3.2.4 shows how the absorption features of the min-

erals in soils (and also other absorption features) are enhanced by the continuum removal process.

Experiments have shown that specific absorption features in the vis-NIR provide information on the chemical, physical and mineralogical composition of the soils (Ben-Dor and Banin, 1995a; Bowers and Hanks, 1965; Viscarra Rossel and McBratney, 1998).

Figure 3.2.4 also explains the main spectral characteristics of soil spectra. The absorption features in the visible range (400 – 780 nm) are associated with iron bearing minerals such as haematite or goethite (Sherman and Waite, 1985; Stenberg et al., 2010) where the second-derivative peaks are observed around 420 and 535 nm (Sellitto et al., 2009). For haematite, absorption features can be observed at 550, 630 and 860 nm and for goethite at 480, 650 and 920 nm (Morris et al., 1985). Iron (Fe) absorption bands are mainly dominant around 500 and 900 nm (Richter et al., 2009). Soil organic matter has a broad absorption throughout the visible and near infrared regions (Matneya et al., 2014) and, therefore, soils with high organic matter content tend to have lower overall reflectance (Bartholomeus et al., 2008; Stenberg et al., 2010). Also, water has a strong influence on the spectra by showing dominant absorption bands at 1400 and 1900 nm (Clark et al., 1990; Viscarra Rossel et al., 2011) and weaker bands in other parts of the spectra (Weidong et al., 2002). Absorption near 2200 to 2250 nm are mainly caused by clay minerals (Viscarra Rossel and Behrens, 2010; Viscarra Rossel et al., 2011) and sheet silicates such as chlorite and biotite (Rodger et al., 2012).

3.2.3 Smoothing (convolution)

After the application the of continuum removal, a high noise level in the lower wavelength region between 400 – 450 nm is still observed (refer to Figure 3.2.2). Therefore, smoothing (convolution) was applied to reduce this noise.

The smoothing was done using convolution with a Gaussian kernel. In MATLAB, the 'gaussmf' function was used to calculate the kernel (bell-shaped Gaussian hump). The Gaussian function depends on two parameters σ and c as expressed in Equation [3],

$$\mathrm{G}(x;\ \sigma, c) = e^{\frac{1(x-c)^2}{2\sigma^2}} \qquad [3]$$

Here, σ can be calculated from the full width at half maximum (FWHM) expressed in Equation [4] where c is the centre point of the spectrum.

$$\sigma = \frac{\mathrm{FWHM}}{2\sqrt{2\ln 2}} \qquad [4]$$

Various kernel values with different FWHM were applied. Note that the ASD spectrometer has an FWHM of 3 nm in the visible range (refer to Table 3.1).

Before applying the smoothing, one must verify that the spectral features at the low wavelength region are a real noise and not a spectral signature of the soil reflectance. To do this, a high-pass filter is applied to the spectra. A high pass filtering can be applied by the Gaussian

smoothing method as in Equation [5].

$$High\ pass\ filtered\ spectrum = Original\ spectrum - Smoothed\ spectrum \qquad [5]$$

Since the noise is occurring in the low wavelength region of the visible spectral range and the resolution of the spectrum at this range is 3 nm (refer to Table 3.1), an FWHM of 3 nm is used for the high pass filter.

Figure 3.2.5 shows the high pass filtered result for randomly selected spectra from Italy. Notice that the wiggling feature (which is assumed to be derived from the low-intensity part of the spectra at short wavelengths) is very systematic and is shown in all spectral measurements not just in the soil spectra, but also in the white reference spectrum (S_{wr}). This indicates that it is not a soil signature. It might be a spectral feature caused by the instrument. To investigate this in more detail, only the white reference spectra measured from different countries (Italy and Hungary) are compared in Figure 3.2.6.

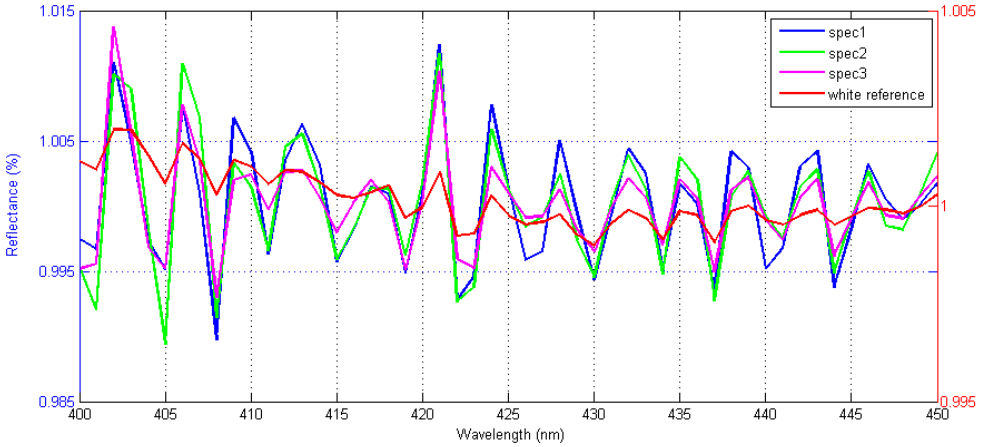


Figure 3.2.5 High pass filtered spectrum in the 400 – 450 nm range. A noise like- systematic pattern is shown in three different spectra from Italy as well as in the white reference spectrum (S_{wr}), which is the measurement of the light source (halogen light) against Spectralon.

The top figure of Figure 3.2.6 indicates that when a number of white reference spectra from Italy and Hungary are compared, there are two major 'wiggling' patterns. Therefore, only the white reference spectra from Italy were chosen in the bottom figure of Figure 3.2.6. It is obvious that the spectra from Italy are almost exactly identical. Similar conclusions are also drawn for the white reference spectra measured in Hungary.

There was an around six-month interval between the Italy campaign and the Hungary campaign. A detailed field description is explained in Chapter 4 for Italy and Chapter 5 for Hungary, but one important thing to mention is that spectral measurements for the sites in Italy were taken in the year 2012 and for Hungary, the measurements were taken in the year 2013. Each field campaign took around a month. Although the wiggling pattern is a most probably an instrumental artefact (either from the spectrometer itself or the light source), it is not iden-

tical for both campaigns. The results presented in Figure 3.2.6 indicate that the pattern does not change over a short-term (probably within one month) but change over a longer term (more than six months).

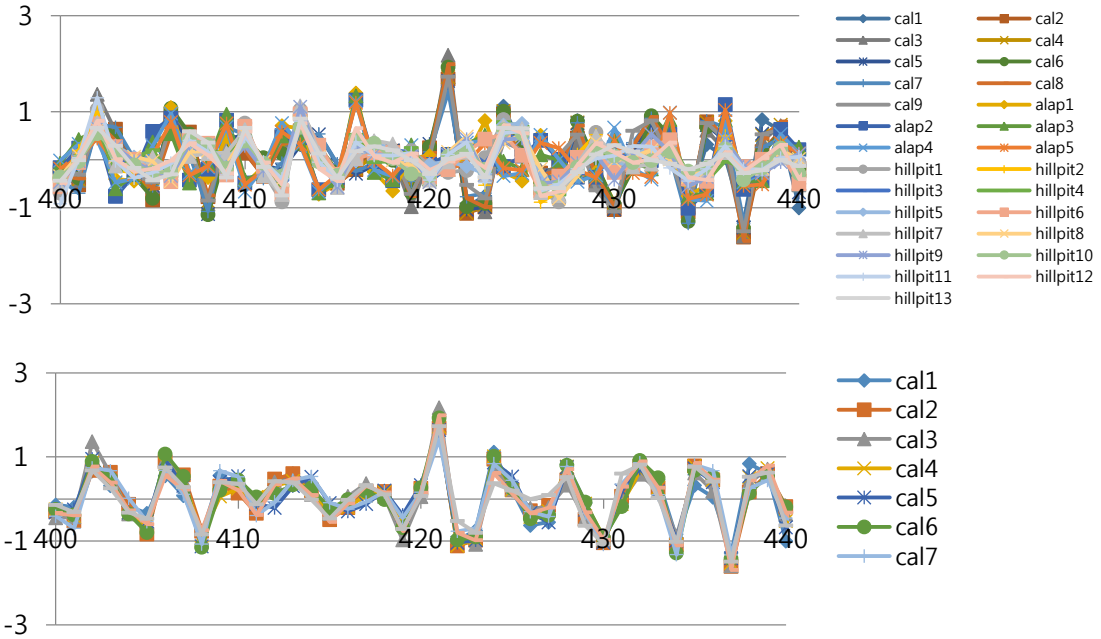


Figure 3.2.6 White reference spectral measurements. Top: White reference spectral measurements made from two different countries Italy and Hungary (measurements taken in different years with the same ASD instrument). Bottom: White reference spectral measurement only from Italy (measurements made within a few days).

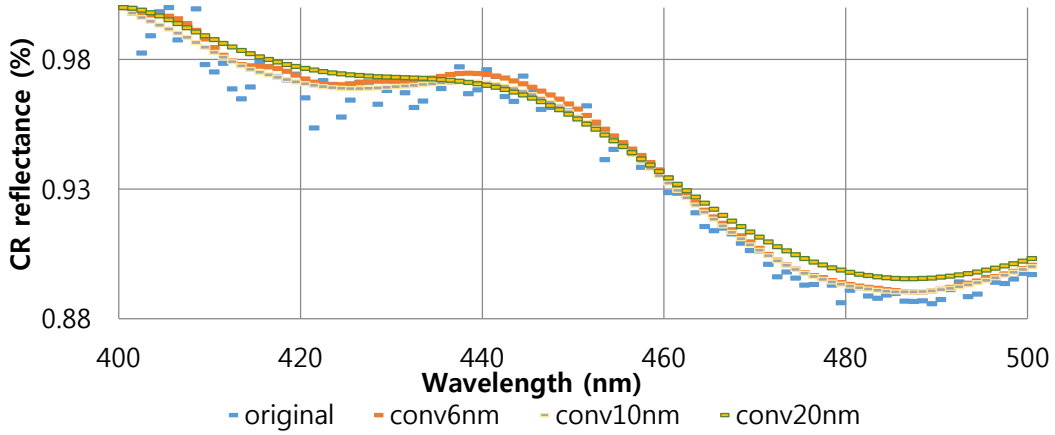


Figure 3.2.7 Continuum removed spectra convoluted with Gaussian kernels of different FWHM values (6 nm, 10 nm and 20 nm). The convolution with a kernel of 6 nm FWHM already effectively reduces the spectral artefacts in the low wavelength region (400 – 500 nm).

It is not possible to correct this systematic instrumental spectral feature by simply subtracting

it from all the spectra because there its amplitude shows some variation. Therefore, to minimise this spectral feature, smoothing is applied to the wavelength region where this is strongest (in the 400 – 450 nm regions). Figure 3.2.7 shows the continuum removed spectra smoothed by different convolution levels. The convolution for smoothing was applied ± 20 nm around the wavelength range: for the spectral range of interest (400 – 450 nm) the convolution is applied to the spectral range 380 – 470 nm. Convolution with different convolution kernels was applied to test which convolution kernel is best in reducing the instrumental artefacts while keeping the true soil spectral signature.

Figure 3.2.7 shows that in the 400 – 420 nm regions, convolution with a Gaussian kernel of an FWHM of 6 nm still shows slight remaining spectral artefacts. These features vanish for a Gaussian kernel with a FWHM of 10 nm. When a kernel of an FWHM of 20 nm is used, the spectrum is strongly smoothed and the spectral reflectance features seem to be strongly suppressed. Therefore, in the following a convolution kernel of 10 nm FWHM is used to correct these 'wiggling' patterns.

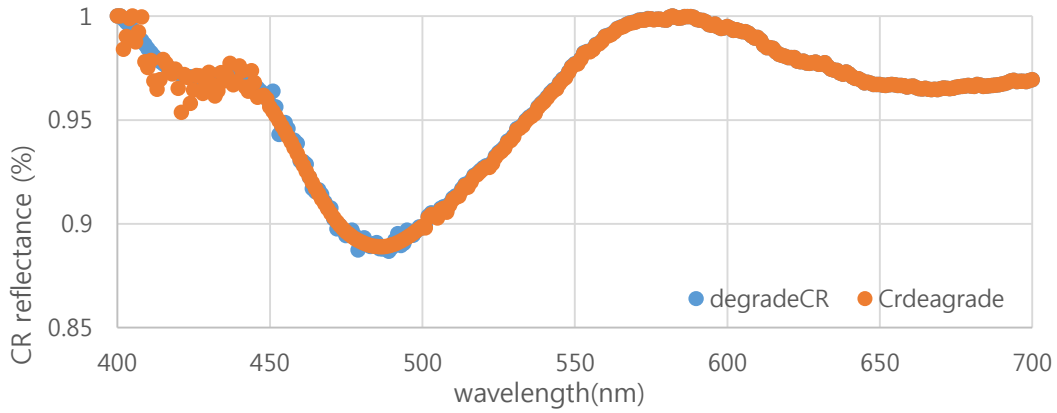


Figure 3.2.8 Comparison of results, for which the order of the continuum removal and smoothing are changed. Blue: Application of the Continuum removal after the spectrum was first convoluted. Orange: Convolution of a spectrum after first the continuum removal was applied.

One important aspect to note it that the spectra should be smoothed before being continuum removed. This is because the continuum removal method is applied to enhance the absorption feature, but it also enhances any other spectral feature (including noise or spectral artefacts), which is not what one would like to use for the further spectral analysis. Figure 3.2.8 shows a comparison of spectra for which the order of smoothing and continuum removal is reversed. Notice that when the smoothing is applied after the continuum removal (orange), the spectral artefacts at the low wavelength region are not corrected at all.

3.3 Principal component analysis (PCA)

3.3.1 Concept of principal component analysis

Principal component analysis (PCA) is a multivariate, statistical technique that decreases the data redundancy by transforming a set of correlated variables into a new set of variables called the principal components (PC) that are uncorrelated. The aim is to identify the strongest patterns in a dataset. Researchers generally apply PCA to reduce the data complexity and extract valid information, not only for soil applications (Linker et al., 2005; Reid and Spencer, 2009; Salehi and Zahedi Amiri, 2005; Singh et al., 2011; Viscarra Rossel, 2008a) but also for archaeological applications (Aqdus et al., 2008; Doneus et al., 2014; Panishkan et al., 2012; Traviglia, 2006; Wells et al., 2007). Due to the overlap of various spectral dependencies of absorption and scattering processes, spectral features in reflectance spectra can usually not unambiguously be assigned to a set of different laboratory reference spectra. Thus for the interpretation of the measured reflectance spectra PCA is a promising approach. Figure 3.3.1 illustrates the flowchart of how the PCA is developed throughout this thesis.

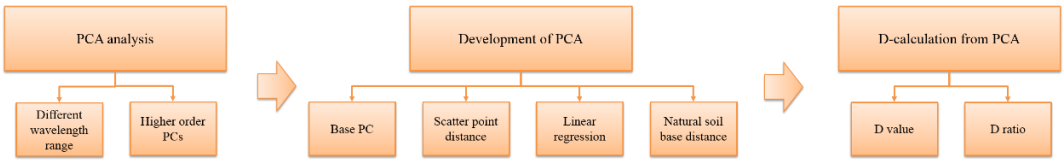


Figure 3.3.1 Flowchart explaining the detailed steps of a modified PCA analysis performed in this thesis.

Figure 3.3.2 shows how PCA creates a new dimension from the original dataset. The first PC (the variable on the first axis) accounts for the greatest variability in the data set and thus containing most information. Every succeeding principal component contains a decreasing amount of variance in the remaining data. Note that these principal components are uncorrelated and therefore orthogonal to each other. The higher order PCs often represent noise, although they still might represent very rare spectral features (Doneus et al., 2014; Traviglia, 2006). Viscarra Rossel et al. (2006a) showed that in soil spectroscopy, the first three PCs accounts for approximately 75% of the variation in the data.

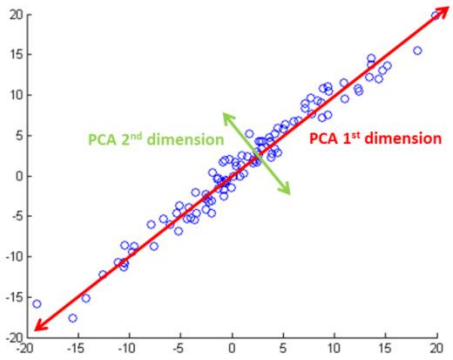
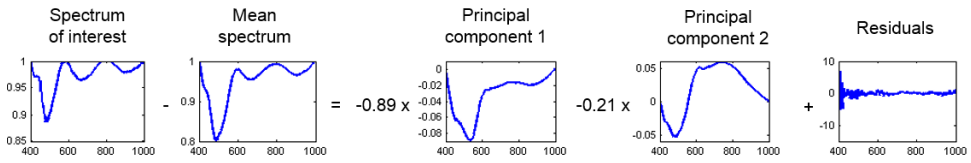


Figure 3.3.2 Artificially created 2D dataset (pairs of integer values varying between -20 to 20) showing the direction of the PCA dimensions. The first PC (red) represents the direction where the dataset is most elongated. The Second PC (green) represents the second most important direction and it is orthogonal to the first PC.

The number of principal components is less than or equal to the number of original variables. A set of spectra *A* has *a* number *i* of samples and *j* variables. Here, the variables are the individual spectral points within the wavelength interval and therefore related to the spectral resolution of the spectrometer. Among a set of spectra *A*, PCA calculates the most common spectral feature (first PC) and second common spectral feature and so on. Figure 3.3.5 illustrates the PCA of a sample spectrum from a set of spectra. As shown in the figure, a spectrum can be explained by the PCA of the most common features of the whole spectral data set and the respective weights of the PC (score values). Figure 3.3.3 shows a simple mathematical expression of the PCA.



$$\text{spectrum} - \text{mean spectrum} = \text{score}_1 * \text{PC}_1 + \text{score}_2 * \text{PC}_2 + \text{residuals}$$

Figure 3.3.3 Example of a PCA of a spectrum within a set of spectra with two PCs. The mean spectrum is the average spectrum of the whole dataset. PC stands for the principal component (also called loading) and the score is the corresponding weight of the PC.

PCA calculation is carried out by calculation the covariance matrix, in a first step, the data are standardised by mean centring (or mean subtraction). This is to ensure that the first PC does not correspond to the mean of the data but describes the direction of the maximum variance. Figure 3.3.4 shows the difference between a mean spectrum and the first principal component of a set of spectra.

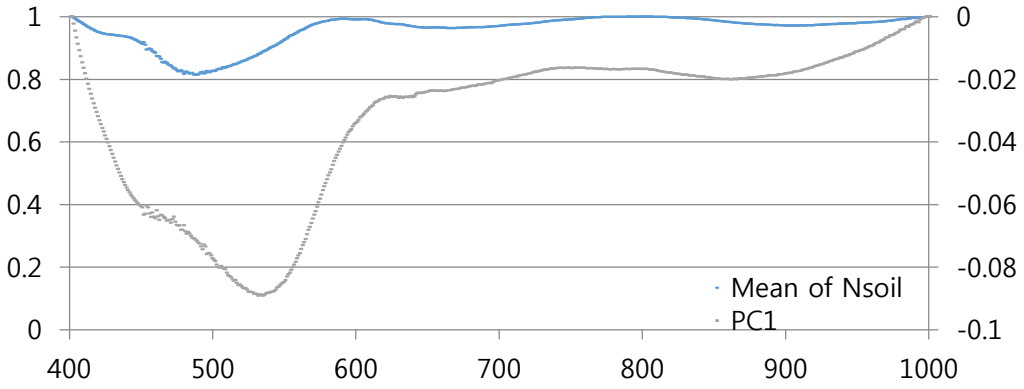


Figure 3.3.4 Comparison between a mean spectrum and the first principal component of a set of spectra. The principal component stands for the direction of the maximum variance but not the mean of the spectra.

From the covariance matrix *S* of the auto-scaled matrix of *A*, eigenvectors and eigenvalues

can be obtained and the principal components (loadings) are derived as;

$$PC_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ij}X_j = \sum a_{ij}X_j \quad [6]$$

PCA transforms the original spectra to a new coordinate system (Figure 3.3.5) where the results of the spectra can be plotted as a function of their score values of the different PCs (Fig. 3.3.5 middle).

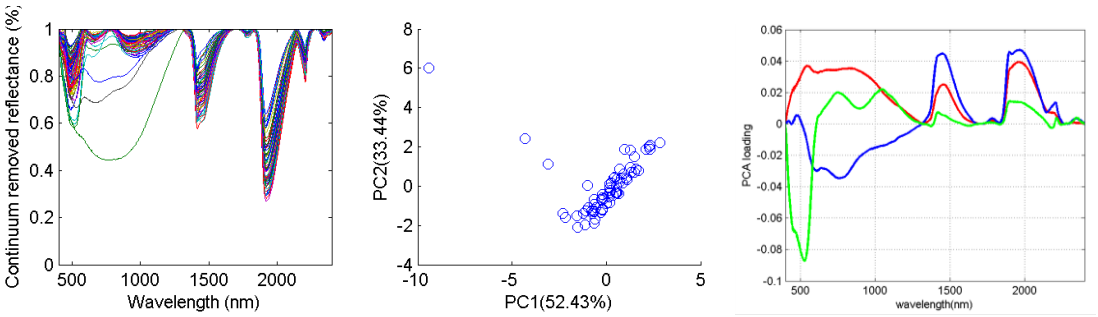


Figure 3.3.5 Illustration of PCA results of measured reflectance spectra. Left: The group of spectra used in the PCA. Middle: PCA score plot of the first two principal components of all spectra. Spectra with similar properties tend to cluster together. Right: The first three principal components representing the most common features among all spectra.

Figure 3.3.5 shows how the PCA results will be represented in this thesis. Already by looking at the score plot (middle of Fig. 3.3.5), one can quickly observe whether archaeological materials are separated from the natural soils.

3.3.2 D value calculation

So far, these PCA results are represented visually. In order to build up an algorithm for the identification of archaeological materials, the distances (D) between the spectra have to be quantified. This method will be called the D value calculation in this thesis.

The basic hypothesis for the D value calculation is that 'all' natural soils share similar spectral signatures and these are mostly represented by the first three principal components of the group of natural soil spectra. Another hypothesis is that any soil spectrum measured represents a natural soil spectrum. Therefore, the measured spectrum should contain similar spectral characteristics (the first three principal components) of natural soils. For a spectrum with different spectral characteristics, it is assumed to be an anomaly, e.g. a non-natural material, which in this case can be assumed to be an archaeological material.

The D value calculation can be summarised in four steps;

1) Obtain the first three PC values for a group of natural soils.

2) Calculate the score values for individual spectra using the PCA.

3) Recalculate the individual spectra using the score values and PCs obtained from the PCA.

4) Calculate the difference between the recalculated spectra and the original spectra through the Euclidean distance method.

The group of natural soil spectra used for calculating the first three principal components are referred as N_{soil} in this thesis (they should not be confused with the original natural soil spectra).

3.3.2.1 Selection of natural soils (N_{soil}) for the D value calculation

Soils have various physical and chemical properties as well as types. Thus the soil spectra differ greatly by region and even vary within the same soil profile. Since the PCs used for the D value calculation depend on the selected group of natural soil spectra, it is essential to know what kinds of natural soils (N_{soil}) should be used. In this thesis, various types of groups of natural soils (N_{soil}) are used, either from the spectral measurements of local soils or global soil spectra databases.

Usually, the local soils (N_{soil}) are gathered around the archaeological site of interest. For example, for investigating site 1 in Italy, N_{soil} contains natural soils from the site in Italy. It is also interesting to use natural soils from a totally different country containing different environmental conditions regarding location, weather and soil type. For the same example (site 1 in Italy), N_{soil} from Hungary is also used.

Figure 3.3.6 shows the variance of each principal component (up to 10^{th} principal component value) and spectra of the first three principal components for natural soils from Italy (A) and Hungary (B). One can observe that the first PC accounts for more than 90% of the variance of the spectral features for the natural soil spectra in both Italy and Hungary. For the natural soils in Italy, from PC4 on, the variance drops below 1% and for Hungary it drops below 1% even from PC3 on. Therefore, the D value calculation only the first three principal components are used, since higher number PCs are mainly representing noise. These first three principal components are shown in the right side of Figure 3.3.6.

The PCA results of natural soil spectra from Italy and Hungary indicate that the first principal component dominates more than 90% of the variance of the spectral features, indicating that natural soils are rather similar for each location. However, it is also found that the first PCs of Italy and Hungary show different spectral patterns (left figures in Figure 3.3.6). For Italy, the first principal component (red) of natural soils shows a strong absorption feature between a 400 – 600 nm range with a peak around 550 nm, which might be related to the characteristic of the red-coloured Mediterranean soils (Costantini and Damiani, 2004). Beyond 600 nm, most of the spectral feature disappears leaving only a slight variation. However, for the natural soils in Hungary, the first principal component is a broadband feature dominating the whole wavelength range (400 – 1000 nm) with a small spectral feature around 500 nm. Interestingly, the strong absorption peak (550 nm) observed in PC1 of the Italian soils is also

found in the second and third PC of the Hungarian soils. However, the overall impression is that the most common spectral features of natural soils differ depending on their type and location. Figure 3.3.7 shows the first three principal components of the natural soils and archaeological materials gathered in Italy. At first sight, these spectra seem to have similar overall spectral patterns. Nevertheless, important differences in the details of the spectral signatures are also observed. These differences are important for the discrimination between natural soils and archaeological materials.

The natural soils shown above are soils from only one specific site and, therefore, represent the spectral features of a specific regional soil type. What if additional various soil types are used as one group of natural soils? Figure 3.3.8 shows PCA results (variance and principal components) for global natural soils: (a) natural soils gathered both from Italy and Hungary (IT+HUN) and (b) the ISRIC spectral library (ICRAF and ISRIC, 2010).

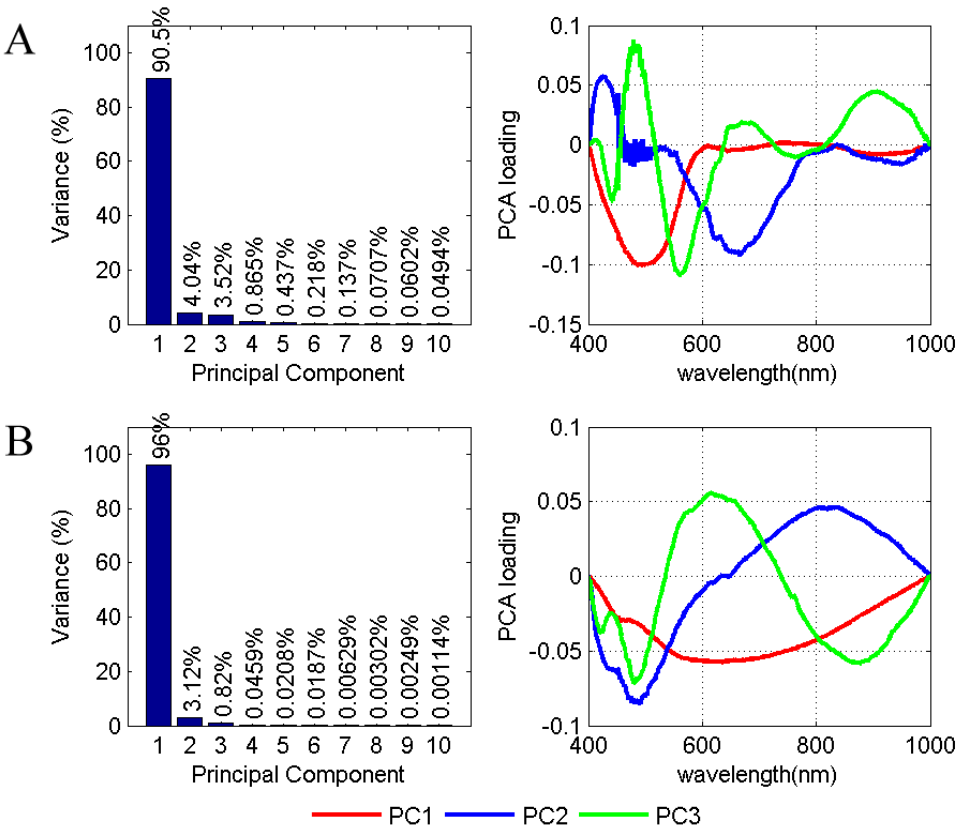


Figure 3.3.6 Variance plot (left) and the first three principal components (right) for natural soils gathered in Italy (A) and Hungary (B). For both countries, the first principal component accounts for more than 90% of the variance and from the third or fourth principal component, the variance drops below 1%. Therefore only the first three principal components are used in the D value calculation. The spectra of the first principal components from Italy and Hungary show different patterns indicating that their physical and chemical characteristics of the natural soils gathered in these sites are different.

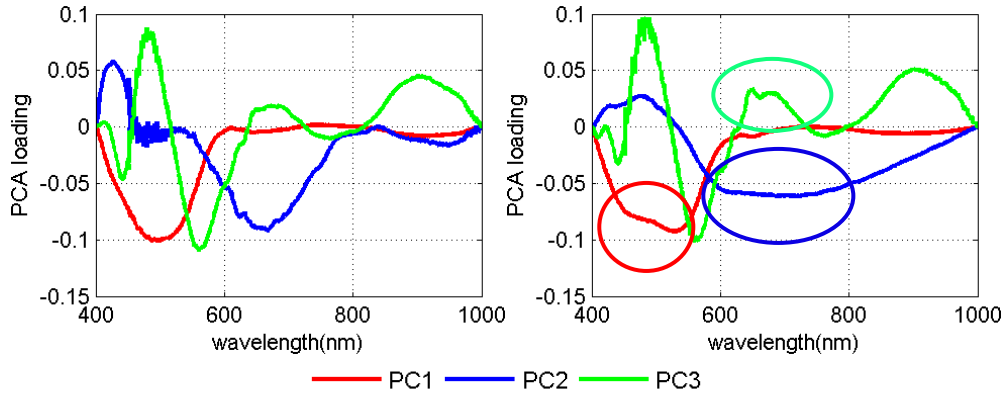


Figure 3.3.7 Spectra of the first three principal components for natural soils (left) and archaeological material (right) from the sites in Italy. Both figures show similar general spectral patterns, but also important differences in the detailed spectral signatures (marked in red, blue and green circles for PC1, PC2 and PC3).

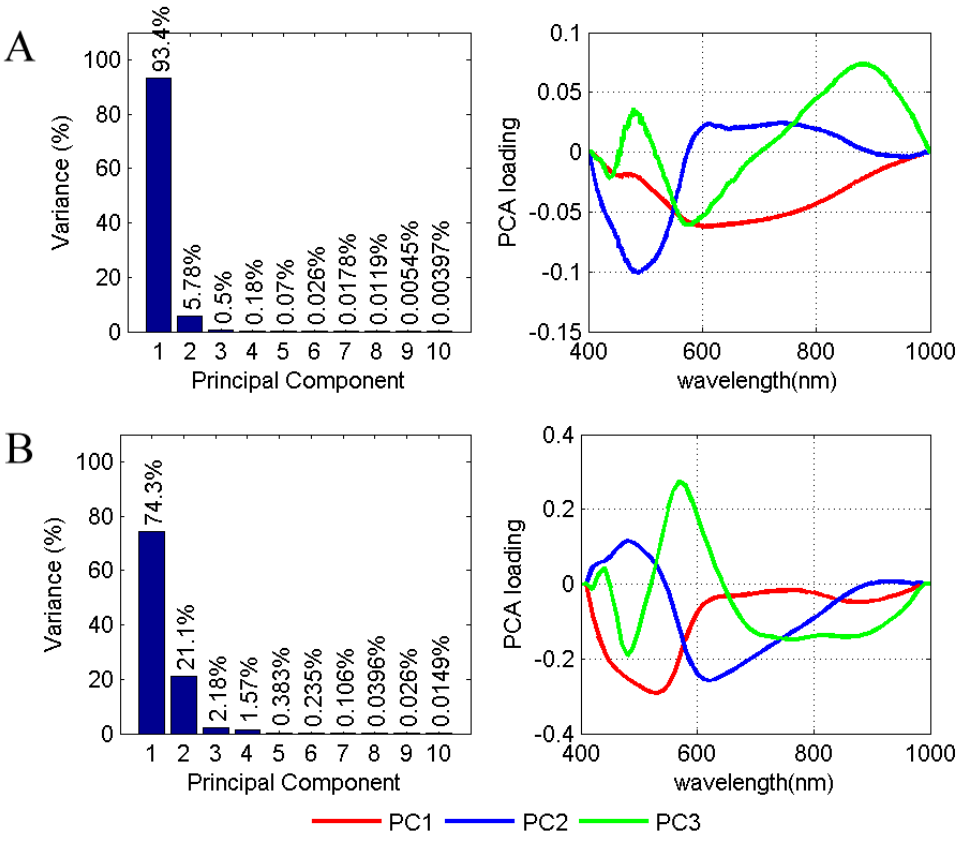


Figure 3.3.8 Variance (left) and the first three principal components (right) for natural soils from (A) the sites in Italy and Hungary (IT+HUN) and (B) the ISRIC spectral library (ICRAF and ISRIC, 2010). The first three principal components account for more than 90% of the total variance for both N_{soil} groups.

N_{soil} of IT+HUN represents a small scale global natural soil group. It includes natural soils from the original site (from Italy) as well as natural soils from another site (Hungary) with an entirely dissimilar spectral pattern. This N_{soil} consists of 171 natural soil spectra in total. In addition, as a global selection of natural soils, N_{soil} from the online spectra library (ICRAF and ISRIC, 2010) are used. Here, the natural soils gathered from the two measurement sites are not included. The spectral library contains 785 soil profile (4437 samples) spectra, which are gathered from 58 countries and four continents. They are measured with the ASD FieldSpec FR spectrometer (Garrity and Bindraban, 2004) over a wavelength range from 350 to 2500 nm. This spectrometer is a previous version of the one used in this thesis which has the same spectral resolution and sampling interval. However, the soil spectra from the ISRIC were re-gridded from the original spectra. The given spectra show reflectance values only at ten nanometre steps (in a total of 216 data points for each spectrum), which leads to a slight undersampling of the original spectra. This leads to 61 data points for the wavelength range 400 – 1000 nm (originally there were 610 data points). Also, the spectra from the ISRIC spectral library are laboratory soil spectra measurements, which were air-dried and sieved under 2 mm (Garrity and Bindraban, 2004). Therefore, in addition to the different spectral sampling, they also represent different properties compared to the spectra gathered in the field for this thesis (for a detailed description of the differences, refer to Chapter 3.1.3).

Figure 3.3.8 shows the PCA results (variance plot and the first three principal components) for the N_{soil} from IT+HUN and the ISRIC spectral library. For the N_{soil} from IT+HUN the broadband feature of the first PC (red) is similar to the first PC of the Hungarian soils (Figure 3.3.6), while the strong absorption feature at the 400 – 600 nm range of the second PC has a similar pattern as the first PC of the Italian soils. Interestingly, the PC1 from the ISRIC spectral library is similar to the PC1 of the Italian soils.

These results indicate that soils can have a very different "most common" spectral feature, but still share some similarities within the first three principal components. Beyond PC3, the contribution is again lower than 1% and most likely represents noise and features that are not well understood at current stage. Therefore, for the D value calculation, only the first three principal components will be used. Such principal components obtained from groups of natural soils will be named *NPC* (natural principal components) in this thesis.

3.3.2.2 Recalculation of the original spectra based on the PCA results of natural soils (N_{soil})

The next step is to calculate the score values according to the selected N_{soil} and use these score values to recalculate the spectrum of interest. First, score of a dataset can be calculated as:

$$T = A_1 * NPC_1 + A_2 * NPC_2 + A_3 * NPC_3 \quad [7]$$

Where the score is the summation of spectra *A* and principal component values of the natural soils (N_{soil}). For the group of spectra of interest, score values are calculated and will be re-

ferred to as *T*. According to the mathematical basis of the PCA as introduced in Chapter 3.3.1, the score of a spectrum is expressed by Equation [7].

After obtaining the score values, the spectrum can be recalculated using the *T* and *NPC* values. A spectrum is composed of the principal components and their weights (score), see also Figure 3.3.3. If only the first three PCs are considered (see above), the recalculated spectrum S' can be expressed by:

$$S' = S_m + T_1 * NPC_1 + T_2 * NPC_2 + T_3 * NPC_3 \quad [8]$$

where S_m is the mean of N_{soil}. This recalculated spectrum *S'* contains the spectral characteristics of the natural soils (N_{soil}) and thus is similar to the original spectrum *S* if that spectrum measures natural soil.

3.3.2.3 Calculation of the difference (D-value)

Finally, the difference between the original (S_λ) and the recalculated (S_λ') spectrum will be calculated. Here the Euclidean distance (referred to as D value) is calculated as expressed as Equation [9].

$$D = \sqrt{\sum_{\lambda} (S_\lambda - S'_\lambda)^2} \quad [9]$$

Figure 3.3.9 illustrates the comparison of original and recalculated spectra for two examples (a natural soil spectrum and an archaeological soil spectrum).

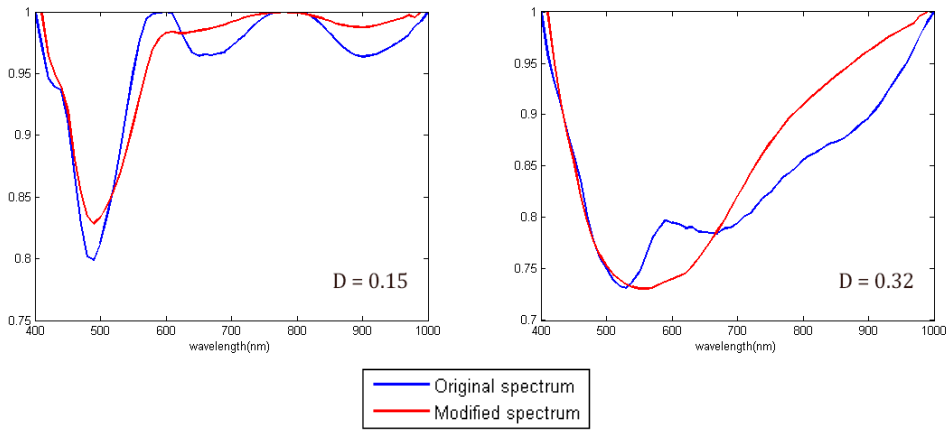


Figure 3.3.9 Examples of original spectra and recalculated spectra of natural soil (left) and burned archaeological material (right). Here N_{soil} from Italy was used for the D calculation. The D value for the natural soil is 0.15 and for the archaeological material is 0.32.

The D values for both types of spectra are different with the higher D value for the archaeological material. This confirms the hypothesis, that the difference between the original and

recalculated spectra is small, only if the selected spectrum is a natural soil. If not, it shows that the spectrum is more likely to be a non-natural soil, probably an archaeological material. However, still a threshold for the D values has to be defined, which separates spectra of natural soils from other spectra. Depending on the site condition and soil type, the D values among natural soils will vary and might eventually be similar to the D values of archaeological materials. One possibility to overcome this problem is to calculate the ratio of the D values of selected spectra.

3.3.3 Calculation of the D ratio

From Chapter 3.3.2, it became clear that it is, in principle, possible to use the D calculation method for the determination whether a sample is likely to be archaeological material or not. However, this D value is very dependent on the archaeological site and the selection of N_{soil}. Since the D value is just a distance of a sample a normalisation has to be applied to derive a more universal. One possibility is to calculate the ratio (D_{ratio}) between the D values of a sample and those of natural soils (D_{nat}).

$$D_{ratio} = \frac{av.D_{arch}}{av.D_{nat}} \quad [10]$$

Equation [10] expresses the D_{ratio} calculation using the averaged D values obtained from a site of investigation. If D_{arch} represents an archaeological material, then the D_{ratio} value should be larger than 1. If not, the method does not work. Also from the obtained D_{ratio} values, one can estimate the range of D_{ratio} values for specific archaeological types.

In the following chapters (Chapter 4 and Chapter 5), the D value and D ratio calculation methods will be applied to five different archaeological sites to 1) show that the methodology works and to 2) obtain D_{ratio} values for specific archaeological sites.

The final aim is to obtain a D_{ratio} value which can be used universally on different sites. Using such a universal D_{ratio} will allow to decide whether a spectrum from an unknown site represents probably an archaeological material or not.

Since in equation [10] the D_{ratio} value is calculated using the averaged D_{arch} and D_{nat} values, the error ranges are calculated by the error propagation method as expressed in Equation [11]. Here, the error range of $_{Darch}$ and D_{nat} values are the maximum and minimum values.

$$\frac{\Delta D_{ratio}}{D_{ratio}} = \sqrt{\left(\frac{\Delta D_{arch}}{D_{arch}}\right)^2 + \left(\frac{\Delta D_{nat}}{D_{nat}}\right)^2} \quad [11]$$

3.3.4 Other methods for the interpretation of the PCA results

In this section other methods, which were not (yet) successful for the application of distinguishing archaeological features from the surrounding soil, are presented. These methods shared the same intention as the D value and D ratio calculation (Chapter 3.4.2) but did not yield as successive results.

3.3.4.1 Base PC

The application of the PCA to discriminate archaeological materials from natural soils is not always successful. For some sites, a clear separation between natural soil clusters and archaeological materials was observed, but for some not (refer to Chapters 4 and 5). Therefore, to improve the PCA application, one can try to use the PC results from a 'well-separated site' (a site where archaeological materials were clearly separated from natural soils) for the site of interest.

The well-separated site will be called the base site and the principal components of this site will be the base PCs. Using the base PC, one can recalculate the spectra from the site of interest according to Equation [12] where only the mean spectrum and the PC are changed from the original PCA equation explained in Figure 3.3.3.

$$spectrum = mean\ spectrum_{base} + SC_1 * PC_{1,base} + SC_2 * PC_{2,base} + residuals \qquad [12]$$

With the new recalculated spectra, PCA can be performed providing new score values that can be plotted in the score plot and probably might improve the PCA results. The results of this method are presented in Chapter 4 and 5 for archaeological sites in Italy and Hungary which produced unsatisfying PCA results.

3.3.4.2 Regression line through the cluster of natural soils

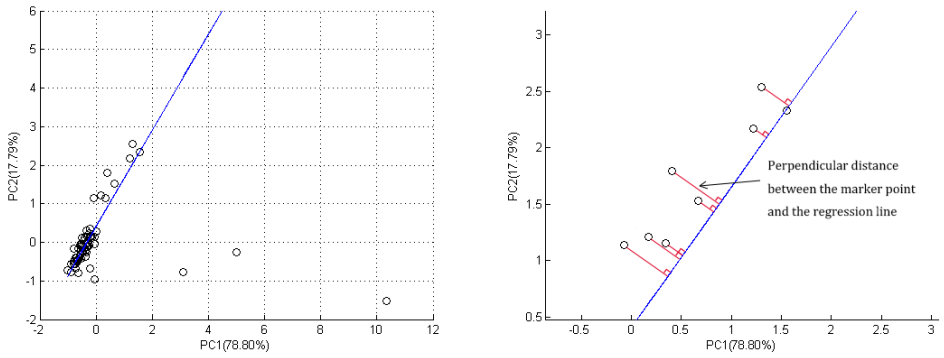


Figure 3.3.10 Left: Score plot of a PCA for a site in Italy. A regression line is fitted with only the results for natural soil spectra. Right: zoom of the figure on the left side. The distances between the individual samples and the regression line are shown.

This method uses a regression line through the natural soil cluster to calculate the minimum

(perpendicular) distance between the samples points (individual spectra) and the regression line. The hypothesis is that, since the regression line is derived from only natural soil spectra, the distance of archaeological samples will be larger than for natural soils. Figure 3.3.10 illustrates this method for a group of soil spectra gathered in Italy.

3.3.4.3 Distances from the centre point

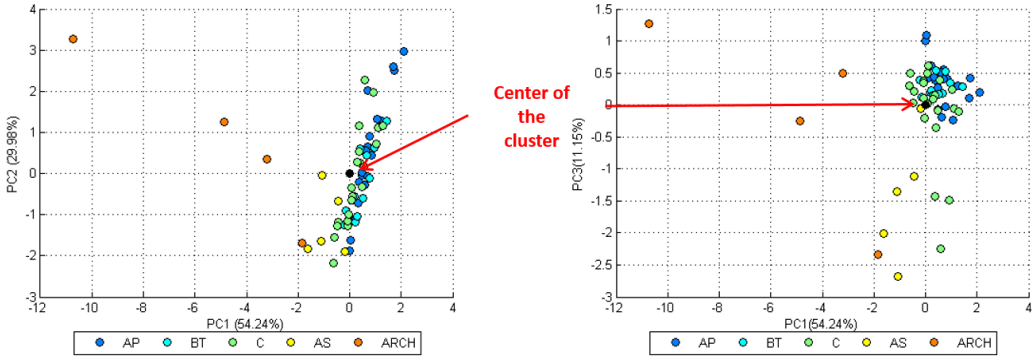


Figure 3.3.11 Score plots of PC2 versus PC1 and PC3 versus PC1 for a site in Italy. The centre of the natural soil cluster is indicated by the black marker point.

In this method the distance between the individual samples and the centre of the cluster of the PCA score plot is calculated. In theory, soil spectra from the same horizon should be clustered together in the PCA score plot because they share a similar spectral pattern. Based on this assumption, a centre point of the cluster of natural soils is calculated using the k-means clustering method (Macqueen, 1967). The distance between the centre point and every scatter point are calculated assuming archaeological materials have larger distances than natural soils. Figure 3.3.11 shows an example of this method applied to the site 1 in Italy. Here, the average distance of the archaeological materials was larger than for the natural soils (Ap, Bt and C clusters).

3.4 Other Analysis methods

3.4.1 Radiometric index

Radiometric indices are used to provide limited information on the soil properties and help to classify the soil type of unknown soils. Among various radiometric indices, this thesis will only investigate the brightness and the colour of the soil as shown in Table 3.2.

This might also be useful in detecting archaeological features since potteries and ceramic pieces tend to be reddish in colour. Also for archaeological pit formation, it is known that due to organic processes in ancient pits, the soils here tend to be darker. Also building walls of archaeological remains are mainly rocks which are brighter in colour than the surrounding natural soil. All these archaeological features depend greatly on their formation, nature, loca-

tion, usage, etc., and radiometric indices can provide a good first overview in identifying archaeological materials. These radiometric indices can be quickly applied to check whether there is an outstanding value but are limited to the visible range and the information in the near infrared region is neglected.

Table 3.2 Radiometric indices calculated for a soil property study (modified from Ray et al., 2004 adopted from Mathieu et al., 1998). R stands for the colour red (620 – 750 nm), B for colour blue (450 – 495 nm) and G for colour green (495 – 570 nm). λ_a and λ_b are the two selected wavelengths of interest.

Index	Formula	Index Property
Brightness Index, BI	$\sqrt{\frac{R^2 + G^2 + B\hat{}2}{3}}$	Average reflectance magnitude
Colour Index, CI	$\frac{\lambda_a}{\lambda_b}$	Soil colours
Redness Index, RI	$\frac{R^2}{B * G^3}$	Hematite Content

3.4.2 XRF analysis

X-ray fluorescence (XRF) spectrometer is an instrument used for relatively non-destructive chemical analysis of rocks, minerals, sediments and fluids. It can analyse major and trace elements in geological materials.

The concept of XRF is simple. It illuminates the sample by an intense X-ray beam and excites the atoms in the sample. Some of the photons are scattered and some are absorbed within the sample in a manner that depends on its composition. XRF is a powerful method which is particularly well-suited for investigations that involve a bulk chemical analysis of major elements (Si, Ti, Al, Fe, Mn, Mg, Ca, Na, K, P) and a bulk chemical analysis of trace elements (Ba, Ce, Co, Cr, Cu, Ga, La, Nb, Rb, Sc, Sr, Rh, U, V, Y, Zr, Zn) in rock and sediment.

However, the XRF analysis is limited to the analysis of 1) relatively large samples, typically greater than 1 gram, 2) materials that can be prepared in powder form and can be effectively homogenised, 3) materials containing high abundances of elements for which absorption and fluorescence effects are reasonably well understood. Thus this method is widely used in archaeology (Caneva and Ferretti, 2000; Oonk et al., 2009b; Speakman and Shackley, 2013; Tuniz et al., 2013) and it is particularly attractive for the analysis of archaeological and museum artefacts (Shackley, 2011; Shugar and Mass, 2012).

In this thesis, XRF was applied with the help from Dr. H. H. from the Johannes Gutenberg University of Mainz. The XRF instrument used was the Thermo Niton XL3t 900S

GOLDD XRF-analyser provided by the group of Prof. Dr. A. V. from the Institute for Geography, Johannes Gutenberg University of Mainz. Samples were measured for 30.5 sec with the calibration mode SOIL. To protect the instrument, all samples were covered by 6 µm thick polypropylene-foil. Some samples were measured with the XRF instrument from the Römisch-Germanische Kommission des Deutschen Archäologischen Institute.

In the following two chapters (Chapter 4 and Chapter 5), analysis results are presented according to the methods explained in Chapter 3. Chapter 4 presents the spectral analysis results for the archaeological sites in Calabria, Italy. Chapter 5 presents the results for the site Sárvíz Valley in Hungary. These sites have contrasting soil compositions, rather complex in Calabria and very homogenous in the Sárvíz Valley. The archaeological remains in these sites represent different types of prehistoric remains.

In Chapter 4 (Italy), all analysis methods explained in Chapter 3 are applied indicating their advantages and disadvantages. Then in Chapter 5 (Hungary) only the best suited methods based on the Italian results are applied and presented. After that, there is a short summary of the overall results and their differences due to the environmental conditions. Based on these results, an attempt is made to develop a universal methodology to identify archaeological features from any unknown site.

Chapter 4 Calabria, Italy

4.1 Site information

4.1.1 Location and background information of the site

The study sites were chosen within the Raganello Valley in Calabria (Figure 4.1.1), which is located at the coast of the Southern Italian peninsula. The Raganello Valley is located between the Southern Apennines to the north and the fold-thrust belt of the Calabrian Arc to the south. The catchment area of the valley is characterised by intensive tectonic activity with a great influence of the Mediterranean climate (strong seasonal contrast with a dry, warm summer and a cool, wet winter) which makes the soil and geology very complex with steep mountain slopes and a high proportion of limestone and calcareous rocks as parent material. This high complexity of the topography of the region makes it difficult to take images of the site from low-flying aircrafts (soil marks are mostly noticeable through airborne images, refer to Chapter 2). This site is a good contrast to the Sárvíz valley in Hungary, where the geology is fairly homogenous and flat landscape.

Along the valley, two specific archaeological sites (sites 1 and 2) were chosen for detailed studies of soil properties and related reflectance spectra. The two sites were chosen out of archaeological interest and had clear evidences of buried archaeological remains, which were

confirmed by previous excavations and geophysical analyses, mainly using magnetic methods.

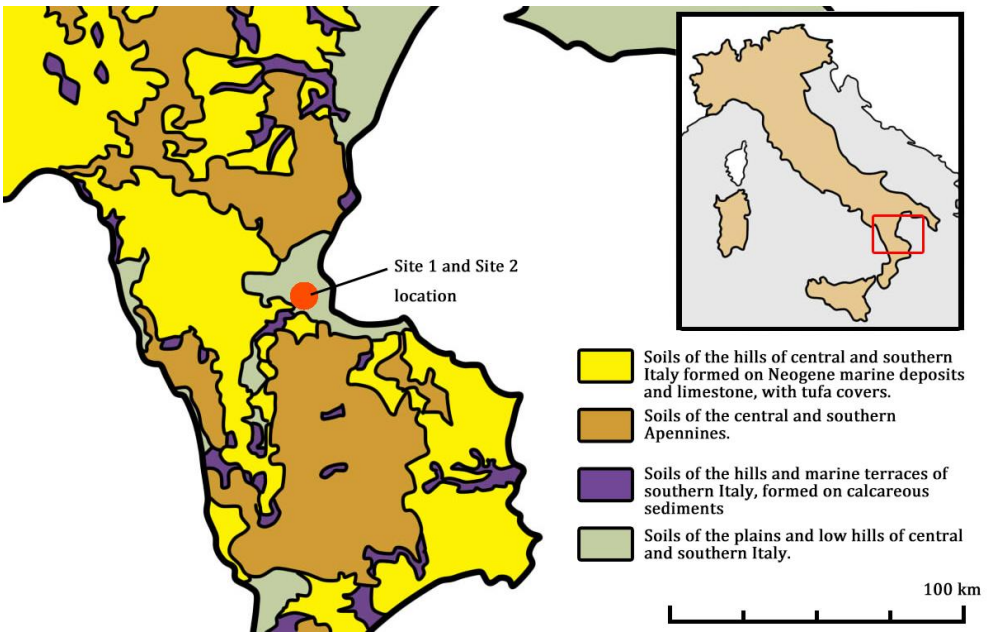


Figure. 4.1.1 Map of the study region, Calabria, Italy (modified from Constantini et al., 2013).

Figure 4.1.2 Images of the ground conditions of the sites in Italy. The soil is subject to agricultural land use but is covered with large stones and dry vegetation making it difficult to observe the bare soil surface.

The field work took place from late summer to autumn 2012 (mid of October to early November), just after the normal ploughing period of the region, in order to get clear soil mark features. Unfortunately, due to the extension of the long hot summer in the year 2012, the ploughing had not started until the research team arrived at the site. This made the ground surface extremely stony and hard. Therefore, the field work took place during the weather changing period between the summer and the winter season. This means that within the field work period there were still warm and sunny days but also dark, cold and rainy days. This influenced both the soil properties and the spectral data. Also, during the field work, there was a short period of heavy rainfall for about a week, which also delayed and influenced the work process. The contrast between the soil spectra taken before the rain (dry and crusted soil surface) is systematically different from the spectra taken after the rain where the soil colour is

darker due to moisture and transportation of soil materials down the slope during the rain. Therefore, the measurements were retaken for sites measured before the rain.

Overall, the site represents a typical Mediterranean environment with dry plants and dry surface soil. The field was covered with dried vegetation and soil in mixture with small and large stones as shown in Figure 4.1.2. The field work took place together with an archaeological team from Groningen University, a remote sensing team from Leuven University and VITO (Flemish Institute for Technological Research).

4.1.2 Soil profiles

Sites 1 and 2 are located on slopes and are, therefore, subjected to a strong downhill transportation of soil. Soil profile images including markers of the spectral measurements for the two sites are presented in this section.

(1) Site 1

For every pit, the soil spectra were measured at the sides of the vertical soil profiles as illustrated in Figure 4.1.3. In this figure, profiles 1 to 5 are vertical soil profiles of so called 'natural soil' pits. These are soil pits which were subjected to little or no ancient anthropogenic processes and for which no archaeological materials were buried beneath the surface. In contrast, an archaeological profile (top left in Figure 4.1.3) is a profile of a pit which contains an archaeological stratum (AS) in the soil horizon.

Different soil horizons (Ap, Bt and C) and spectral measurement points are illustrated in Figure 4.1.3 Ap stand for horizons of ploughed top soil. Bt is a transformation horizon, in other words, subsoil, whereas C is the parent soil. Pits were dug until the parent soil (C) was reached, which is usually around 1 m depth. C1 and C2 indicate the same parent soil but with different textures. AS is not a terminology used in soil science, but an abbreviation used in this thesis denoting archaeological soils. These soils are within the archaeological stratum but do not show clear characteristic archaeological materials (such as potteries or burned materials). Nevertheless, they display different soil colours in comparison to the surrounding natural soils. More detailed descriptions of different soil horizons are explained in the method chapter (Chapter 3.1.1).

The archaeological profile (top left in Figure 4.1.3) is located near a terrace edge. The Ap horizon in this profile is about 30 cm thick with materials transported by slope movement rather than from the parent soil. Therefore, it is unlikely to find any archaeological material or features in this Ap soil. Below the Ap horizon is a thick (around 20 cm thick) red stratum which shows clear evidence of anthropogenic activities with a thin (approximately 1 cm thick) layer of burned materials. In this stratum, no artefacts were found and there were only soils with altered soil colours due to archaeological activities.

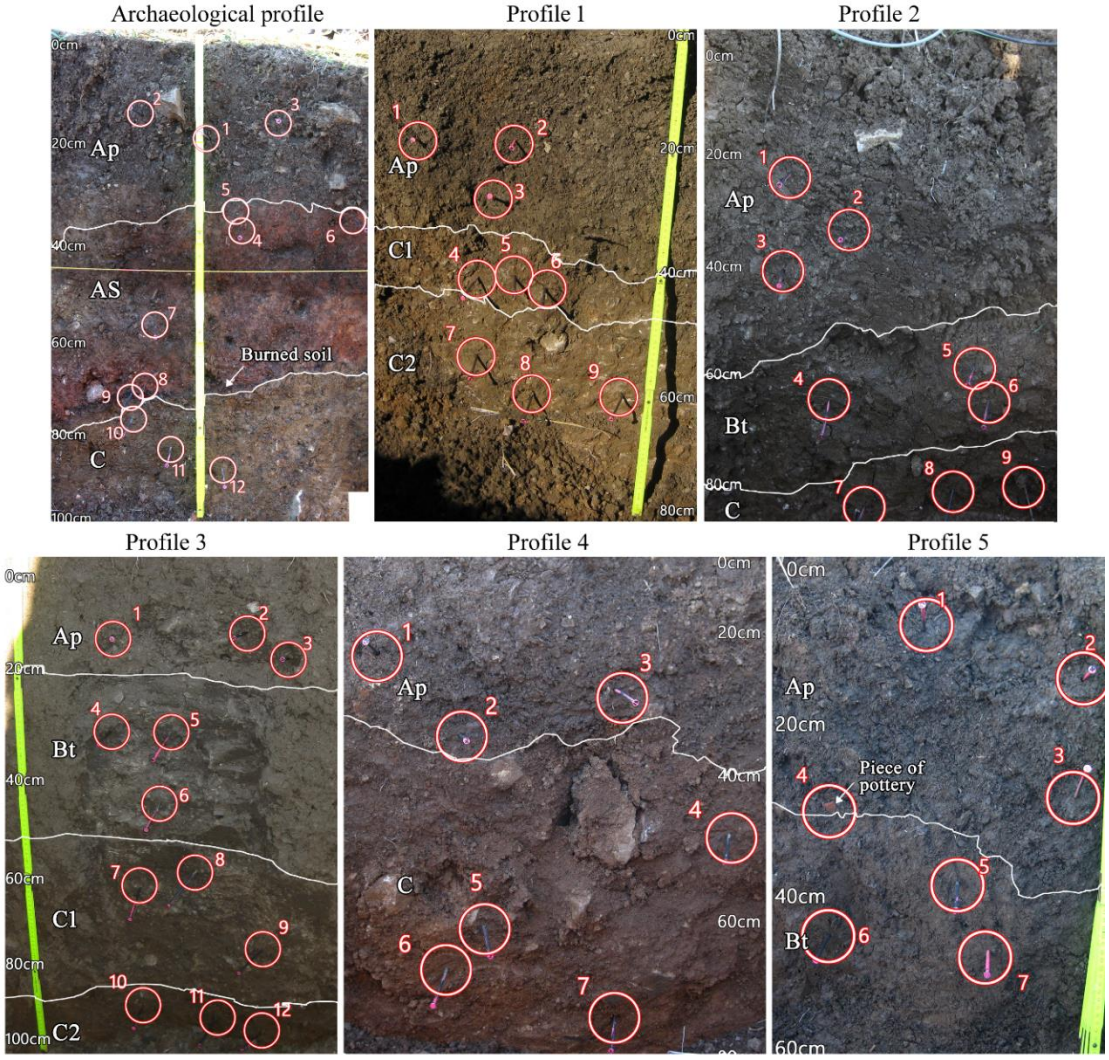


Figure 4.1.3 Images of different soil profiles taken in the pits of site 1. The term 'Archaeological profile' indicates the pit profile where the archaeological stratum was found. The profiles 1 to 5 represent the natural soil pits surrounding the archaeological profile. The red circles indicate the location where spectral measurements were taken. For every measurement point, the spectral measurements were repeated at least 3-4 times. Soil horizons are separated by the white lines.

Figure 4.1.4 Images of different profiles from the pits of site 2. Profiles 1 to 5 represent natural soils surrounding the archaeological profile. The 'archaeological' profile is the profile of a pit where archaeological features were found. Here, archaeological stratum was not clearly observed in the profile. However, the archaeological materials were visible on the floor of the pit. The red circles indicate the locations where spectral measurements were taken. Different soil horizons are separated by the white lines. For every point, the spectral measurements were repeated at least 3-4 times.

Profile 1 does not have a Bt horizon. The Ap horizon is around 10 cm thick and in the C1 horizon, one could observe leaching of clay downwards with organic material and calcium carbonates in situ. The C2 horizon is about 60 cm thick and contains weathered parent material with yellow marl and aggregate state of calcium carbonate under *in situ* conditions.

Profile 2 is located downhill from the archaeological profile and is therefore subject to colluvium thickening and deposition of marl. This profile is located in a more stable region and one can observe constant input of colluvium due to the unsorted rocks within the profile. Here, below the 60 cm thick Ap horizon, the Bt horizon has a very high clay content with calcium concretions. The C horizon is weathered and, due to bioturbation, is diminishing with depth and therefore the soil colour brightens with depth. In this horizon, extensive clay leaching is observed with shrinking and expanding amounts of marl and smectites.

Profile 3 is located between the transition zones of two different parent materials and therefore one can observe both colluvium and the marine terrace formation. Below the 30 cm thick Ap horizon is the Bt horizon, with concretions formed around the stones. Below the colluvium horizon (Ap and Bt horizon), is the marine terrace where the horizon can be divided into the C1 and C2 horizons as indicated by the different textures. Profile 3 is located on a fairly gentle slope where the erosion decreased and the amount of deposition increased.

Profile 4 is divided into the Ap and C horizons. Here the Bt horizon does not exist and the soil horizons are thinner due to the lack of the colluvium process.

Profile 5 is located near the archaeological pit and, a piece of pottery was observed in the Ap horizon. Since pit 5 is located more uphill than the archaeological pit, the pottery is not naturally transported from the archaeological remain of this site. However, since it is observed in the Ap horizon, it could have been transported during the ploughing activity, but the origin of the pottery is not known.

(2) Site 2

Figure 4.1.4 shows profile images of site 2. Profiles 1 to 5 are obtained from the pits around the buried archaeological remain and assumed to have no archaeological influence (in other words, the pits only contain 'natural' soils). Originally, 7 natural soil pits were made at this site. However, due to an instrumental problem, which distorted the spectra taken and could not be corrected, the spectra from only 5 pits will be used for analysis (Chapter 3.1.1).

The image of the archaeological profiles (Figure 4.1.4) is shown with the Ap horizon removed and therefore, the 0 cm depth is not the surface of the ground but rather a starting level (around 20 cm deep from the surface) of another horizon. However, in this AS profile, no change in soil colour indicating the influence of archaeological activities was observed. In this site, archaeological materials showed up around 60 cm depth (floor of the archaeological pit, see Figure 4.1.4), and therefore the AS and ARCH spectral measurements were only taken at the floor of the pit. Instead of digging through these archaeological materials, the archaeologists decided to preserve the remains and only take records on the surface. Large ceramic pieces (around 20 cm in width) and burned materials were found indicating that it was a kitchen formation. The white coloured feature in the top-left side of the figure of the floor of

the pit might be part of the parent geology or a wall structure of this buried remain.

In this site, the Bt horizons were not observed in most of the soil profiles. The soil profiles mostly contain a parent soil and an Ap horizon above the parent soil due to constant ploughing (remember that the site is under agricultural land use). The lack of the Bt horizons may be due to the fact that the site is located on a hill side, and most of the soils are transported downhill. One thing to notice is that the ploughing seemed to be done with the same machine over a long period since the Ap horizon had a constant thickness of around 25 cm.

4.1.3 Archaeological material

Archaeological materials buried at sites 1 and 2 are fairly identical in both the time scale of its origin and in the type of material. Both sites contain remains from prehistoric settlements with clear traces of kitchen features (soils subjected to fire activities). As shown in Figure 4.1.3 and Figure 4.1.4, both sites contain archaeological artefacts such as ceramics (parts of pottery, mostly orange in colour) and black burned material. At site 1, the soil profile showed a clear horizon, a strong red coloured stratum of soil highly influenced by the ancient anthropogenic activities. At site 2, a clear AS horizon was not observed in the profile but soils with various colours which were different from the soil colour of natural soils were identified. Therefore, they are assumed to be soils highly influenced by archaeological activities and these were mainly observed at the floor of the pit. In this thesis 'ARCH' is used to represent archaeological materials which are assumed to be non-soil materials and 'AS' represents soils which are highly influenced by archaeological processes (non-natural process).

4.2 Soil analysis

4.2.1 X-ray Fluorescence (XRF)

For the X-ray fluorescence (XRF) measurement, soil samples were taken for every spectral measurement point at the site (refer to Chapter 3.1.1 and Chapter 3.4.2 for more details on how the samples were taken and treated). The results of the XRF measurements of these soils were averaged for each horizon. These measurements provide information on whether the composition of elements changes between the natural soil and the archaeological material. Also, from the comparison of the specific composition of the elements, it might be possible to observe which element is dominant for the archaeological materials in these sites.

The top figure of Figure 4.2.1 shows the dominating major elements, K, Ca, Ti and Fe, of the top soil (Ap horizon) from site 1. The composition of both types of profiles is very similar indicating that there are no specific elements that stand out in the Ap soil from the archaeological pit against the Ap soil from natural soil pits. This is reasonable because site 1 was not ploughed and, therefore, there was no transportation of buried archaeological material to the overlying natural soil. The bottom figure shows the element level for each Ap horizon of the different pits. Notice that there are no large differences observed between Ap above the archaeological remains and natural soil except for copper (Cu) where soils above archaeological remains have relatively low Cu amounts.

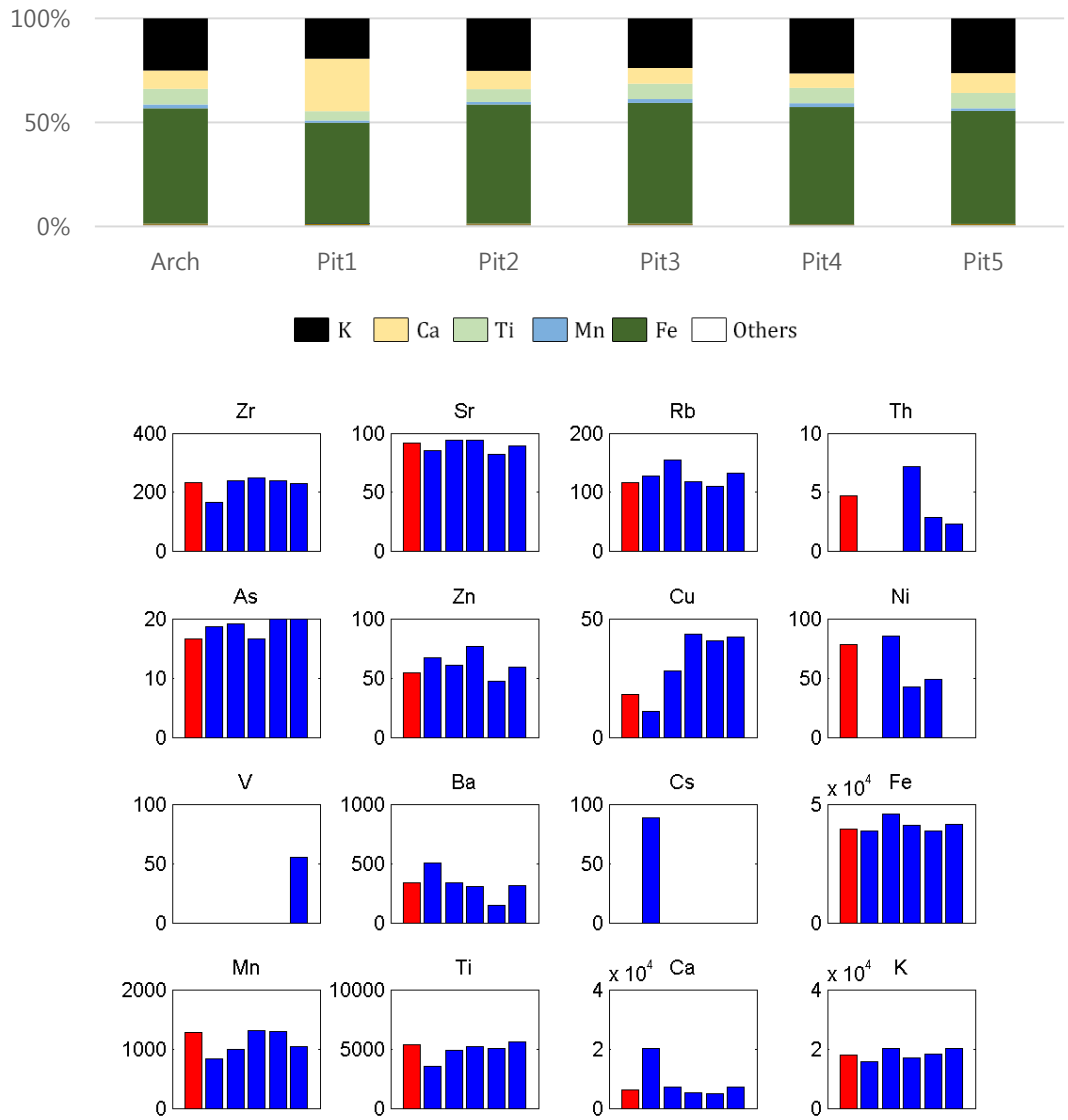


Figure 4.2.1 XRF results of the surface soils (Ap horizon) from site 1.
Top: Relative partitioning of the different elements. For the exact XRF values refer to Appendex 1. Arch stands for the archaeological pit and pit 1 to pit 5 are the natural soil pits. Dominant elements are potassium (K), calcium (Ca), titanium (Ti), manganese (Mn) and iron (Fe).
Bottom: Bar plots showing each element (Zr, Sr, Rb, Th, As, Zn, Cu, Ni, V, Ba, Cs, Fe, Mn, Ti, Ca and K) levels (in parts per million) in soil features from the Ap horizon (red: on-site, above archaeological remain; blue: off-site, surrounding natural soil) for site 1. For the exact XRF values refer to Appendex 1.

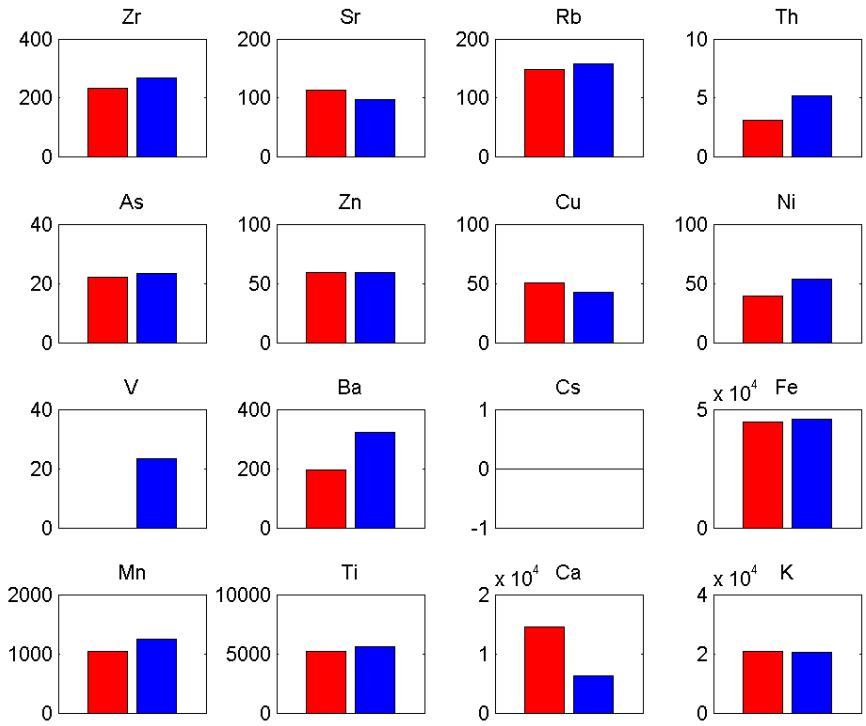


Figure 4.2.2 Bar plots showing the average element levels in ppm (parts per million) in archaeological materials (red) and natural soils (blue) for site 1. Here caesium (Cs) is not detectable. For the exact XRF values refer to Appendex 1.

Figure 4.2.2 displays the average element levels for archaeological materials and natural soils found at site 1. Archaeological materials have fairly high calcium (Ca) and low barium (Ba) amounts compared to natural soils.

Figure 4.2.3 shows the XRF results for individual pits from site 1. For the archaeological pit, surprisingly, there are no large differences between the element contents of the archaeological materials (AS and burned material) and the natural soils (C). The only noticeable features are the high Ca and Sr contents for burned materials and the low Ni and Ba contents for other archaeological materials (AS and ceramics). When each of these elements (Ca, Sr, Ni and Ba) is compared with the XRF results of the natural soils from other pits, it becomes obvious that these elements are not useful as unambiguous indicators of archaeological materials. For example, the burned materials have Ca mixing ratios of around 20000 ppm and the parent soil in this pit has a Ca mixing ratio of 7000 ppm where previously, it was noticed to be characteristic of burned materials.

However, in pit 1, the Ap and parent soil (C) have Ca values of 20000 ppm and 10000 ppm, respectively. Therefore, a high Ca value is not a characteristic element of burned materials (although in other pits the Ca value is only around 5000 ppm). Similarly, the Ni content in the burned materials is around 40 ppm, which is lower than in parent soil (96 ppm). However, natural soils from other pits have Ni values around 30 to 40 ppm (Pit 2 and Pit 4) and sometimes even below the level of detection limit (Pit1). Therefore, a low Ni content is also not a characteristic feature of archaeological materials.

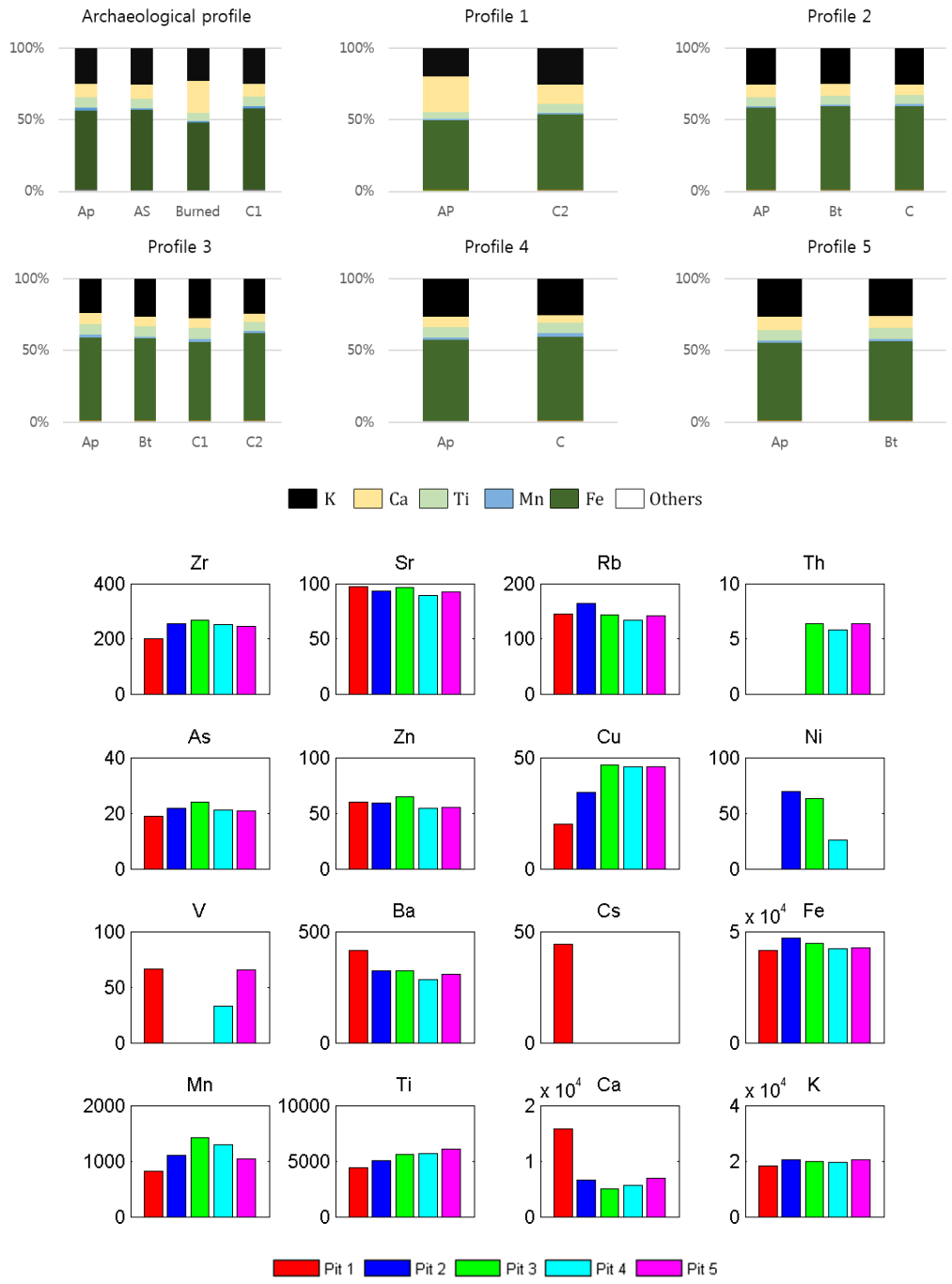


Figure 4.2.3 Top: XRF result of relative partitioning of the different elements for the different horizons for site 1. Each horizon value represented is an averaged XRF value of all soil sample measurements within that horizon. Dominant elements are potassium (K), calcium (Ca), titanium (Ti), manganese (Mn) and iron (Fe). Profiles 1 to 5 are the soil profiles of the natural soil pits.
Bottom: Bar plots showing the average element levels in ppm for individual pits at site 1. The archaeological pit is not included. For the exact XRF values refer to Appendex 1.

Some elements such as Sr and Ba show outstanding values for the archaeological materials. The Sr values for the burned materials of AS are over 100 ppm, while the natural soils (in all pits) are below 100 ppm. Similarly, the Ba content for archaeological materials is below 300 ppm, while nearly all the natural soils have Ba contents over 300 ppm (except for the Ap soil in pit 4). There are slight exceptions, but one could argue that archaeological materials tend to have Sr values above 100 ppm and Ba values below 300 ppm.

Note that in Figure 4.1.3, different soil horizons that can be clearly distinguished with the naked eyes. However, the XRF results show that there is no large difference in the element concentrations between different horizons. One interesting thing to notice is that the amount of Rb, Cu and Fe increases with depth. For example, the Fe content in Pit 2 increases from 45000 ppm in the Ap horizon to 46000 ppm in the C horizon.

The XRF results for site 1 show that it is difficult to distinguish archaeological materials from natural soils their element contents. Archaeological materials and natural soils contain very similar elemental compositions and the difference of the element contents fall within the error range of each other. Because of the ineffectiveness in using the XRF to distinguish natural and archaeological soils, for site 2, the XRF measurements were only made for a few selected soil samples.

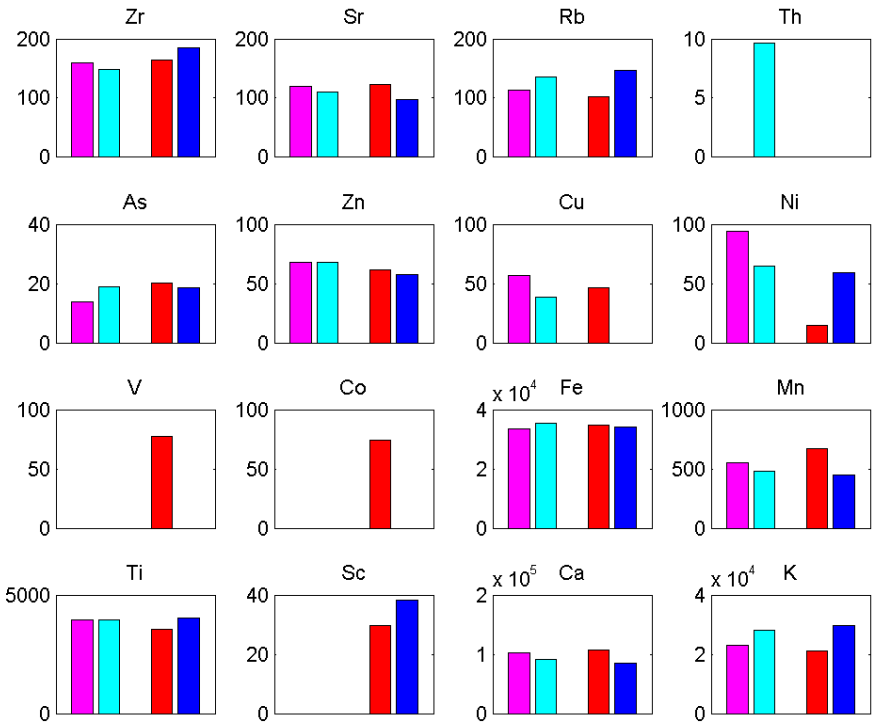


Figure 4.2.4 Bar plots showing the element levels in soil features from the Ap horizon above archaeological remains (magenta), Ap of natural soil (cyan), averaged archaeological remains (red) and average natural soils (blue) for site 2. Notice that V and Co are only observed in archaeological materials. For the exact XRF values refer to Appendex 1.

For site 2 (Figure 4.2.4), only 8 soils were used for the XRF calculation. These are the Ap soils from the natural soil pit (pit 1) and the archaeological pit, expecting to observe any dif-

ference between them (this site was ploughed). However, no significant difference between the two Ap soils was found. Other samples are one parent soil (C horizon) and various ARCH and AS materials. One interesting thing to notice is that the AS showed some Co content of around 350 ppm. This element was never observed in any other soils measurements (even at site 1) and perhaps can be an indicator to find such similar archaeological strata. Also, at site 1, a difference between the barium (Ba) content was observed for archaeological materials and natural soils. However, at site 2, the amounts of Ba were below the detection limit. Beside these features, no significant difference was observed between the archaeological materials and the natural soils.

In some publications, XRF was shown to be a powerful tool in identifying archaeological materials such as obsidians (Cecil et al., 2007; Frahm, 2013; Moholy-Nagy et al., 2013; Parnell et al., 2002), potteries (Hunt and Speakman, 2015; Neff, 1992), skulls (Carter, 2009) and for detecting pigments in paint layers (Mantler and Schreiner, 2000). However, due to certain fundamental characteristics of the XRF technique, it is not suitable for some projects which would seem at first sight to present no problems (Hall, 1960).

If a certain element is above or below the average background level, it can be an indicator of certain types of human activities. However, in this thesis, the XRF results did not show unique or outstanding elemental amounts in the archaeological materials.

4.2.2 Radiometric indices

Radiometric indices give a first glance of the spectral colour of the soil samples of interest (Ray et al., 2004). There are various radiometric indices, but this thesis focuses on the redness index (RI) and brightness index (BI). The redness index represents the haematite content in the soil. Torrent and Barron (2003) showed a nice linear relationship between the haematite content and the redness index for Mediterranean soils. A large RI value indicates a large haematite content and, therefore, can give a reddish colour to the soil (but not always). BI is sensitive to the brightness of the soils, which is highly correlated with humidity and the presence of salts (Escadafal, 1989). For the definition of the radiometric indices and wavelengths selected for the red, green and blue bands, see Chapter 3.4.1.

Figure 4.2.5 shows the results of the radiometric indices (redness index and brightness index) at site 1. One obvious result for site 1 is the high redness index for the archaeological soils (AS) compared to the natural soils (Ap, Bt and C). Here it is interesting to note that for the archaeological pit, a clear red archaeological stratum was already identified by eye (see Figure 4.1.3). However, the main archaeological artefacts (ARCH) found in this site were burned materials showing dark blackish colour for which no clear correlation with the RI value is found. The high RI values for AS indicate that AS contains a larger haematite content than natural soils or archaeological artefacts. In addition, since the iron oxide content is also linearly related to the RI, the AS might also contain a large Fe content. However, this expectation is not confirmed by the XRF results. The AS has a Fe mixing ratio of 47000 ppm for the burned materials and the parent soil (C) a mixing ratio of 43000 ppm and 45000 ppm are found, respectively.

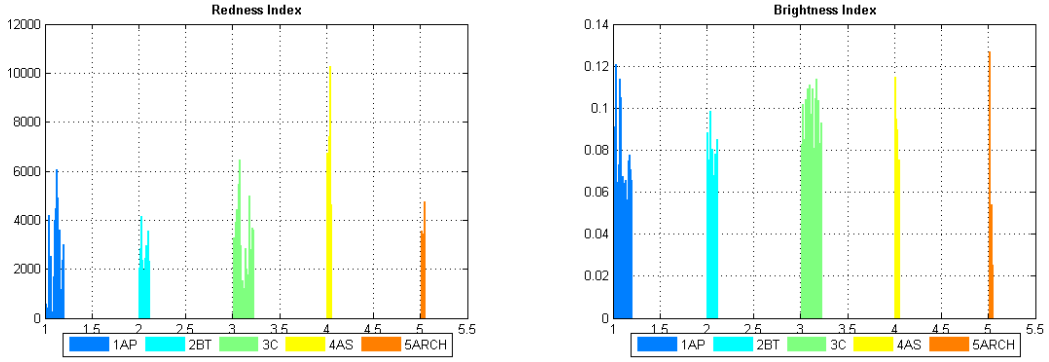


Figure 4.2.5 Radiometric indices (left: redness index, right: brightness index) for all spectral measurements gathered from site 1. Each bar represents one spectral measurement, and different colours represent soils from different horizons. Blue: Ap, cyan: Bt, green: C, yellow: AS, orange: ARCH.

The brightness index (BI) showed that soil colour brightens with depth (right sides in Figure 4.2.5 and 4.2.6), with especially low BI values for the Ap soil and high BI values for the parent soil (C). This confirms the findings from Chapter 4.1.2 that the C horizon is weathered and due to bioturbation diminishing with depth, which leads to increasing brightens with depth.

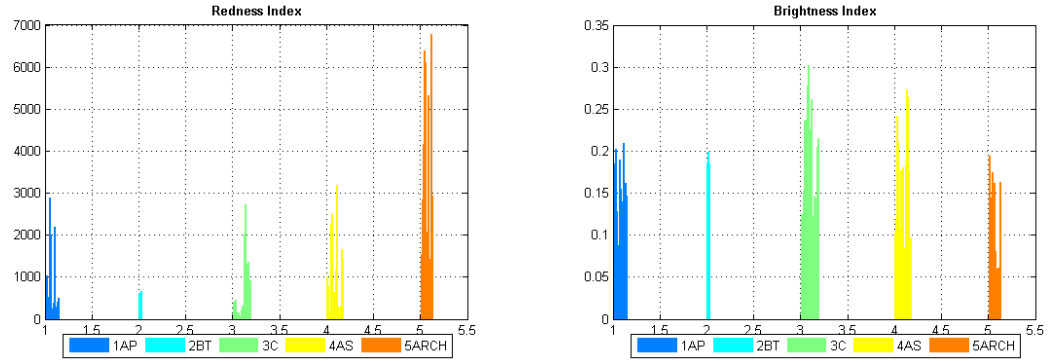


Figure 4.2.6 Radiometric indices (left: redness index, right: brightness index) for spectral measurements gathered from site 2 (Italy). Each bar represents one spectral measurement and different colours represent soils from different horizons. Blue: Ap, cyan: Bt, green: C, yellow: AS, orange: ARCH.

For site 2, the redness index of ARCH is extremely high compared to the natural (Ap, Bt and C) and archaeological soils (AS). This is reasonable since most of the measured archaeological materials were orange-coloured ceramics and the AS soils were not as reddish as the AS at site 1 (refer to Figure 4.1.3 and 4.1.4 for comparison of the AS colour in the sites 1 and 2).

The overall RI values at site 2 are much lower than those at site 1, which have a high RI value for the AS horizon ranging from 7000 to 10000, while for site 2 the highest RI value for ARCH is less than 7000. Also the remaining soils (which did not appear to be reddish for the naked eye) at site 1 have RI values around 4000, while at site 2 they vary around 2000. This

indicates that site 1 has a higher haematite content than site 2 and the archaeological stratum (AS) at site 1 has a stronger red colour than potteries.

The brightness index of the soils at site 2 is fairly constant. The parent soils (C horizon) are brighter than the soils from the Ap horizons, which can be again explained by diminishing due to bioturbation. Also this may be related to the increasing humidity of the soil with depth. Compared to site 1, the BI values at site 2 are around two times brighter. One interesting thing to note is that site 1 was excavated before the heavy rain events and site 2 was excavated after these events. High humidity is expected to lower the BI value (because it darkens the soil), but the result showed high BI for site 2 although it was moister.

Overall, the RI and BI give a first impression of the soils and in some cases archaeological materials could be identified by their strong red colour. However, not all archaeological materials are identified (for example, burned materials) and the result can strongly vary even within the same site. Therefore, based on these indices universal application classification of archaeological materials is not possible.

4.2.3 Additional colour indices

In addition to the established radiometric indices, colour indices (CI) were calculated to get a quick look whether archaeological materials can be separated from natural soils (see Chapter 3.4.1). According to the colour index formula in Table 3.2, two wavelengths (λ_a and λ_b) can be selected for the CI calculation. Several λ values were chosen for sites 1 and 2. Various λ combinations were tested, but only selected colour index results will be presented in this chapter because of the vast amount of combinations (note that the ASD spectrometer measures the 400 – 2400 nm range with 1 nm interval). The λ values presented in this chapter are the wavelengths which showed 'fairly' clear difference between archaeological and natural soils. Especially, the 550 nm region, which represents the redness of soil, is often used since the red colour was the major colour signature of the archaeological materials at the Italian sites (refer to the Appendix 2 for other various λ combinations).

First, the λ_a and λ_b values were chosen from the visible range (400 nm and 700 nm). The corresponding results for site 1 are illustrated in Figure 4.2.7.A. For most of the cases when λ values were chosen from the visible region, the CI values of the archaeological materials did not stand out with respect to natural soils. However, in the previous radiometric index section (Chapter 4.2.2), it was shown that a high haematite content in the AS horizon leads to a high RI because haematite has an absorption band around the 550 nm region. Even small amounts of haematite can have a strong colouring effect (Hill and Mégier, 2007). Therefore in the next step, the wavelength range 550 nm was specifically investigated.

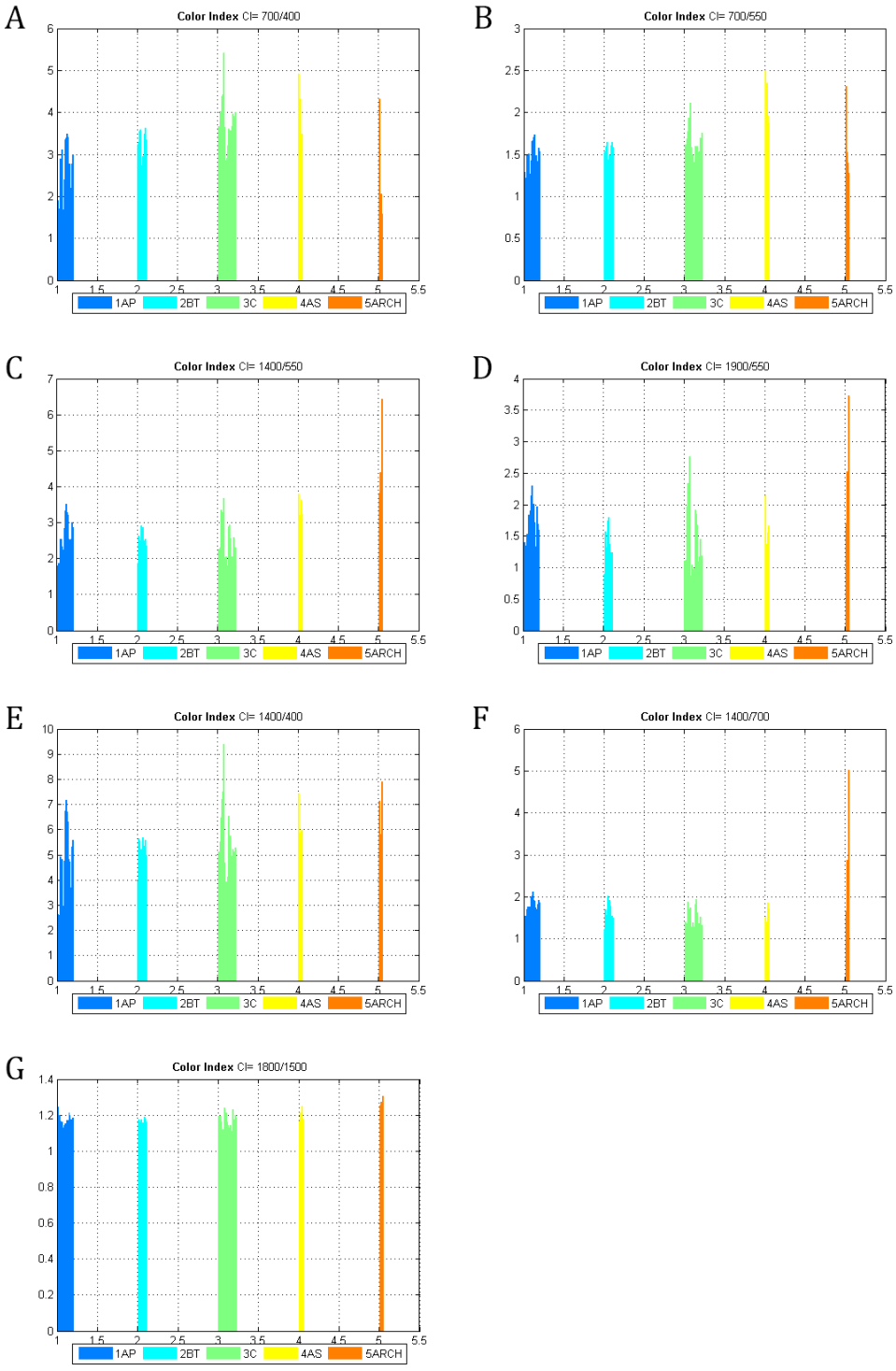


Figure. 4.2.7 Colour indices for site 1. Each bar indicates a spectral measurement of the soil and each colour represents different soil horizons. The Ap, Bt and C horizons (blue, cyan and green in colours) are natural soils, AS and ARCH (yellow and orange colours) represent archaeological materials. The wavelengths chosen for the λ values are written at the top of each plot.

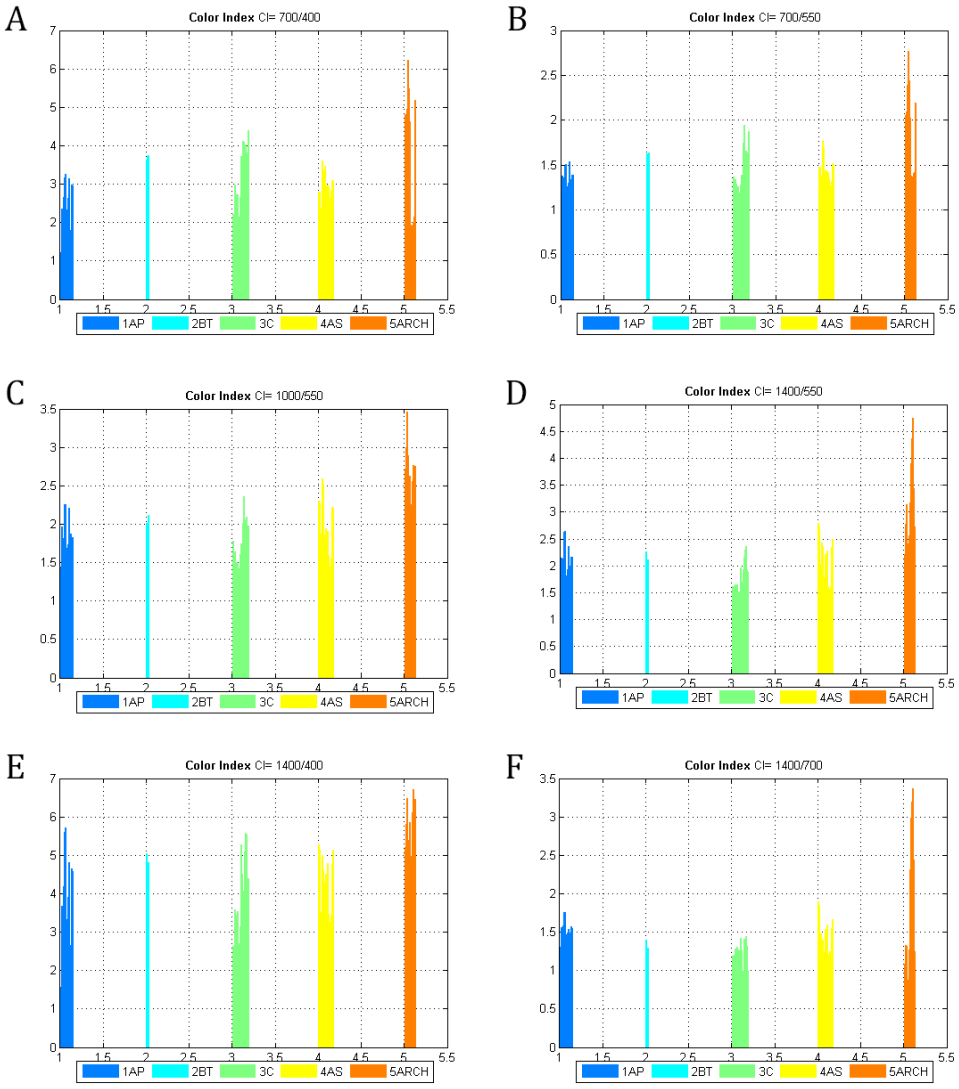


Figure. 4.2.8 Colour indices for site 2. Each bar indicates a spectral measurement of the soil and each colour represents different soil horizons. The Ap, Bt and C horizons (blue, cyan and green) are natural soils, AS and ARCH (yellow and orange colours) represent archaeological materials. The wavelengths chosen for the λ values are written at the top of each plot.

Figure 4.2.7.B shows the CI results for λ_a at 700 nm and λ_b at 550 nm (both still within the visible region). Here the results show that the CI values for AS and ARCH are slightly higher than for natural soils. When λ_a is chosen in the NIR region (in this case, 1400 and 1900 nm) and λ_b at 550nm (Figure 2.4.7. C and 2.4.7.D), CI values are always strongly enhanced for AS and ARCH compared to natural soils. This is an interesting finding, because in ARCH no outstanding amount of haematite was observed (refer to Chapter 4.2.2). Another thing to notice is that, in soil spectra, the 1400 nm and 1900 nm wavelength region is influenced by strong water absorption bands (Stoner and Baumgardner, 1981), see also Figure 3.2.4. This may indicate that the high CI values in Figures 4.2.7.C and 4.2.7.D are caused by high water content.

Figure 4.2.7.E, however, shows an example indicating that combinations of wavelengths (λ_a

and λ_b) from the NIR (1400 nm) and VIS (400 nm) region do not always produce a satisfactory result. Figure 4.2.7.F shows a CI value for λ_a at 1400 nm and λ_b at 700 nm. Here the ARCH values are outstanding compared to the soils (note that AS is also an archaeological soil and ARCH are archaeological artefacts). This is consistent for any λ_a value in the NIR region with λ_b of 700 nm. However, when λ_a is fixed to 700 nm and λ_b varied within the visible spectral range, these exaggeration patterns in ARCH values are not observed. This suggests that only a combination of λ_a of 700 nm and λ_b from the NIR region will enhance a certain feature in the archaeological artefacts found at site 1. Figure 4.2.7.G shows CI values when λ values were chosen only from the NIR region. Some λ from the NIR give a slightly higher value for ARCH, but this is not always the case.

Figure 4.2.8.A and 4.2.8.B show that unlike site 1, at site 2, the CI values using λ_a and λ_b from the visible range give slightly higher CI values for archaeological artefacts (ARCH). Notice that site 2 did not have a strong reddish archaeological soil (AS) colour as site 1. In addition, the archaeological artefacts found at site 2 contained a large amount of orange coloured potteries whereas at site 1 only a small piece of pottery was observed and most of the ARCH spectra represent burned materials. In Chapter 4.2.2 it was shown that at site 2, ARCH has large RI values indicating a strong reddish soil colour and a large haematite amount. Especially large CI value for ARCH is found when 700 nm and 550 nm are used for the CI calculation (Figure 4.2.8.B).

When λ_a is as 550 nm and λ_b is selected from the NIR region, for site 1 high CI values for ARCH were always observed. This pattern is also seen for site 2 (Figures 4.2.8.C and 4.2.8.D). However, one thing to notice is that at site 1, the haematite content (at 550 nm) in ARCH was low and at site 2, the haematite content in ARCH was high, while for both sites similar CI results were obtained. The same finding is obtained for 700 nm and various combinations of NIR wavelengths (Figure 4.2.8.E). Also, not all combinations of λ from the visible and NIR region result in high ARCH values (Figure 4.2.8.D). For both sites, high CI values for ARCH were obtained when 700 nm (for λ_a) and λ_b from the NIR region were used for CI calculation. This indicates that the archaeological materials in the Calabria region investigated in this thesis have different haematite content, but similar other properties leading to similar CI values when λ_a is either 550 nm or 700 nm and λ_b is a wavelength in the NIR region.

Not all combinations with λ_a from the visible range and λ_b from NIR region lead to different CI for archaeological materials and natural soils (Figure 4.2.8.E). However, certain wavelengths in the visible (VIS) range, especially λ_a at 700 nm, together with λ_b from NIR region, can cause significantly enhanced CI values for ARCH (Figure 4.2.8.F).

Similar to site 1, for site 2, the investigation of the CI using only NIR bands gave a somehow unclear result. Some combinations showed high ARCH values indicating that the visible range is not the only wavelength region in which archaeological material can be separated from natural soils. Also the NIR region contains useful information, which should be investigated in more detail in future studies.

Overall, the colour index results indicate that in principal ARCH can be identified among natural soils by the CI method. However, it should be noted that from measurements at only two wavelengths only very limited information can be derived.

4.2.4 Laboratory and field spectra comparison

Some collected soil samples were air-dried and sieved under 2 mm and then crushed again with the rock crusher to obtain pure soil conditions in order to enable the measurement of the spectral properties of the pure soil material. The spectra of these soils were measured with the same instrument as used in the field, and these spectra will be referred to as the laboratory spectral measurement.

A comparison between both sets of spectral measurements is shown in Figure 4.2.9. Here, soil samples from 3 different soil horizons were compared: burned material (ARCH), top soil (Ap) and parent soil (C) which were all gathered from site 1.

Overall, by looking at the original spectra (plots on the left side of Figure 4.2.9), the laboratory spectral measurements have a higher intensity than the field measurements. This is a reasonable result, since the soils were dried for the laboratory measurements and less moisture content leads to a brighter soil colour. Also just by looking at the original spectra (left side of Figure 4.2.9), it seems like all absorption features are kept and except for the intensity not much has changed compared to the field measurements. However, the continuum removed spectra illustrates that there are, in fact, significant differences between the laboratory and field spectral measurements. In the spectra of the burned material, there is a strong absorption pattern difference in the VIS range (500 – 900 nm). In the Ap spectra, some absorption features around 700 nm and 1000 nm disappear (or decrease in intensity) in the laboratory measurements, which may indicate that some information is lost in this wavelength region.

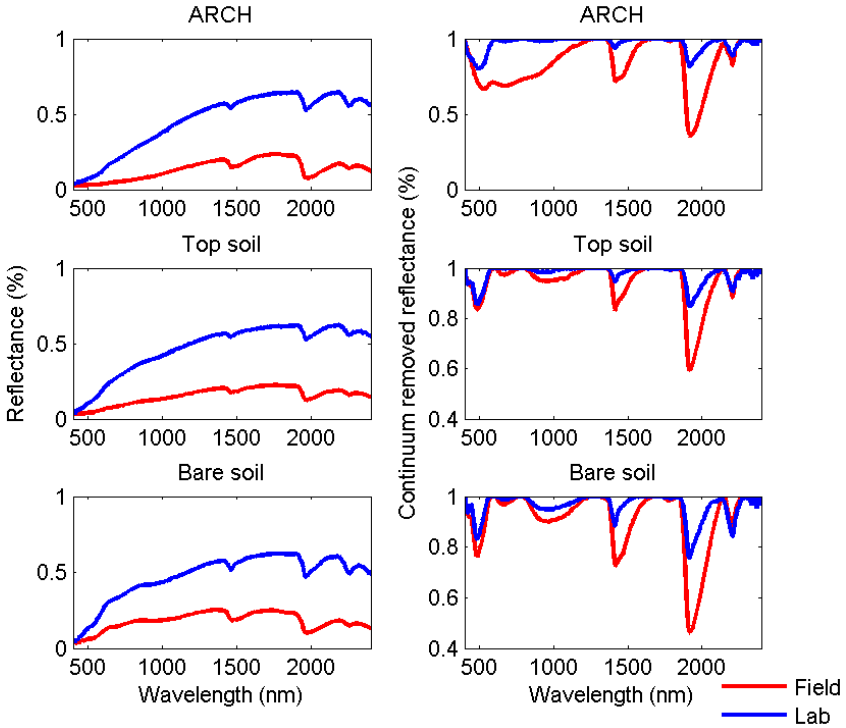


Figure 4.2.9 Comparison between the spectra measured in the laboratory (red) and in the field (blue). From the top to bottom: archaeological (burned) material, top soil (Ap) and parent soil (C). Right side: original spectra; left side: continuum removed spectra.

For the parent soil (C), the spectrum measured in the laboratory and the field are very similar but the depths of the absorption peaks decrease (this was also mostly found for the other materials). Since the soils prepared for the laboratory measurements were dried, they contain less water content and, therefore, the water absorption peaks at 1450 nm and 1950 nm are relatively small.

Overall, the laboratory measurements occasionally show the same absorption patterns as the field measurement, but also a loss of information (especially in the lower wavelength region) is observed. Reduced water absorption features in the laboratory measurement may allow to extract any archaeological spectral information in this range. However, there are also disadvantages in using laboratory spectral measurements. As shown in the results above, some spectral information is lost although at the moment it is not clear which spectral range contains most archaeological information. In addition, if data collected directly from the field are used, they will probably better fit to airborne or satellite data. Therefore, in this thesis, all spectral analysis will use spectra measured directly at the field.

4.2.5 Mixing of soil and archaeology for investigation

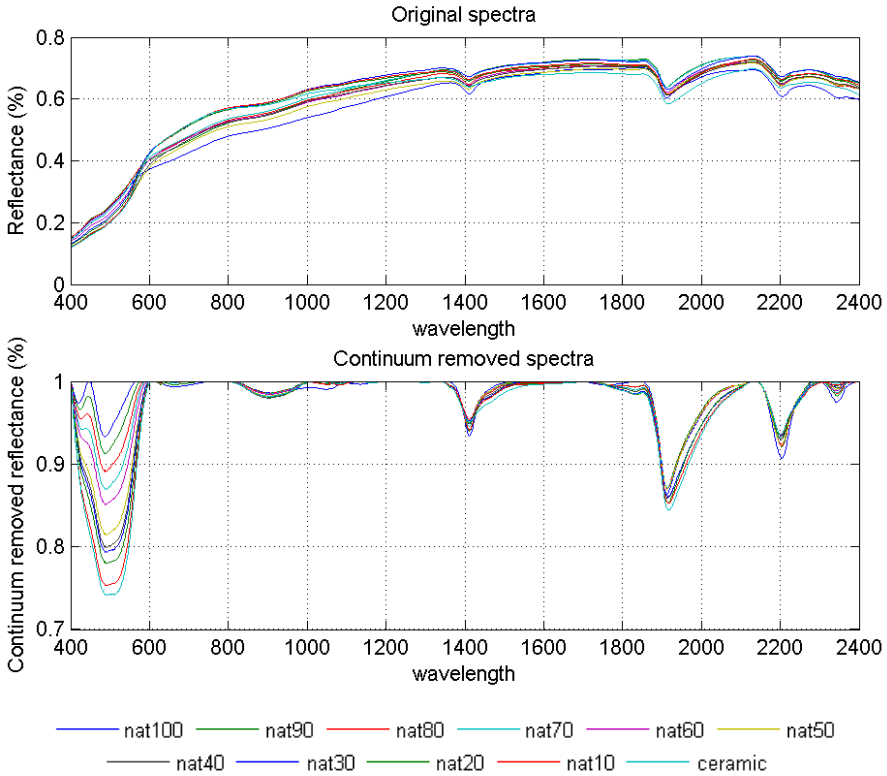


Figure 4.2.10 Original spectra (top) and continuum removed spectra (bottom) for different mixtures of natural soil and archaeological material (ceramic). The legend shows the percentage of the archaeological material mixed into the natural soil. 100% natural soil is represented in blue and 100% archaeological material (pottery) in cyan.

In this thesis, samples and spectra of soils and material of archaeological remains and natural soils were gathered from the pits which were least influenced by recent human activities. With these soils, one can mix the soils and observe how the spectra change for different fractions of archaeological materials and natural soils. From such spectral data, one can get an idea of how much archaeological soil is needed to be distributed on the surface to be detectable as soil mark.

A controlled amount of natural soil and archaeological material was mixed together to see how the spectra changes as the two different soils are mixed together. According to Buck et al., 2003 who observed the spectra of a specific area with a different amount of obsidian, around 80% of obsidian coverage is needed in order to identify spectra with obsidian features.

Figure 4.2.10 shows how the spectrum of a natural soil changes as more and more archaeological material (pottery) is mixed into the sample. The archaeological material used for this experiment is a piece of crushed ceramic. Just by looking at the spectra it is difficult to identify any pattern between the different spectra.

Here, as more archaeological material is mixed into the natural soil, the spectrum gradually becomes more and more identical to the pure ceramic spectrum. In the visible wavelength, until Nat 60% Arch 40%, the absorption band is very similar to the pure natural soil. Only from Nat 10% Arch 90% the absorption feature becomes similar to the spectrum of pure ceramic. Also there is a large change between the 70% archaeology mixture and the 60% archaeology mixture at the wavelength range of 400 to 1300 nm. The former spectrum follows the absorption pattern of the pure archaeological spectrum and the latter follows the pattern of the pure natural soil spectrum. Interestingly, this is not observed in the continuum removed spectra (Figure 4.2.10 bottom) indicating that this difference is only caused by the intensity of the spectra (brightness). The absorption bands at 1400 nm, 1900 nm and 2200 nm are due to water absorption, thus they don’t strongly change for different fractions of archaeological materials and soils.

Figure 4.2.11 shows the score plots of the first and second principal components for the spectra mixed between archaeological and natural soils. The application of the PCA is explained in Chapter 3.3.1 where it clusters spectra of similar behaviour together. PCA results are shown from Chapter 4.3 onwards, but this figure is provided to give an impression of how much archaeological coverage is needed to be detected. This figure illustrates that the 400 to 1000 nm wavelength range is most suitable to distinguish archaeological materials and natural soils (this agrees with the later results in Chapter 4.4). It also indicates that PC1 distinguishes most of the differences between the two different materials. When only the x-axis (PC1 axis) is considered (Figure 4.2.11.A and Figure 4.2.11.B), the marker points follow the degree of archaeological soil mixture.

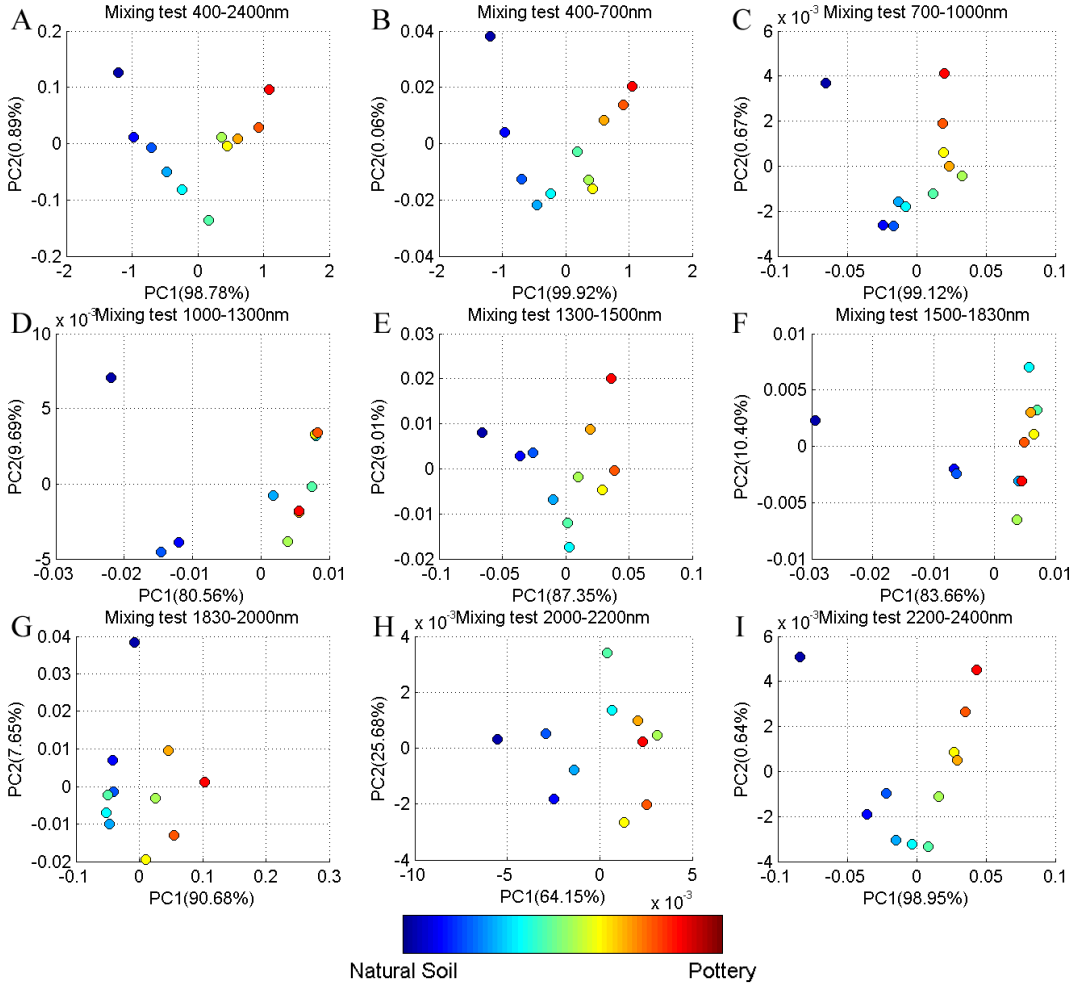


Figure 4.2.11 PCA score plots of the spectra (from Figure 4.2.10) for different mixtures of natural soil and archaeological material (ceramic). Light green indicates a mixture of exactly 50 % to 50 %. Notice that the score plots for 400 – 2400 nm, 400 – 700 nm, and 2200 – 2400 nm show a clear steady transformation of natural soil to archaeological material. However, beyond 1000 nm, the markers are randomly scattered, and therefore difficult to distinguish. This indicates that most of the spectral information determining archaeological features is found below 1000 nm.

4.3 PCA

The Principal Component Analysis (PCA) was applied to the spectra gathered in Calabria in order to separate archaeological spectral features from those of the spectra of the surrounding natural soils. Classification of the soil horizons followed the rule outlined in Chapter 3.1.1, where the different abbreviations stand for: top soil (Ap), subsoil (Bt), parent material (C), archaeological artefact (ARCH) and archaeological soil (AS). Figure 4.3.1 shows the score plot of the first two principal components for sites 1 and 2. The PCA was applied over the whole wavelength range of the spectrometer (400 – 2400 nm). Clear differences between the archaeological materials (ARCH), marked in the red circle and natural soils are found for both

sites. An interesting point to notice is that, although the archaeological soils (AS) at site 1 have a characteristic red colour (shown in yellow markers in Figure 4.1.3), the corresponding results of the PCA present the AS within the cluster of the natural soils (the cluster on the right side of Figure 4.3.1.A). This finding indicates that in this wavelength range soil colour is not the main factor that determines the principal components.

Another thing to notice in Figure 4.3.1 is that the separation between the archaeological artefacts (ARCH) and the soils (both natural and archaeological soils) is mostly dominated by the PC1 for site 1 and the PC2 for site 2. At site 1, the cluster of the natural soils is centred on 0 for the PC1 and the ARCH values vary from -5 to -10 for the PC1. This indicates that the archaeological spectral features have high contributions to the first PC while the natural soils have much smaller contributions. For site 2, the soils have PC2 values around 0 and ARCH has PC2 values from 0 to 6. This result indicates that the spectral features of the PC2 are more influenced by archaeological artefacts than by the soils and thus it is possible to separate archaeological features from natural soils. To investigate this in more detail, the spectra of the first three principal components are displayed in Figure 4.3.2.

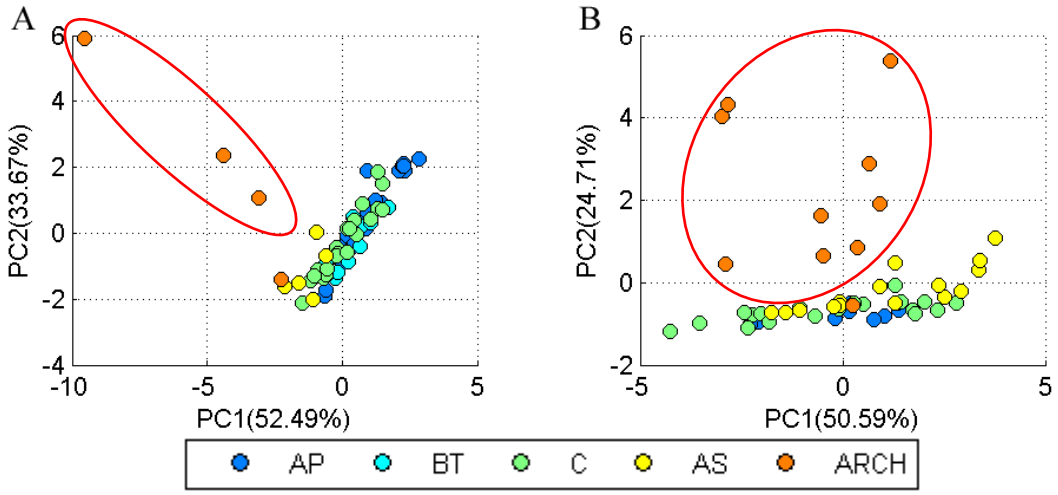


Figure 4.3.1 Results of the PCA analysis (contributions of the PC1 against those of the PC2) obtained for the wavelength range 400 – 2400 nm.
At site 1, three out of the four archaeological materials (marked in red-lined circle) are well separated from the natural soil cluster. These three spectra represent the spectra of black burned materials which were located beneath the red archaeological stratum. The orange point which is located close to the natural soil cluster represents the spectrum of a ceramic piece (pottery).
At site 2, most of the orange markers (archaeological artefacts) are separated from the natural soil cluster showing higher PC2 values. The one orange point within the natural soil cluster represents the spectrum of a ceramic piece. One assumption for these features is that the point measurement on the pottery was made on its surface where soils were not clearly removed. If the artefact is covered by thick soil, it may not be recognisable.

Figure 4.3.2 shows the spectra for the first three PCs. Here for both sites, in the NIR region, the spectral features are dominated by the water absorption bands (absorption peaks around 1400 nm and 1900 nm, refer to Figure 3.2.5). Thus, it is probably difficult to distinguish other spectral features in that spectral range. In the visible range, for site 1, the PC1 is dominated by

a large broadband absorption feature with a slight peak around 550 nm. The PC1 for site 2 also shows a similar but weaker peak at 550 nm. There is also a broad spectral signature from 600 – 1300 nm in the PC1 for site 2, which might be a characteristic archaeological feature. However, the peak at 550 nm and the broad spectral signature between 400 – 1300 nm are observed in nearly all three PC plots (Figure 4.3.2). Haematite content or iron oxides in soils can be the reason for peaks in the region from 450 – 500 nm (De Oliveira et al., 2015). The 550 nm peak cannot be defined as an archaeological spectral feature since it is observed in all soil spectra measured within this thesis.

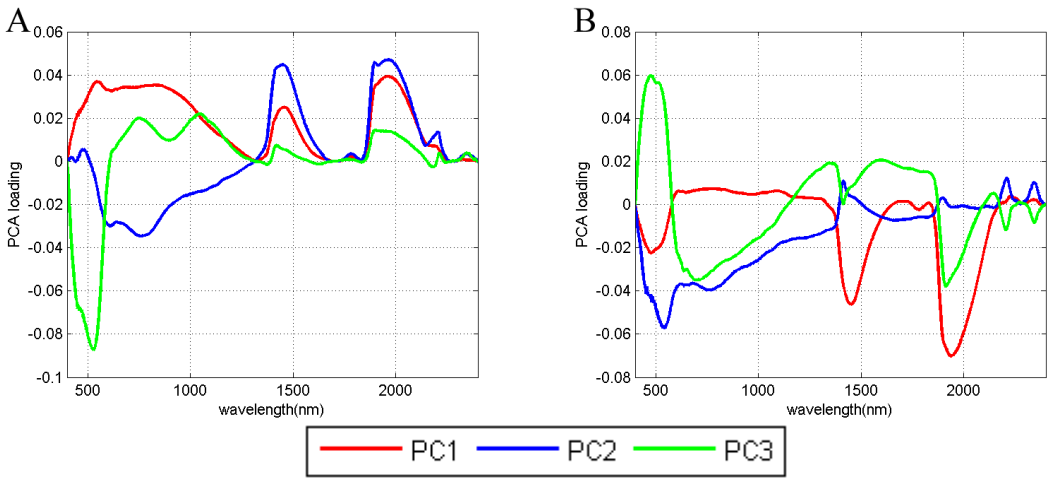


Figure 4.3.2 400 to 2400 nm spectra of the first, second and third principal component (PC1, PC2 and PC3) for sites 1 (A) and 2 (B).

The soil spectra are largely dominated by water absorption features in the NIR range. Investigating the PCA results without the NIR region might provide some useful information. Also, in the previous section (Chapter 4.1.2), each soil horizon could be distinguished by eye indicating that the PCA in the visible spectral range might give a similarly good separation. Therefore, the next step is to observe the PCA over different wavelength ranges separately.

4.4 PCA results for different wavelength ranges

Figure 4.4.1 and Figure 4.4.2 show the PCA score plots of the first two PCs for various wavelength ranges for sites 1 and 2. The wavelength ranges are first divided by visible range (400 – 700 nm) and near infrared region (700 – 2400 nm). However, the NIR region is also further divided according to the strong water absorption bands (1300 – 1500 nm, 1830 – 2000 nm and 2000 – 2200 nm) and the spectrometer boundaries at 1000 and 1830 nm (the ASD spectrometer is separated in to three spectrometers, refer to Table 3.1). Therefore the wavelengths investigated are 400 – 700 nm, 700 – 1000 nm, 1000 – 1300 nm, 1300 – 1500 nm, 1500 – 1830 nm, 1830 – 2000 nm, 2000 – 2200 nm and 2200 – 2400 nm.

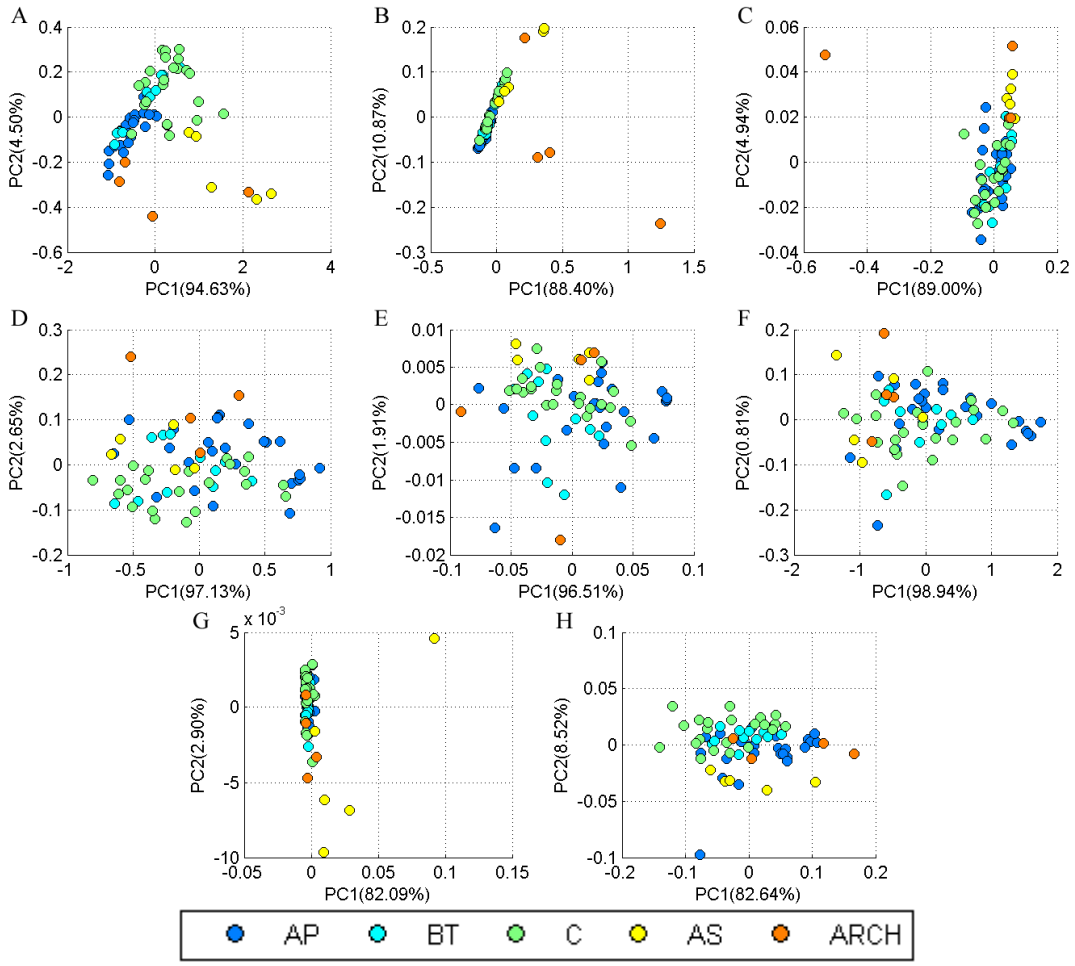


Figure 4.4.1 PC1-PC2 score plots for site 1 in Calabria for different wavelength ranges. A: 400 – 700 nm (visible range), B: 700 – 1000 nm, C: 1000 – 1300 nm, D: 1300 – 1500 nm, E: 1500 – 1830 nm, F: 1830 – 2000 nm, G: 2000 – 2200 nm and H: 2200 – 2400 nm. AP, BT and C represent different soil horizons (natural soils). AS represents spectra of archaeological soils and ARCH spectra of archaeological artefacts.

For site 1 (Figure 4.4.1), all archaeological features (including ARCH and AS) are partly separated from the natural soils (Ap, Bt and C) for wavelength ranges below 1000 nm. However, in the NIR range (beyond 1000 nm), archaeological artefacts (ARCH) are mixed together with natural soils.

Interestingly, in the 2000 – 2200 nm range, most archaeological soils (AS) were isolated from the rest of the spectra (natural soils and archaeological artefacts like pottery and burned material). This finding indicates that there might be a specified material in AS which is not observed in natural soils or archaeological artefacts in this region. However, according to the XRF results in Chapter 4.2.1, there are no outstanding minerals in AS. Thus, the reason for this separation at the 2000 – 2200 nm range is unknown. Overall, the results indicate that the archaeological materials are best separated from natural soils in the wavelength range between 400 – 1000 nm.

Figure 4.4.2 shows the results of the PCA for site 2, which is similar to the result for site 1. In

the 400 – 1000 nm spectral range, archaeological features (ARCH and AS) are well separated from the cluster of natural soils. However, beyond this range (1000 nm) it becomes difficult to observe any clear separation. For example, the 2000 – 2200 nm spectral range showed a clear division between natural soils and AS for site 1, but such a separation is not observed at site 2.

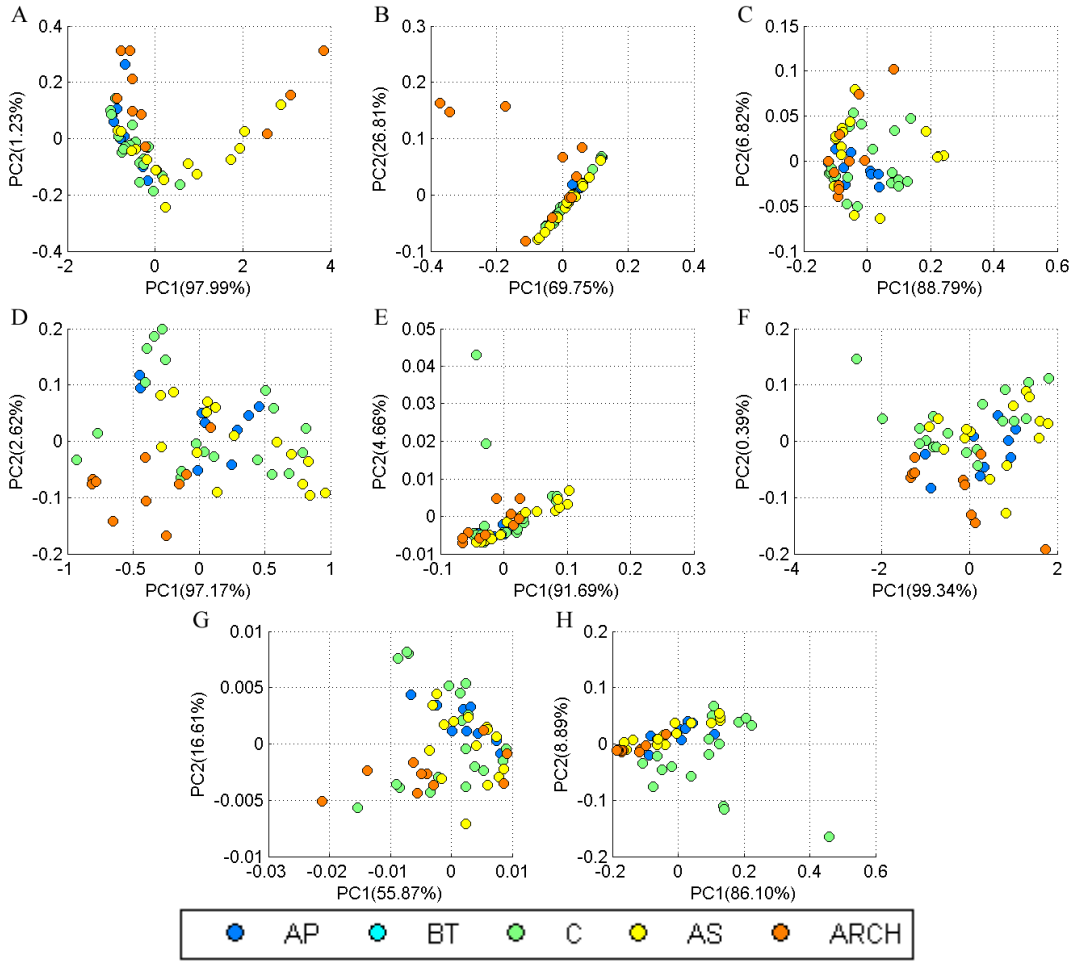


Figure 4.4.2 PC1-PC2 score plots for site 2 in Calabria for different wavelength ranges. A: 400 – 700 nm (visible range), B: 700 – 1000 nm, C: 1000 – 1300 nm, D: 1300 – 1500 nm, E: 1500 – 1830 nm, F: 1830 – 2000 nm, G: 2000 – 2200 nm and H: 2200 – 2400 nm. AP, BT and C represent different soil horizons (natural soils). AS represents spectra of archaeological soils and ARCH spectra of archaeological artefacts.

One interesting finding observed at both sites is that in the visible spectral range (400 – 700 nm), the PCA does not clearly separate the different horizons. Note that these horizons were identifiable by bare eye although the human eye can only observe three wavebands (red, green and blue). Compared to the human eye, spectrometers provide a higher spectral resolution and thus more detailed information. This indicates that the PCA is not just controlled by the colour of the soil, but there are also other factors which control the differences between the soil spectra.

Before concluding that spectral ranges beyond 1000 nm range do not provide useful archaeological information, one should note that there are strong water absorption bands which dom-

inate the spectral features in this region (refer to Figure 4.3.2 where the spectra of the first three PCs in the NIR region are dominated by the absorption features of water). Therefore, if the spectral measurement were made in a dry region with less water content in the soil, the results might provide more useful information in this wavelength range. This hypothesis is supported by the dried soil samples in Chapter 4.2.4, where laboratory soil spectra measurements showed small water absorption features in the 1400 and 1900 nm regions. However, for soil spectra gathered from Italy, the spectral range 400 – 1000 nm contains the most useful information and, therefore, from now on the PCA is applied only in the 400 – 1000 nm spectral range.

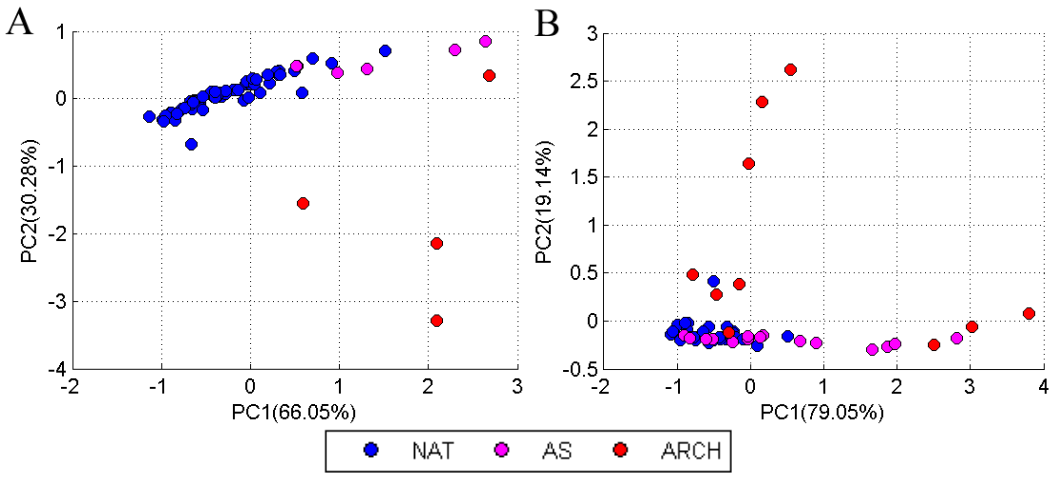


Figure.4.4.3 Score plot of the first two principal components for sites 1 (A) and 2 (B) over 400 – 1000 nm range. NAT represents the spectra of natural soils (including Ap, Bt and C), AS represents spectra of archaeological soils and ARCH of archaeological artefacts. Notice that the combination of the first and second principal components clearly separates ARCH features and parts of AS features. Most of the natural soils (NAT) are clustered together.

The PCA was repeated for the wavelength range 400 – 1000 nm in Figure 4.4.3, similarly as Figures 4.4.1 and 4.4.2, but in Figure 4.4.3 the results for the 3 soil horizons are now represented by the same (blue) colour for clear visualisation.

4.5 Results for higher order PCs

The analyses so far focused only on the first two PCs which already accounted for 90% of the total variance. In other words, all soil spectra from sites 1 and 2 can be explained by the first two principal components. Although the first two PCs clearly dominate, higher order PCs (PC3 and PC4), which account for the remaining 10%, might also contain information which could be used to separate archaeological features from natural soils. Figure 4.5.1 shows the contributions of the different PCs for sites 1 and 2. It is found that the contributions of PC3 and PC4 are close to or below 1%. However, to test whether there is still any valuable information in this 2% of variance, higher order PCs up to PC4 are investigated.

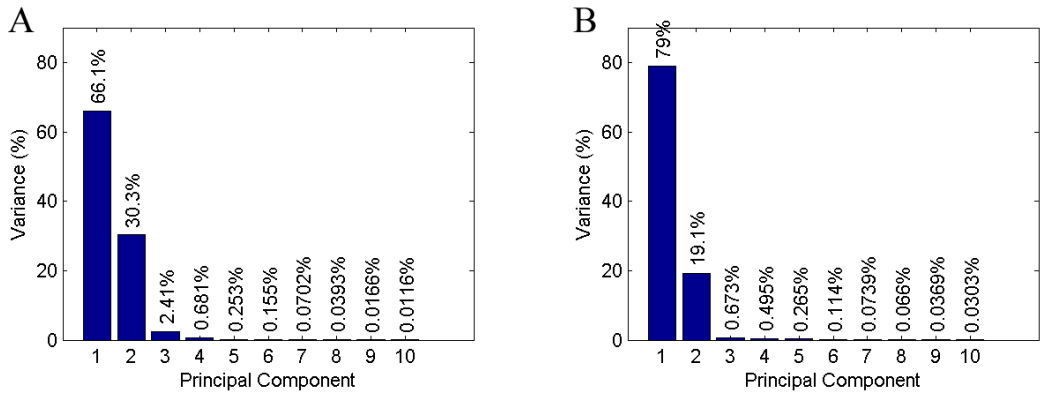


Figure 4.5.1 Contributions of the first 10 principal components for the wavelength region 400 – 1000 nm for sites 1 (A) and 2 (B).

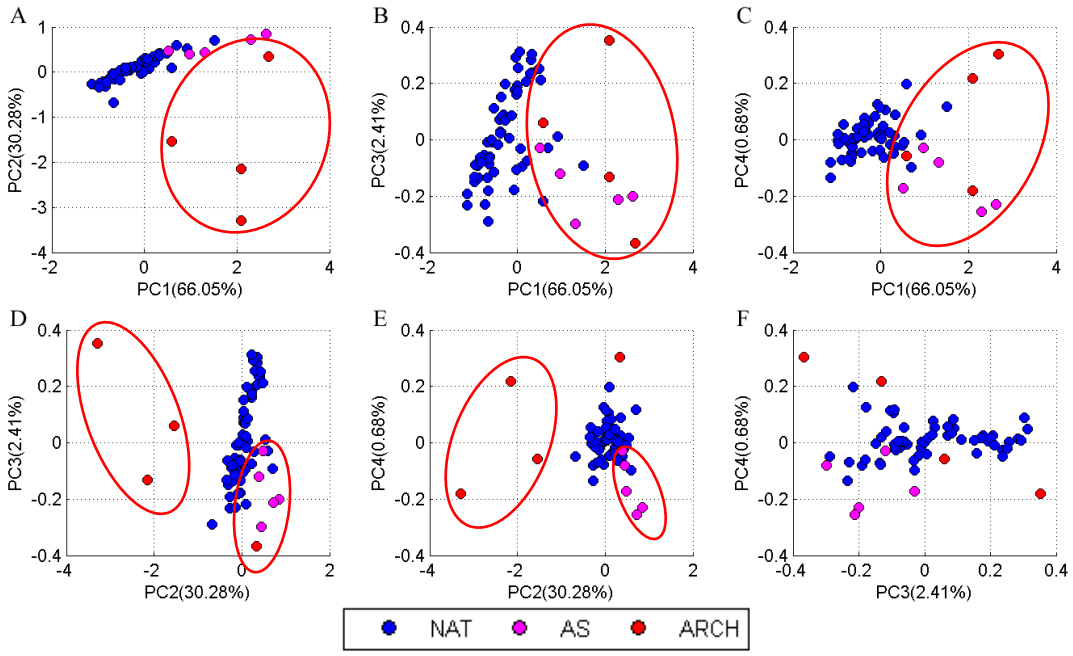


Figure 4.5.2 PCA results (score plots of PC1 to PC4) for site 1. NAT represents natural soils (including Ap, Bt and C), AS represents archaeological soils and ARCH archaeological artefacts. The red circles indicate the archaeological spectra which are clearly distinguished from the natural soil cluster.

Figure 4.5.2 illustrates results (score plots of various PC combinations from PC1 to PC4) for site 1. All score plots for PC1 and other PCs (PC2 to PC4) provide clear separations between archaeological materials and natural soils since PC1 already showed large difference between archaeological materials and natural soils (refer to Figure 4.4.3). The combinations of PC2 with the higher order PCs (PC3 or PC4) also present separated ARCH, NAT and AS results (Figures 4.5.2.D and 4.5.2.E). Especially the score plots of PC2 against PC3 or PC4 (Figure 4.5.2.E) show a separation of different types of archaeological materials where burned materials are on the left side of the natural soil cluster (negative PC2 values) and reddish archaeo-

logical soils (AS) are on the right side of the natural soil cluster (positive PC2 values). Here, the natural soils have PC2 values of approximately 0 indicating that natural soil spectra are not represented by the PC2. Finally, the PC3-PC4 score plot does not provide a clear distinction between archaeological materials and natural soils. These results indicate that the score plots of either PC1 or PC2 and higher order PCs can provide useful information for the separation of archaeological materials, archaeological soils and natural soils, which was not provided by the combination of PC1-PC2 alone. However, score plots using only higher order PCS (Figure 4.5.2.F) do not provide any additional information.

Figure 4.5.3 shows the score plot of different PCs (PC1 to PC4) for site 2. Compared to site 1 (Figure 4.5.2), the archaeological materials are less clearly separated from the natural soil cluster. ARCH features are outstanding in some plots whereas AS spectra are often clustered within the natural soil cluster when higher order PCs are used. For site 2, PC2 is representative for large parts of the spectra of archaeological artefacts (ARCH), but much less representative for soils (including both natural and archaeological soils).

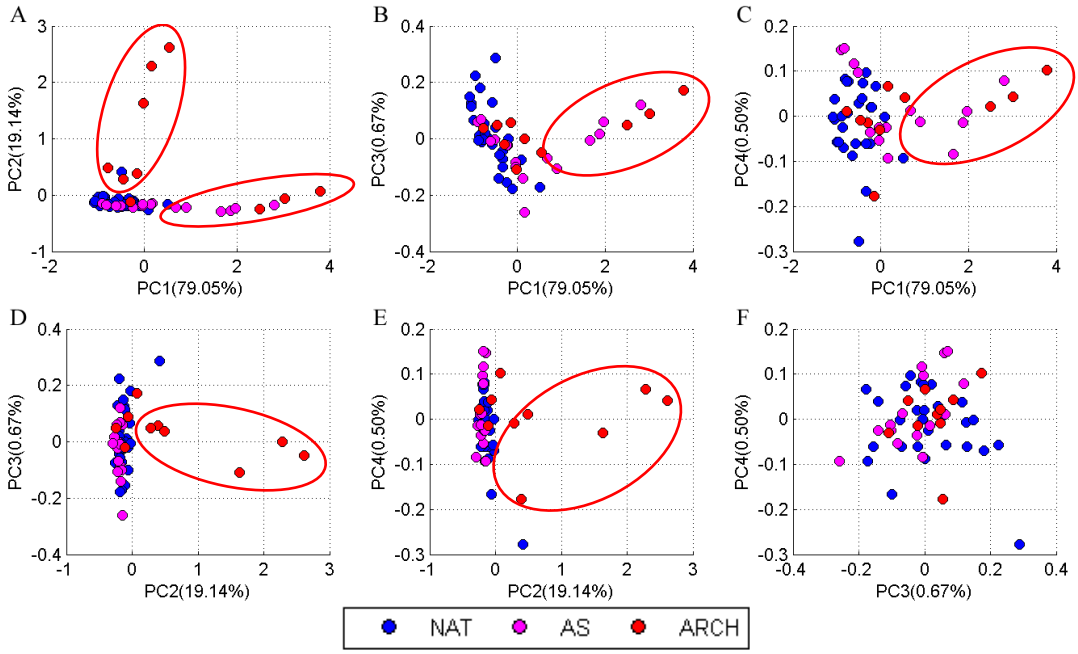


Figure 4.5.3 PCA results (score plots of PC1 to PC4) for site 2. NAT represents natural soils (including A, Bt and C), AS represents archaeological soils and ARCH archaeological artefacts. The red circles indicate the archaeological spectra which are clearly distinguished from the natural soil cluster.

Spectra of the first three principal components for sites 1 and 2 for the wavelength range 400 – 1000 nm are shown in Figure 4.5.4. As mentioned above, for both sites, the second principal component (PC2) largely represents spectral features of archaeological materials. In Figure 4.5.4, PC2 (for both sites) has a broad spectral signature in the 520 – 1000 nm range, which is probably a characteristic spectral feature for archaeological materials.

In summary, the results (score plots of higher order PCs) provided some combinations where archaeological materials are separated from natural soils. However, some archaeological spectra are still mixed with the cluster of natural soils (especially for the archaeological soils), and

this effect is more clearly seen for site 2 than for site 1.

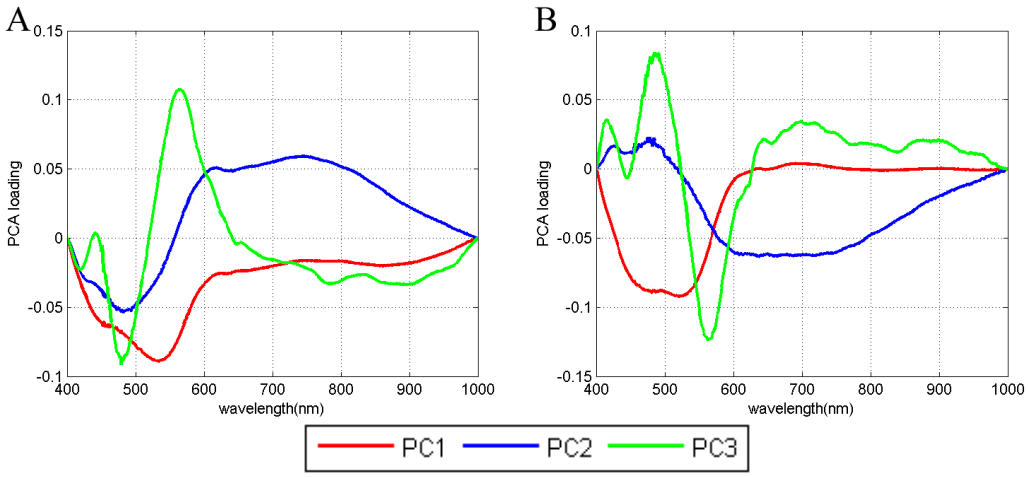


Figure 4.5.4 Spectra of the first 3 principal components for sites 1 (A) and 2 (B) for the wavelength range 400 – 1000 nm.

4.6 Base PC: Results for using PCs from a known dataset

In the previous section, the PCA of site 1 gave a clear separation between the archaeological materials (both AS and ARCH) and the natural soils at the wavelength range 400 – 1000 nm (Figure 4.5.2.A). Also for site 2 most of the archaeological materials were distinguishable, but still a number of spectra of archaeological features remained at the centre of the natural soil cluster (Figure 4.5.3.B).

To improve the separation of the spectra from site 2, a set of principal components (PC) of a dataset where archaeological materials and natural soils were well separated is used. A detailed description of this method can be found in Chapter 3.3.4.1. To avoid confusion, principal component values from a 'well-separated' site will be referred as 'base PC'.

Here, the PC results from site 1 are used as the base PC since a clear separation between archaeological materials and natural soils was observed for this site. Figure 4.6.1 shows the application of the base PCs to site 2. It is found that the PC1-PC2 score plots (Figure 4.6.1.A and Figure 4.6.1.B) are almost unchanged (besides a slight clockwise rotation). However, for the PC2-PC3 score plot, the application of the base PC improves the result (Figure 4.6.1.C and 4.6.1.D). In the original PC2-PC3 score plot (Figure 4.6.1.C), only the burned archaeological spectra were separated and the remaining archaeological artefacts, such as potteries, were mixed into the natural soil cluster. When the base PCs are used (Figure 4.6.1.D), archaeological soils are also detached from the natural soil cluster. In addition, the application of the base PC allowed a division even within the archaeological artefacts. The burned materials are located in the left side of the natural soil cluster while AS and potteries are located on the right side of the cluster. The natural soils have PC2 values around 0, while the archaeological features have PC2 values of ±2.

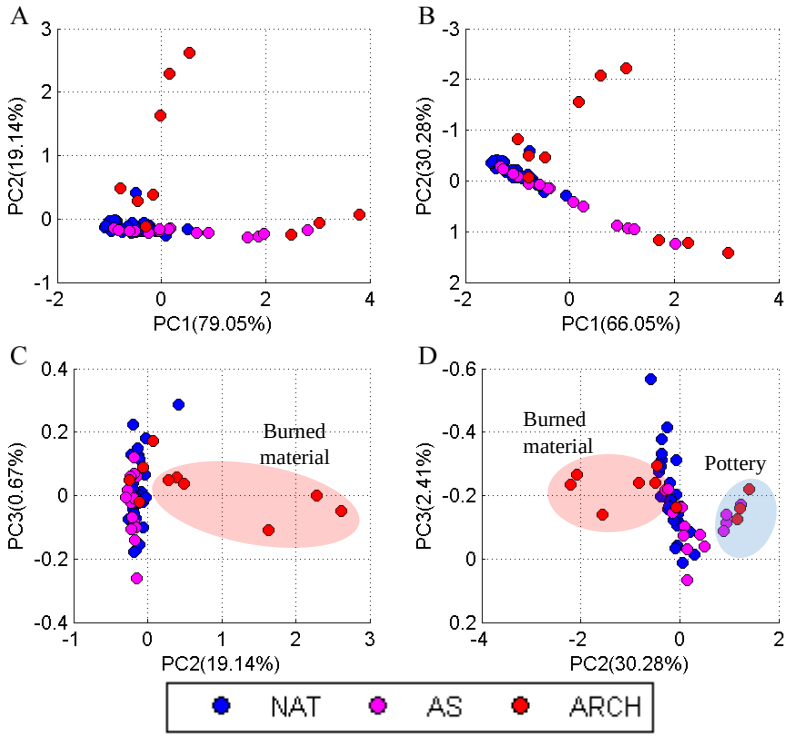


Figure 4.6.1 Score plots for site 2 with either the original PCs (left) or the base PC from site 1 (right). NAT represents the spectra of natural soils (including AP, Bt and C), AS those of archaeological soils and ARCH those of archaeological artefacts. Notice the y-axes are reversed for better comparison.

The base PC application seems to improve the result in some cases, at specific PC combinations, but does not always improve the result. In addition, it is difficult to define a dataset which can be referred to as the 'best' separated dataset (base PC). Even site 1, which was used as the base PC, showed some archaeological soils mixed within the natural soil cluster.

4.7 Numerical representation of the PCA score plot result

4.7.1 Distance to the centre of the cluster

The PCA results indicate that archaeological features can be distinguished from natural soils by looking at the score plot of, mainly, the first two PCs in the wavelength range 400 – 1000 nm. However, by representing such results as a scatter plot of the first two PCs they may depend on personal interpretation of determining what the boundary of the natural soil cluster is. In the following chapter methods, which aim to analyse the PCA results in a more quantitative way, developed in Chapter 3.3.4, are explained.

The representation of the PCA results in a scatter plot is a visual representation, and personal interpretation of how close a data point (for an individual spectrum) is to a cluster of a certain group decides the result. To obtain a more quantitative result, the so-called distance to the centre of the natural soil cluster is introduced. A centre point of the cluster will be used as a reference point to calculate the distance between each data point and this centre point. The

underlying assumption is that the spectra of natural soils are grouped together and the archaeological materials are not within the group of the natural soils. Therefore the centre of the cluster should be close to the natural soil spectra but far away for spectra of archaeological materials.

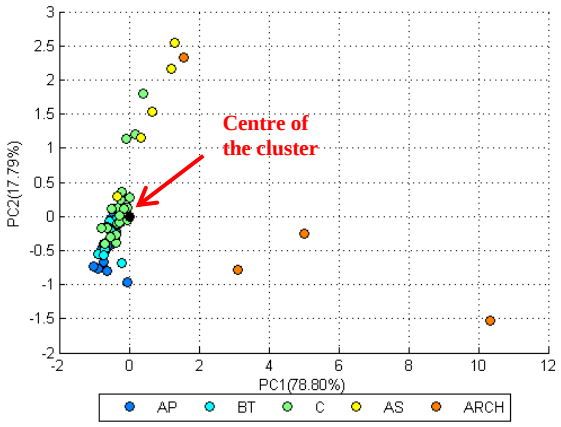


Figure 4.7.1 Score plot of PC1 against PC2 for site 1 for the wavelength range 400 – 1000 nm. The centre of the natural soil cluster is marked in black.

Table 4.1 Averaged distance value for each horizon to the centre of the cluster (black marker) for the results in Figure 4.7.1.

Horizon	Av. Distance
Ap	0.7654
Bt	0.6100
C	0.5372
AS	1.7322
ARCH	5.3728

Table 4.1 shows the distances between each horizon and the centre point. The results indicate that for natural soils, the average distance to the centre point is around 1 while for archaeological materials the distances vary from 1 (for AS) to 5 (for ARCH). The AS are slightly further away from the centre point than the other natural soils, but it is difficult to clearly identify the AS as non-natural soils if one does not have any information about where the spectra were taken. Although for ARCH an average distance value of 5 is found one can notice that ARCH spectra are widely scattered and some of the spectra are close to the natural soil cluster. Also by looking at Figure 4.7.1, depending on the shape of the cluster, some natural soil spectra are as far away from the centre point as some archaeological spectra. When the nature of the spectra is not known, it is difficult to identify whether such a spectrum is natural soil or not.

Overall, this method assigns a distance value to each spectrum, which in principal allows the separation of archaeological material from natural soils. However, still for some archaeological samples, no clear distinction from natural soils is possible. Also the method does not provide ranges of values indicating natural soils and depending on the shape of the cluster a natural soil sample may have distance as large as an archaeological spectrum.

4.7.2 Improvement of the separation based on a regression line

Based on Chapter 3.3.4.2, a regression line through the scatter points of the natural soils can be plotted. The perpendicular distance between the regression line and each scatter point (spectral measurement) can be calculated. Since the regression line only considers natural soil

spectra as variables, the perpendicular distances of the natural soil (Ap, Bt and C) and the scatter points should be smaller than those of the archaeological materials (AS and ARCH).

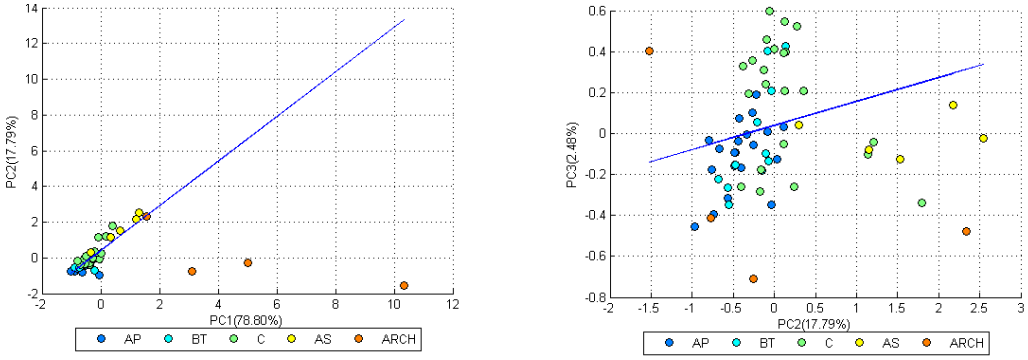


Figure 4.7.2 Regression lines plotted to the score plots of PC1 against PC2 (left) and PC2 against PC3 (right) for site 1 for the wavelength range 400 – 1000 nm. The regression lines only considered natural soils as variables.

Figure 4.7.2 and Table 4.2 show the regression line results as an image and in numerical format respectively, for the wavelength range 400 – 1000 nm for site 1 for PC1-PC2 and PC2-PC3 combinations. As expected, the orthogonal distances from the regression lines of the natural soil is much smaller than those of the archaeological materials (AS and ARCH). For both, PC1-PC2 and PC2-PC3 combinations the distance of archaeological materials is around five times larger than those of the natural soils.

Table 4.2 Average perpendicular distances between the regression lines and the data points for the different soil horizons for the results shown in Figure 4.7.2.

Score plot	Ap	Bt	C	AS	ARCH
PC1-PC2	0.1137	0.1097	0.1615	0.209335	4.1960
PC2-PC3	0.1345	0.2103	0.3121	0.2281	0.5990

The regression line method gives similar results as the centre point distance method (Chapter 4.7.1), but instead of measuring the distance from just one point, it measures the orthogonal distance from a regression line which goes through the natural soil cluster. Thus for most archaeological samples an improved separation from natural soils is found. However, as one can observe from Figure 4.7.2, some of the natural soil spectra are also as far away as the ARCH and AS spectra and these natural soils will thus not be identified as natural soils by this method. Another problem is that this method can only be applied when natural soils are identified independently, which makes it difficult to use when there is a set of unknown spectra.

4.8 D calculation: Calculation of the Euclidean distance

The methods used in the previous sections are partly successful in separating archaeological features from natural soils but also have their limitations. Therefore, this chapter focuses on identifying archaeological features among ‘common’ soils based on the PCA results. The soils are represented by the first two principal components of a group of natural soil spectra. Here,

it is assumed that although the spectral features of natural soils vary depending on the soil type and sampling location, they share a similar spectral pattern compared to the spectral features of archaeological remains. Using the method explained in Chapter 3.3.2, a difference (D) between the original spectrum (S) and the modified spectrum (S'), which represents the principal component (PC) values of natural soils, is calculated. If the difference (D value) between the two spectra is small (close to 0), then the spectrum is similar to the spectral features of natural soils (D_{nat}). If the difference is large, then the spectrum is more likely to belong to a non-natural soil, probably an archaeological material (D_{arch}). To make the results independent form the absolute D values, a D_{ratio} is calculated, which is the ratio between D_{arch} and D_{nat} (refer to Chapter 3.3.3 for detailed mathematical explanation on the D_{ratio}). Large D_{ratio} values (larger than 1) probably represent archaeological materials. If the D_{ratio} value is less than 1, the measured spectrum is probably a non-archaeological material.

Refer to Chapter 3.3.2 for a detailed explanation of the D calculation method. Also the definition and detailed information on different types of N_{soil} used are explained in Chapter 3.3.2.1.

4.8.1 D calculation results for local N_{soil}

The D_{ratio} results are calculated for the archaeological sites 1 and 2 from Calabria, Italy using the N_{soil} from either the original site (site 1) or from Hungary (Figure 3.3.6), The results for N_{soil} from the original site are expected to give low D_{nat} values (since identical natural soil spectra are) and thus high D_{ratio} results. The results for N_{soil} from Hungary should give higher D_{nat} values because N_{soil} contains spectral features different to those in Italy as shown in Figure 3.3.6, and thus also lower D_{ratio} results.

Figure 4.8.1 shows the D values for site 1 when N_{soil} is used from the original site (left) and Hungary (right). Figure 4.8.1 shows that the D_{nat} values are reduced by a factor 5 when N_{soil} is used from the original site (site 1) compared to when N_{soil} from Hungary is used, but the D_{arch} values (0.6 to 1) are fairly similar for both N_{soil} combinations.

When the N_{soil} is obtained from the site where the measurements were taken (site 1), the D_{nat} values vary between 0.07 and 0.10 and the D_{arch} values vary between 0.70 and 1.10, depending on the number of PCs used. The D_{arch} values are around a factor of ten larger than D_{nat} resulting in D_{ratio} values around 10. In the following the number of considered PCs is indicated by the summation symbol (for example, $\sum_1^2 \mathrm{PC}$ indicates that the first two PCs are used (for the S' calculation, refer to Equation [8]). For $\sum_1^2 \mathrm{PC}$ the D_{ratio} is slightly larger than for $\sum_1^3 \mathrm{PC}$, but this difference is fairly small. The error ranges in this case are rather large, but depending on the type of archaeological material, the D_{ratio} value may also be large (up to > 40). The average D_{ratio} value for N_{soil} from the local site is around 10, which is extremely large and therefore indicates that the spectra of archaeological materials do stand out among the spectra of natural soils.

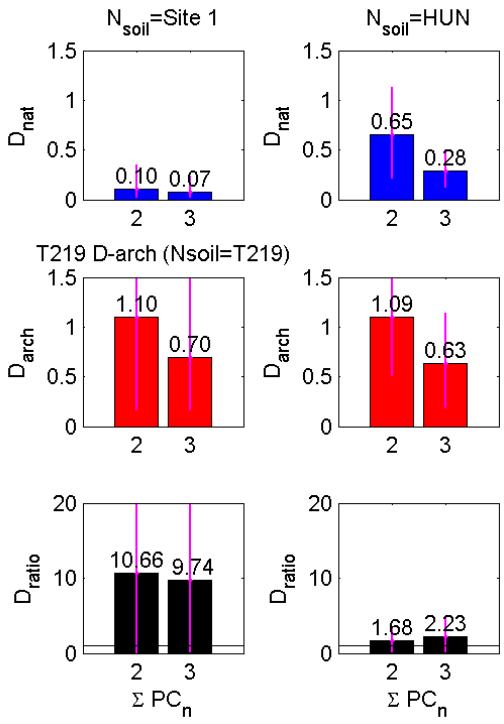


Figure 4.8.1 D_{nat}, D_{arch} and D_{ratio} values for site 1, Italy for different sets of N_{soil} (either from the original site (site 1) or Hungary (HUN)). The error bars are calculated by the maximum and minimum D values using the error propagation method (refer to Chapter 3.3.3). The x-axis represents the number of PCs used for the D calculation. Here the results from only $\sum_1^2 PC$ and $\sum_1^3 PC$ are shown since for PC4, the contribution drops below 1% (refer to Figure 3.3.6 and Figure 4.5.1). For all cases, D_{nat} values are much smaller than D_{arch} values causing D_{ratio} values larger than 1.

When N_{soil} is calculated from natural soil spectra gathered at a totally different country (HUN), the D_{arch} values are similar to the D_{arch} obtained for the original site (site 1). However, the D_{nat} values are around a factor of 6 larger indicating that natural soil spectra of site 1 and spectra of Hungarian soils are not similar. However, the average D_{ratio} values are still larger than 1, which once more emphasises that the difference between natural soils gathered in various regions is smaller than the difference between spectra of soils and archaeological features. Also these results indicate that even if N_{soil} does not include any natural soils from the original site, the method still works. N_{soil} from HUN gives better result when PCs are added up to the third PC. This suggests that the first two principal components of the Hungarian soils do not contain many spectral features of the natural soils at site 1 and therefore another common feature might be represented by PC3.

Figure 3.3.6 reveals that the natural soil spectral features (N_{soil}) are different depending on the site where the samples were gathered. Thus, when N_{soil} is used from a completely different location (country), this will result in a large D_{nat} value indicating systematic spectral differences. However, these spectral differences between natural soils are not larger than the difference to the archaeological material and, therefore, D_{nat} values are still smaller than D_{arch} values.

Figure 4.8.2 illustrates the D_{ratio} values for individual spectra from site 1, with D values larger than 1 are represented by red markers. The first impression is that when N_{soil} is used from the original site, all of the archaeological spectra are identified with large D_{ratio} values. However, some natural soil spectra also show D_{ratio} values slightly higher than 1, indicating that some false detection can occur. This is more extreme when N_{soil} is used from the Hungarian soil. Here even some of the dominant archaeological materials are not identified (burned material,

marked by the red circle) and many natural soils gave large D_{ratio} values. This result is almost independent from the number of PCs added, although there are slight differences in which archaeological spectrum does not have D_{ratio} higher than 1.

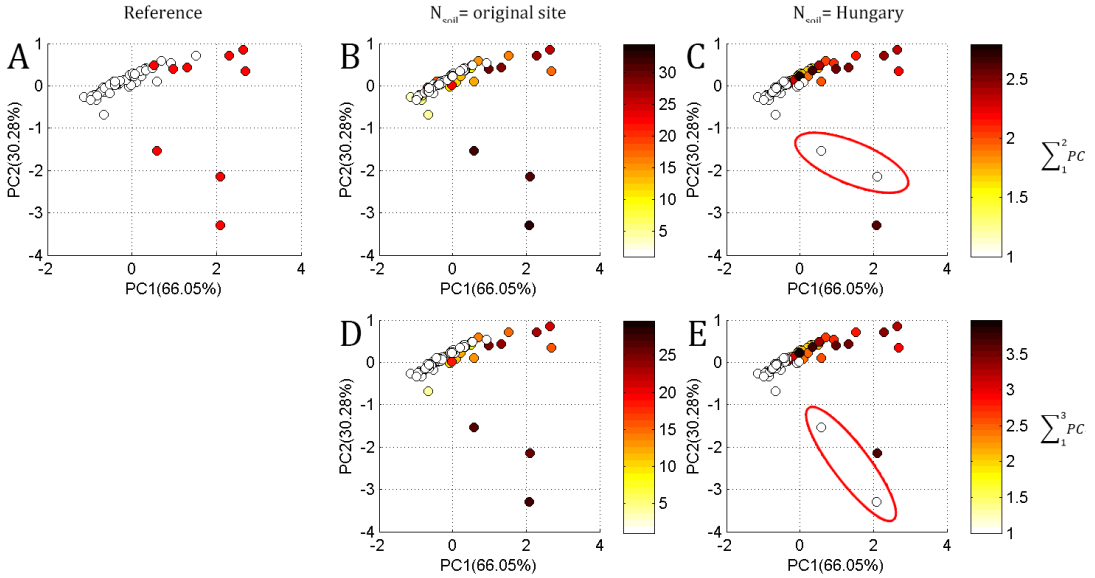


Figure 4.8.2 D_{ratio} values represented as colour on the PC1 against PC2 score plot of site 1. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) Results for N_{soil} used from the original site (site 1) with $\sum_1^2 PC$ (first two principal components used from N_{soil}, refer to Equation [7] and [8]). C) Results for N_{soil} used from Hungary with $\sum_1^2 PC$. D) Results for N_{soil} used from the original site (site 1) with $\sum_1^3 PC$(first three principal components used from N_{soil}). E) Results for N_{soil} used from Hungary with $\sum_1^3 PC$.

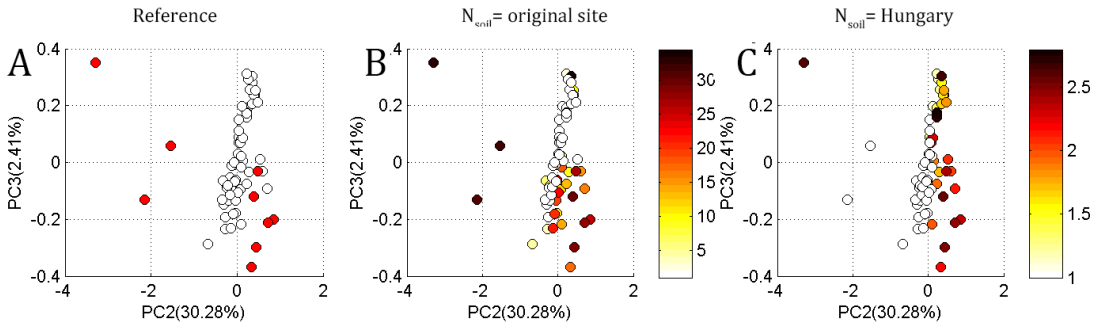


Figure 4.8.3 D_{ratio} values represented as colour on the PC2 against PC3 score plot of site 1. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) Results for N_{soil} used from the original site (site 1). C) Results for N_{soil} used from Hungary. Here, only the first two principal component values ($\sum_1^2 PC$) for N_{soil} are presented, refer to Appendix 3.1 for $\sum_1^3 PC$.

Figure 4.8.3 shows the same results as shown in Figure 4.8.2 but now the score plots for PC2

against PC3 are shown. Similar to the results from Figure 4.8.2, all archaeological spectra are identified when N_{soil} is from the original site, but several false detections in the natural soils are made (Figure 4.8.3.B). When N_{soil} from Hungary is used, most of the burned materials are not detected, but many of the reddish archaeological strata are identified.

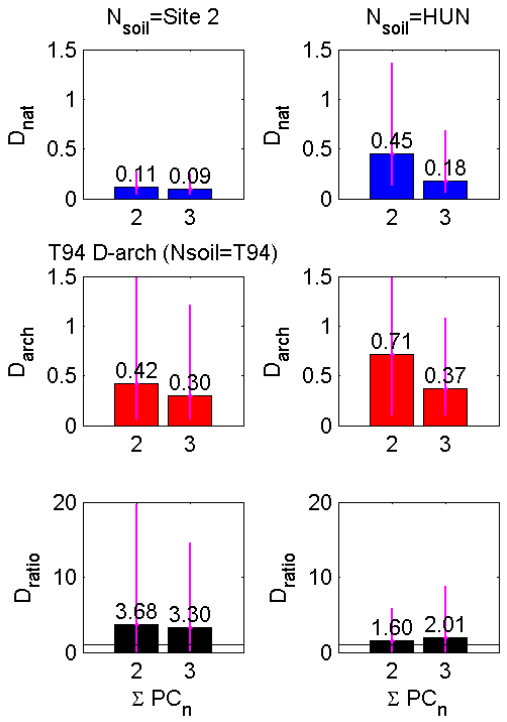


Figure 4.8.4 D_{nat}, D_{arch} and D_{ratio} values for site 2, Italy for different sets of N_{soil} (the original site (site 2) and Hungary (HUN)). The error bars are calculated by the maximum and minimum D values using the error propagation method (refer to Chapter 3.3.3). The x-axis represents the number of PCs added to the D calculation, here only the results from $\sum_1^2 PC$ and $\sum_1^3 PC$ are shown, since for PC4 the contribution drops below 1% (refer to Figure 3.3.6 and Figure 4.5.1). For all cases, D_{nat} values are much smaller than D_{arch} values resulting in D_{ratio} values larger than 1.

Next, the same steps are applied to site 2 from Calabria, Italy. Figure 4.8.4 shows the D values for site 2 with N_{soil} from the original site (site 2) and from Hungary. The overall D value results of site 2 are similar to those of site 1. When N_{soil} is from the original site (site 2), the D_{nat} value is around 0.10, which is similar to the D_{nat} value for site 1. However, the average D_{arch} value is around 0.4, which is much smaller than that of site 1 (0.7 to 1.1). Also the average D_{ratio} is smaller than for site 1. Although the two archaeological sites (site 1 and site 2) are from different locations, they share a similar archaeological type (burned materials and potteries of prehistoric kitchen formation) and are located in the same region. Both sites have preserved remains of burning activities such as pitch black burned materials, reddish soils and ceramic pieces. However, the different D_{arch} values for the two sites suggest that although similar archaeological materials are found, they are spectrally different.

When N_{soil} is taken from Hungary, the D_{nat} value is around 0.18 to 0.45 depending on the number of PCs added. However, the corresponding D_{arch} values are still larger than D_{nat}. Similar to the result from site 1 (Figure 4.8.1), the D_{ratio} value increases when $\sum_1^3 PC$ is used. Here the D_{arch} value is around twice as large as D_{arch} for the original site when $\sum_1^2 PC$ is used. This indicates that the archaeological material gathered at site 2 has similar spectral patterns to the natural soils gathered within the site, but different compared to the natural soils gathered in Hungary. Here, the D_{ratio} values are larger than 1 but not as high as for site 1, where the D_{ratio} values exceeded 10. When N_{soil} is used from the original site (site 2), the D_{ratio} value is around

3, and when N_{soil} is used from Hungarian soils, the D_{ratio} value varies between 1.6 and 2. The latter result is similar to the D_{ratio} value for site 1 when N_{soil} from Hungary was used.

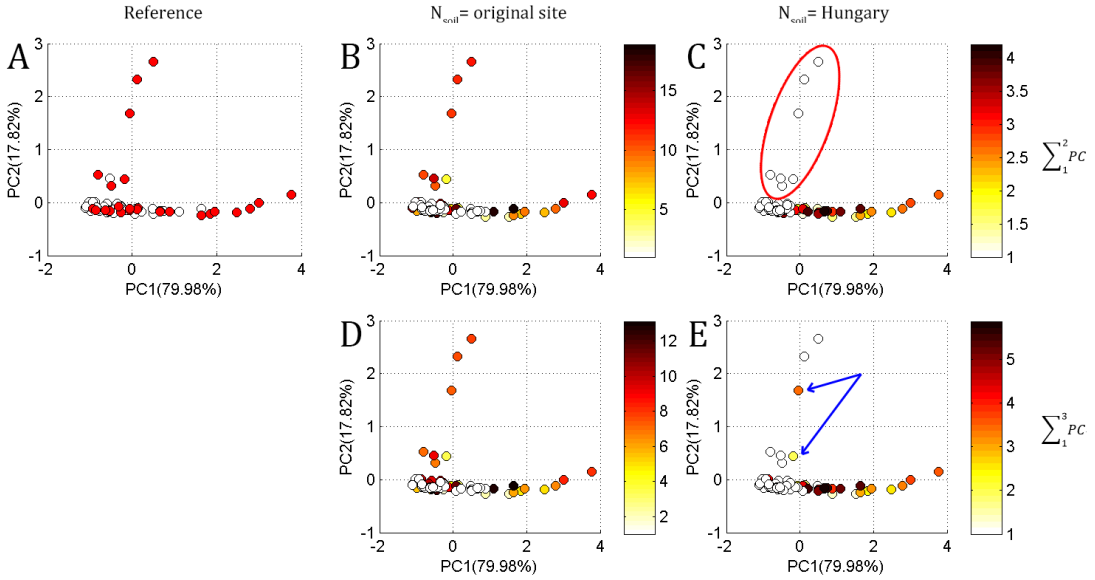


Figure 4.8.5 D_{ratio} values represented as colour on the PC1 against PC2 score plot of site 2. Dratio values below 1 are represented by white (empty) markers and Dratio higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) Results for Nsoil used from the original site (site 2) with $\sum_1^2 PC$ (first two principal components used from Nsoil, refer to Equation [7] and [8]). C) Results for Nsoil used from Hungary with $\sum_1^2 PC$. D) Results for Nsoil used from the original site (site 2) with $\sum_1^3 PC$(first three principal components used from Nsoil). E) Results for Nsoil used from Hungary with $\sum_1^3 PC$.

Figure 4.8.5 shows the score plot of site 2 with D values coloured. Unlike site 1 where all the archaeological features showed D_{ratio} values higher than 1, at site 2, many of the 'archaeological soils' (which were not reddish in colour like at site 1, but rather showed bright grey and yellowish colours) are not identified when N_{soil} is used from the original site (site 2). Among 26 archaeological spectra, 20 of them resulted in a D_{ratio} larger than 1. The archaeological materials, marked by the red circle (Figure 4.8.5.C), are not detected at all and many of the natural soil spectra showed a large D_{ratio} value. However, when N_{soil} from Hungary is used, some of the archaeological spectra, those which were not identified when N_{soil} from original site was used, showed large D_{ratio} values eventually resulting also in higher average D_{ratio} values. When up to the third principal component is used ($\sum_1^3$ PC), some archaeological spectra which were not identified previously give a D_{ratio} larger than 1 (marked by the blue arrows) indicating that the results improved.

The overall results show that when the N_{soil} is collected from the original site, it is more likely to obtain high D_{ratio} values regardless how many PCs are used ($\sum_1^2$ PC or $\sum_1^3$ PC does not make large difference). However, when the N_{soil} is used from a completely different site, then using up to $\sum_1^3$ PC is more likely to produce better results.

4.8.2 D calculation results for global N_{soil}

The previous section showed D results when N_{soil} were collected from either the local original site or another site with a limited number of natural soil spectra used as N_{soil}. In this chapter, the calculations are repeated for a 'universal' N_{soil} dataset. Two types of global datasets are used: 1) all natural soil spectra gathered from both Italy and Hungary (IT+HUN) and 2) natural soils from the ISRIC spectral library (ICRAF-ISRIC, 2010). As mentioned in Chapter 3.3.2.1, the ISRIC spectra are undersampled by only providing values every tenth nanometre. Therefore, to test the effect of this undersampling, the N_{soil} from Italy and Hungary (IT+HUN) is used to investigate the differences of the results for the original spectra and the mathematically degraded spectra (resampled spectra at every tenth-nanometre).

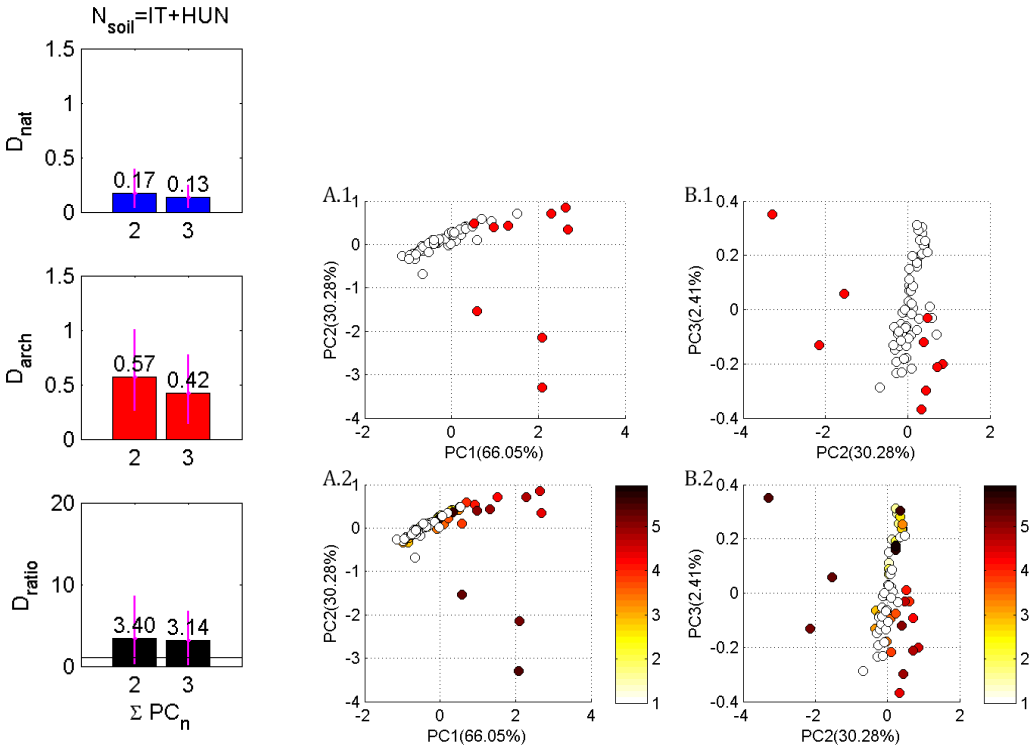


Figure 4.8.6 D calculation results for site 1 when a combined N_{soil} from Italy and Hungary (IT+HUN) is used. The bar chart on the left side shows the D_{nat}, D_{arch} and D_{ratio} values for $\sum_1^2 PC$ and $\sum_1^3 PC$.
The PCA score plots on the right sides are D_{ratio} values represented as colour, where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. Figure A) shows the PC1 against PC2 score plot and B) represents the PC2 against PC3 plot. Top figures (A.1 and B.1) are the reference figures where archaeological materials are indicated by red markers. Here, only the first two principal component values ($\sum_1^2 PC$) for N_{soil} are presented.

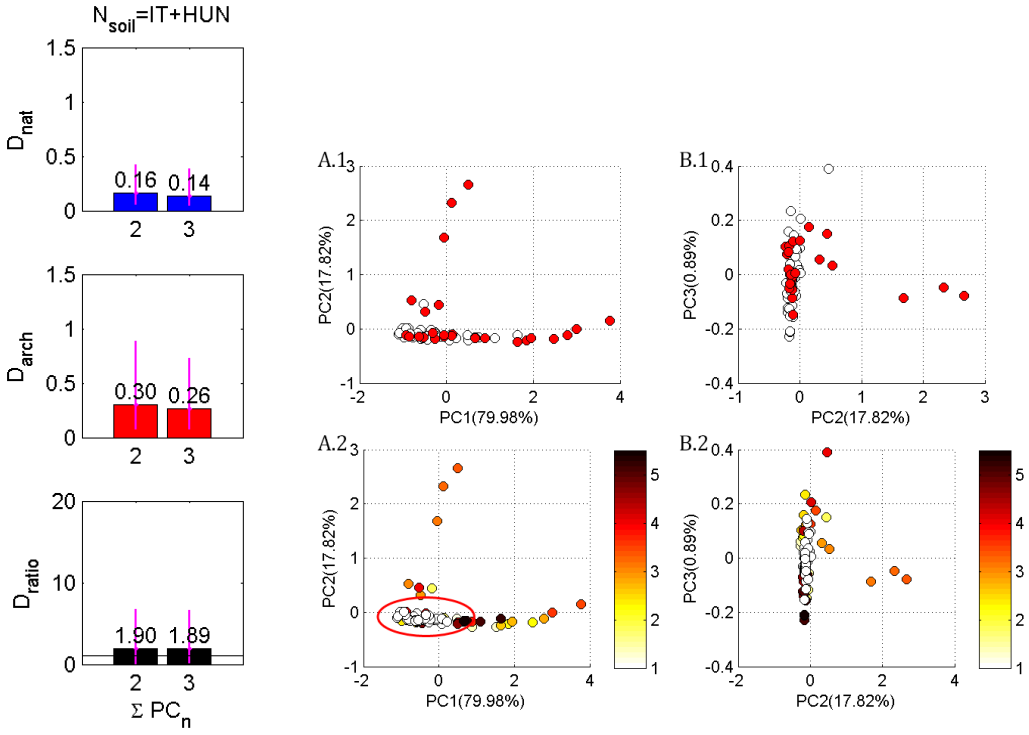


Figure 4.8.7 D calculation results for site 2 when a combined N_{soil} from Italy and Hungary (IT+HUN) is used. The bar chart on the left side shows the D_{nat}, D_{arch} and D_{ratio} values for $\sum_1^2 PC$ and $\sum_1^3 PC$.
The PCA score plots on the right sides are D_{ratio} values represented as colour, where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. Figure A) shows the PC1 against PC2 score plot and B) represents the PC2 against PC3 plot. Top figures (A.1 and B.1) are the reference figures where archaeological materials are indicated by the red markers. Here, only the first two principal component values ($\sum_1^2 PC$) for N_{soil} are presented.

Figures 4.8.6 and 4.8.7 show the D values for sites 1 and 2 when N_{soil} is used from both Italy and Hungary. The D_{nat} values of both sites vary between 0.13 and 0.17 indicating that the natural soils from these two sites are fairly identical. However, the D_{arch} value from site 1 is around 0.4 to 0.5 which is slightly larger than for site 2 (0.2 to 0.3). This instantly results in a higher D_{ratio} value of 3.4 for site 1. This value is much smaller than that of the N_{soil} used from the original site (D_{ratio} of 10), but still larger than for the case that N_{soil} was gathered only from Hungary (D_{ratio} of 1.7 to 2.2, refer to Figure 4.8.1). For site 2, the D_{ratio} varies around 1.90 which is smaller than for N_{soil} taken only from the original site (D_{ratio} of 3.3 to 3.7) but larger than for the case that when only spectra from Hungary were used (D_{ratio} of 1.6 to 2, refer to Figure 4.8.4).

The PCA score plots with coloured data points (D_{ratio}) are shown in the left part of Figure 4.8.6 and 4.8.7. For site 1, both PC1-PC2 and PC2-PC3 plots indicate that all archaeological materials have D_{ratio} values larger than 1, but there are still some natural soils with high D_{ratio} values. One interesting thing to notice is that burned archaeological material tends to reveal

higher D_{ratio} values than the reddish soil stratum and potteries. However, for site 2, not all of the archaeological materials show D_{ratio} values higher than 1. Especially, the spectra which were mixed into the natural soil cluster (marked by the red circle) show D_{ratio} values less than 1. Thus also a low average D_{ratio} value of 1.9 is obtained.

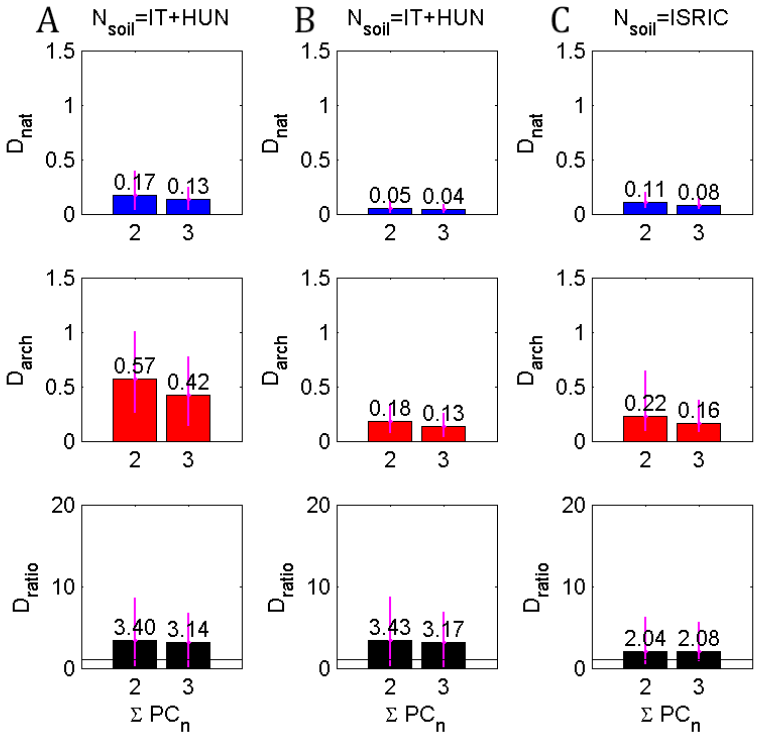


Figure 4.8.8 D value (D_{nat}, D_{arch} and D_{ratio}) results for site 1 for different 'universal' N_{soil} data sets.
A: N_{soil} from both Italy and Hungary. B: N_{soil} from both Italy and Hungary but for spectra resampled at a ten-nanometre grid. C: N_{soil} from the ISRIC spectral library (spectra are resampled at a ten-nanometre grid).

Results so far have shown that the D calculation method works sufficiently well to identify archaeological spectral features among a group of natural soils, although there are several false detections. The D method works best when the N_{soil} is collected from the site where archaeological materials are investigated (local site). However, the HUN and IT+HUN combination for N_{soil} have also shown that the method can work for cases where natural soils are not gathered from the original site. Therefore, although one is unclear whether a measured spectrum represents natural or archaeological soils, the method can still work. To prove this in a more global scale, the D calculation will be applied with N_{soil} gathered from the ISRIC spectral library. Before applying N_{soil} from the ISRIC spectral library, a simple test is performed to check whether resampling at every tenth-nanometre dramatically influence the D values.

The first two rows in Figure 4.8.8 represent D values for the IT+HUN combination, either at the original spectral sampling (left) or at a degraded sampling (10 nm), which is similar to the spectra from the ISRIC spectral library. Since this resampling process to a loss of spectral information, one would expect to obtain low D_{ratio} values. However, the results indicated that

there is not much difference between the two types of spectra. The D_{nat} and D_{arch} value for the resampled dataset are around a factor of three smaller than the original D values. However, still the D_{arch} (0.13 to 0.18) values of the resampled data are about three times larger than the D_{nat} value (0.04 to 0.05), which is similar to the original data. This leads to a similar D_{ratio} value of around 3.17 to 3.4.

This result is not completely unexpected, because the original spectral resolution of the measured spectra (FWHM) is of 3 nm and 10 nm depending on the spectral range (see Table 3.1). Moreover, because of the low signal too noise ratio of the spectra at short wavelengths a smoothing with a 10 nm convolution kernel was applied (Chapter 3.2.3). Thus the loss of information by the re-sampling is relatively small. But still the results indicate that those spectral archaeological signatures (and also the signatures of natural soil spectra) are mainly contained in the 'broad' spectral features (with bandwidth ≥ 10 nm). The dependence of the D_{ratio} values on the spectral resolution of the measurements is further investigated in Chapter 4.9. Overall, it is an important result that the spectral re-sampling has a negligible effect on the D_{ratio} values, because it justifies the use of the ISRIC spectral library.

When N_{soil} is used from the ISRIC spectral library, the D_{nat} value is around 0.08 to 0.11, which is similar to the D_{nat} value when N_{soil} is used from the local original site (D_{nat} value of 0.07 to 0.10, refer to Figure 4.8.1). This is can be understood by the fact that the first principal component of the ISRIC natural soils is similar to the first principal component of the natural soils from Italy, although there are some dissimilarities in the shape of the absorption feature at 400 to 600 nm (refer to Chapter 3.3.2.1). Unlike the D_{nat} values, the D_{arch} values from the ISRIC (0.16 to 0.22) are fairly small (although still larger than the D_{arch} values for N_{soil} from Italy and Hungary which is around 0.13 to 0.18). It is around five times smaller than the D_{arch} of the local soil, which was around 0.70 to 1.10 (refer to Figure 4.8.1), indicating that it becomes more difficult to identify archaeological signatures when global natural soils are used. However, the resulting D_{ratio} values of around 2 are still sufficiently high to identify archaeological. Also $\sum_1^3 PC$ gives a slightly higher D_{ratio} value than $\sum_1^2 PC$, because as shown in Figure 4.8.9, some archaeological materials which had a D_{ratio} value lower than 1 for the $\sum_1^2 PC$ case provide D_{ratio} values larger than 1 when $\sum_1^3 PC$ is used.

Figure 4.8.9 shows the score plots with D_{ratio} values coloured. Here, characteristic archaeological materials (such as potteries and burned materials) always provide D_{ratio} values larger than 1. When for N_{soil} from the combined IT+HUN is used, 27 natural soil spectra give D_{ratio} larger than 1. However, when N_{soil} is taken from ISRIC, there is less false detection of natural soil spectra. Therefore although the D_{arch} values are smaller (Figure 4.8.8) when the ISRIC spectra are used, the problem of false detection of natural soils as archaeological materials is reduced.

Figure 4.8.10 shows the D_{ratio} values for site 2. Similar to the results from site 1, when N_{soil} is used from Italy and Hungary and only the spectral sampling was changed, the overall D_{ratio} values are identical. Here, the resampled data have D_{nat} and D_{arch} values of around a factor three less than the original value (which is exactly the same as the results shown in Figure 4.8.8 for site 1) but the ratio between D_{nat} and D_{arch} remains similar leading to a D_{ratio} value of 1.90.

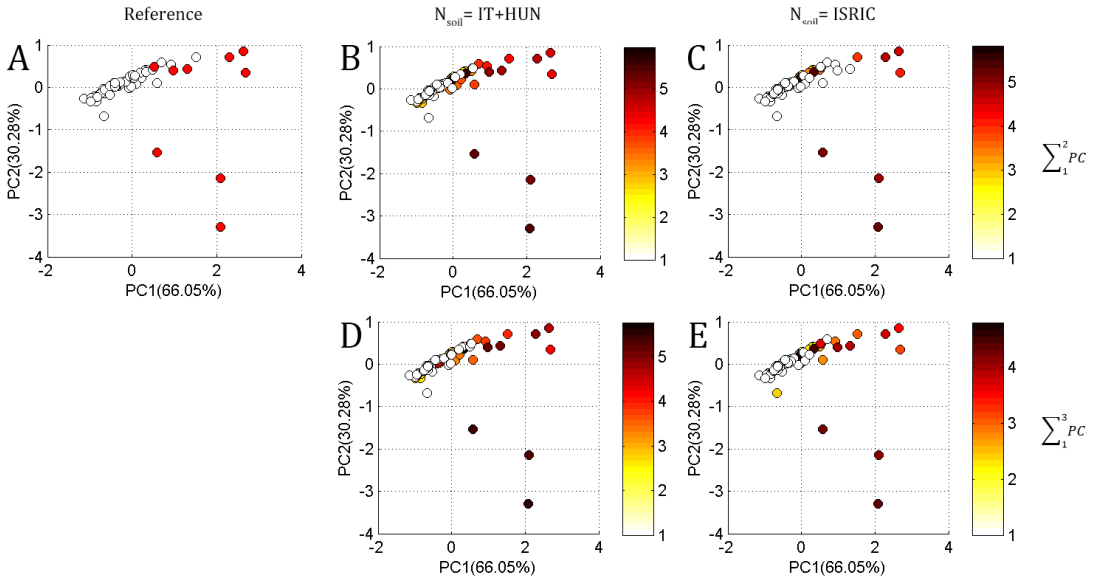


Figure 4.8.9 D_{ratio} values represented as colour on the PC1 against PC2 score plot of site 1. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) Results for N_{soil} used from both Italy and Hungary with $\sum_1^2 PC$ (first two principal components used from N_{soil}). C) Results for N_{soil} used from ISRIC with $\sum_1^2 PC$. D) Results for N_{soil} used from both Italy and Hungary with $\sum_1^3 PC$(first three principal components used from N_{soil}). E) Results for N_{soil} used from ISRIC with $\sum_1^3 PC$.

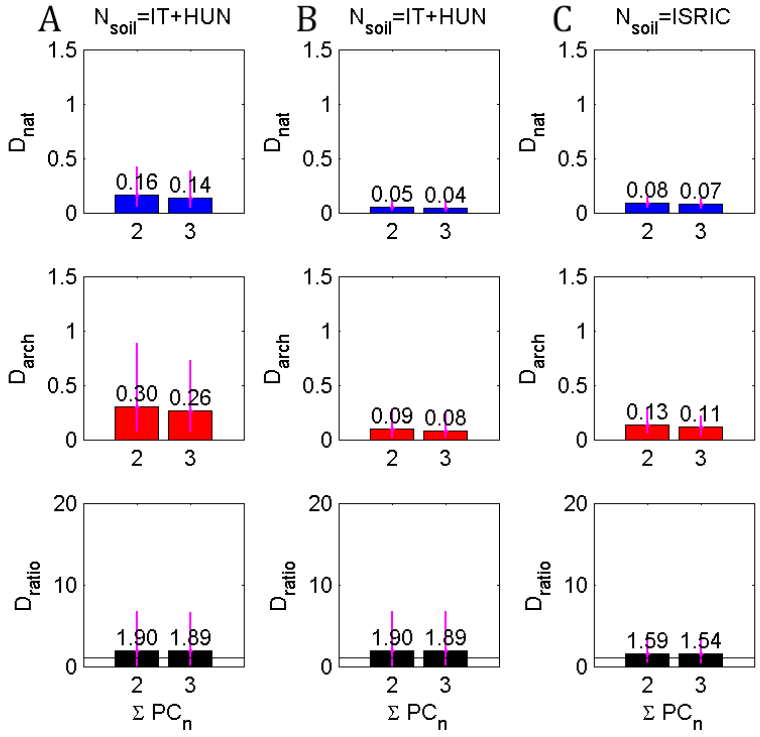


Figure 4.8.10 Global N_{soil} D value (D_{nat}, D_{arch} and D_{ratio}) results for site 2.
A: N_{soil} from both Italy and Hungary. B: N_{soil} from both Italy and Hungary for spectra were resampled on a ten-nanometre grid. C: N_{soil} from ISRIC spectral library.

When the global natural soil data set from the ISRIC spectral library is used, the overall results are once again similar to those from site 1. The D_{nat} values are fairly similar to the D_{nat} values of the local soil but have relatively smaller D_{arch} values leading to smaller D_{ratio} values of around 1.55. However, the D_{ratio} is still larger than 1 illustrating that the method works well for the archaeological sites investigated here (kitchen remains).

Figure 4.8.11 once again shows the D_{ratio} values colour coded on the PCA score plot. Here, among 25 archaeological material spectra, 18 are detected when N_{soil} from IT+HUN is used and 15 spectra are detected when N_{soil} from the global ISRIC data set is used. The burned material (marked by blue arrows) is identifiable when natural soils from Italy and Hungary are used but its identification becomes less clear when the ISRIC data set is used. The natural soils from the ISRIC spectral library do provide good D_{ratio} values, but sometimes might miss important archaeological features.

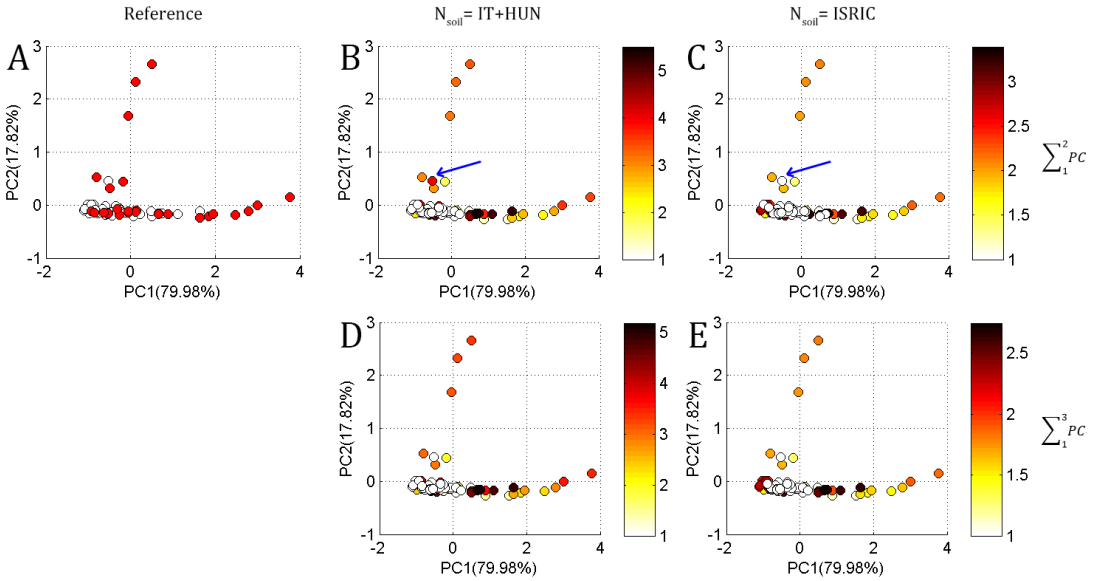


Figure 4.8.11 D_{ratio} values represented as colour on the PC1 against PC2 score plot of site 2. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) Results for N_{soil} used from both Italy and Hungary with $\sum_1^2 PC$ (first two principal components used from N_{soil}, refer to Equation [7] and [8]). C) Results for N_{soil} used from ISRIC with $\sum_1^2 PC$. D) Results for N_{soil} used from both Italy and Hungary with $\sum_1^3 PC$(first three principal components used from N_{soil}). E) Results for N_{soil} used from ISRIC with $\sum_1^3 PC$.

From the D value results from sites 1 and 2, one can conclude that the D calculation approach works well and allows the determination archaeological features among natural soil spectra. The method performs best when known natural soil spectra are collected from the archaeological site. In this case N_{soil} is preferred to be calculated with only the first two principal components to produce larger D_{ratio} values. The results also show that when the D_{ratio} value is larger than 1.5 for an unknown site using N_{soil} from the ISRIC, then that spectrum is likely to be an archaeological feature. This is an important finding as it indicates that the method can potentially be applied universally.

Additionally, the false detected natural soil spectra at site 1 were investigated carefully to observe whether these spectra were collected from the soil horizon just below or around the archaeological stratum (site 2 had archaeological remains on the floor of the pit, refer to Figure 4.1.4). These false detected natural soil spectra were not collected from the archaeological soil pit, but are actually bare soil spectra (mainly from the C horizons, and only few from Ap or Bt horizons) measured from natural soil pits around the site. This once again indicates that the variety of natural soils in the Calabria region is large and that sometimes natural soil can have a spectral characteristic which are fairly different from other natural soil spectra.

The results also gave some indications that possibly no high resolution spectral measurements are needed to distinguish between archaeological and natural soil spectra. The following Chapter 4.9 will investigate the dependence of the D_{ratio} values on the spectral resolution in more detail.

4.9 Dependence of the results on the spectral resolution

In this section, the PCA and D calculation method are applied to 'degraded' spectra. The spectrometer used in this thesis has a spectral resolution of about 3 nm in the visible range and 10 nm in the infrared region (Chapter 3.1.2). This section investigates which minimum spectral resolution is needed for detecting archaeological materials using the modified PCA method introduced in this thesis. In the previous section (Chapter 4.8.2), when only every tenth nanometre is selected from the spectrum, the results did not change dramatically compared to the original spectra. Hence, a 'high resolution' spectrometer might not be needed for the identification of the archaeological features.

This section investigates the effect of various spectral resolutions by smoothing the spectra and thus tries to identify the minimum resolution level which is necessary to detect archaeological materials.

Figure 4.9.1 represents the original and smoothed spectra (after the continuum was removed) for the spectral range 400 – 1000 nm for different convolution kernels. The method of degrading (or smoothing) is explained in Chapter 3.2.3. Notice that many small features are smoothed away, and the shape of the spectra starts to change as it is smoothed by kernel levels of ≥ 50 nm.

Figure 4.9.2 illustrates the PC1-PC2 score plots when different smoothing kernels were applied for site 1 in Italy. Notice that AS and ARCH are always separated from the natural soil cluster and only the orientation of the score plot changes as the spectra are smoothed. When the smoothing kernel is 200 nm (Figure 4.9.2.F), the ARCH samples are still clearly separated. This may suggest that narrow absorption features do not play a great role in separating archaeological material from natural soil. Instead, broadband features of the soils determine the characteristic feature of archaeological materials. However, it should be noted that for a smoothing kernel of 200 nm the AS spectra are mixed into the natural soil cluster. Overall, the results indicate that, for the distinction of archaeological materials from natural soils a rather low spectral resolution (≤ about 50 nm) is probably sufficient. This finding also explains t he results in Chapter 4.8.2 where the D_{ratio} for the degraded spectra (resampled at a ten na-

nometre grid) are similar to the original dataset.

Figure 4.9.3 shows the D values and D_{ratio} values for site 1for the different levels of smoothing when global natural soils are used. Notice that the D values (D_{nat}, D_{arch} and D_{ratio}) do not change until the spectra are smoothed by the 20 nm kernel and from smoothing kernels ≥ 50 nm the D_{ratio} values gradually decrease. This can be understood by the spectral features shown in Figure 4.9.1 where the shape of the spectrum starts to smooth out small absorption features for smoothing kernels ≥ 50 nm.

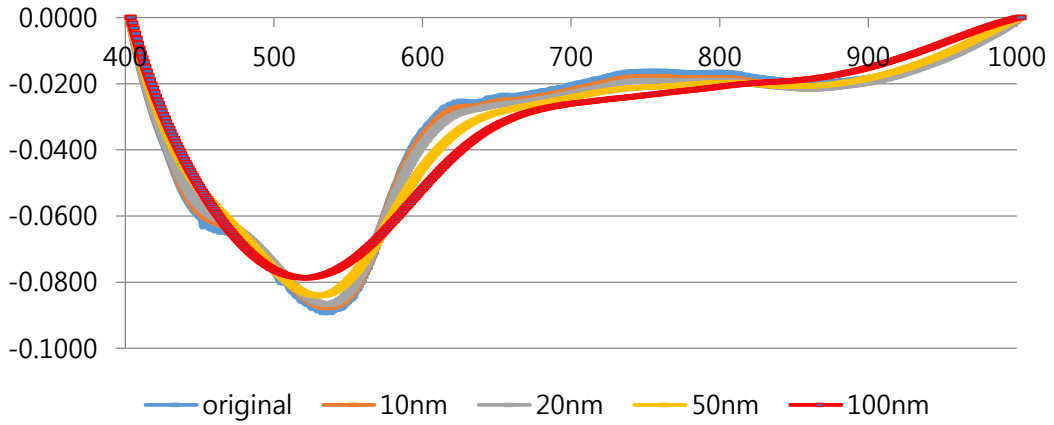


Figure 4.9.1 Original spectrum and smoothed spectra for various convolution kernels.

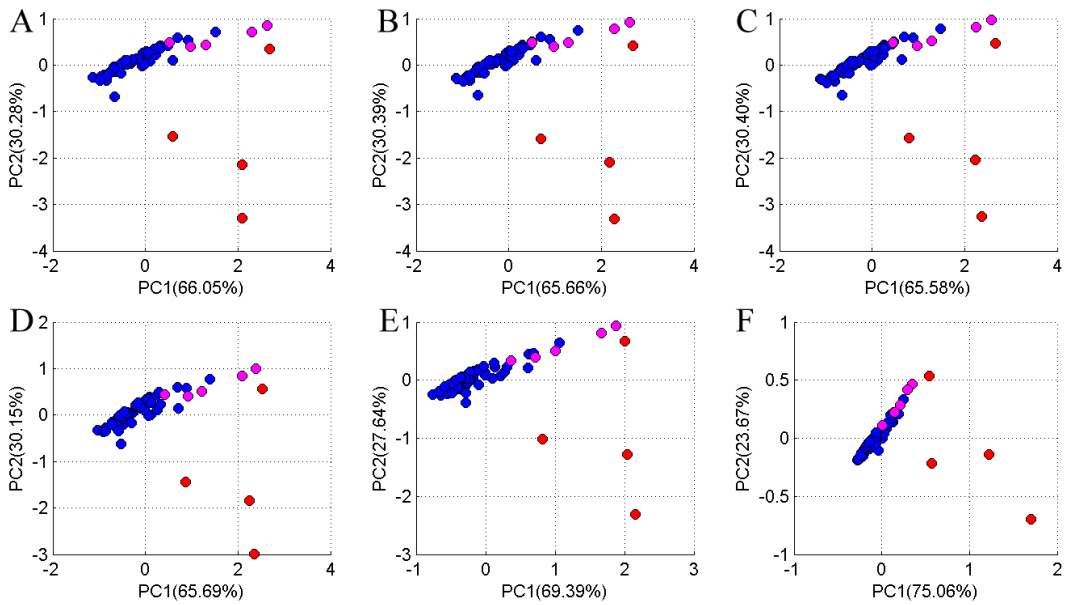


Figure 4.9.2 Score plots for the first two PCs for site 1 when different smoothing kernels were applied. The convolution kernels used for smoothing are; A: original, B: 10 nm, C: 20 nm, D: 50 nm E: 100 nm and F: 200 nm.

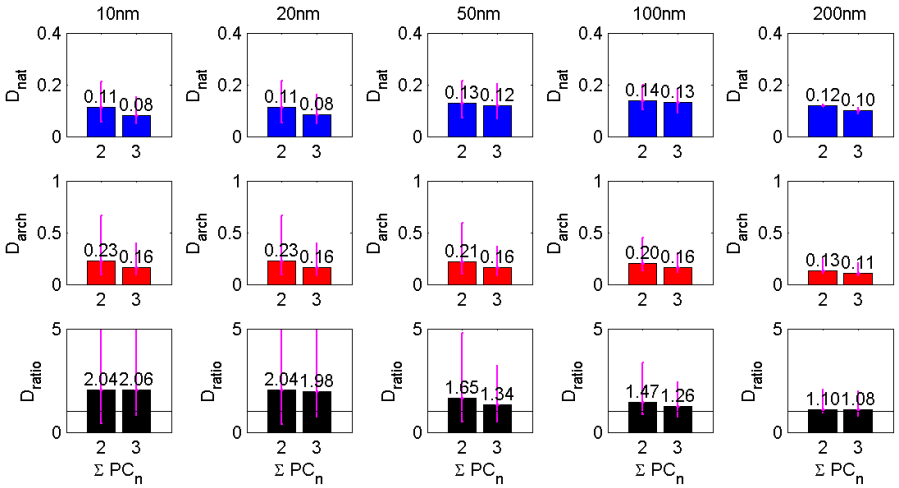


Figure 4.9.3 D values for the site 1 where spectra are smoothed between 400 – 1000 nm at different convolution kernel (10 nm, 20 nm, 50 nm, 100 nm and 200 nm). Here, N_{soil} is from the natural soils of ISRIC spectral library.

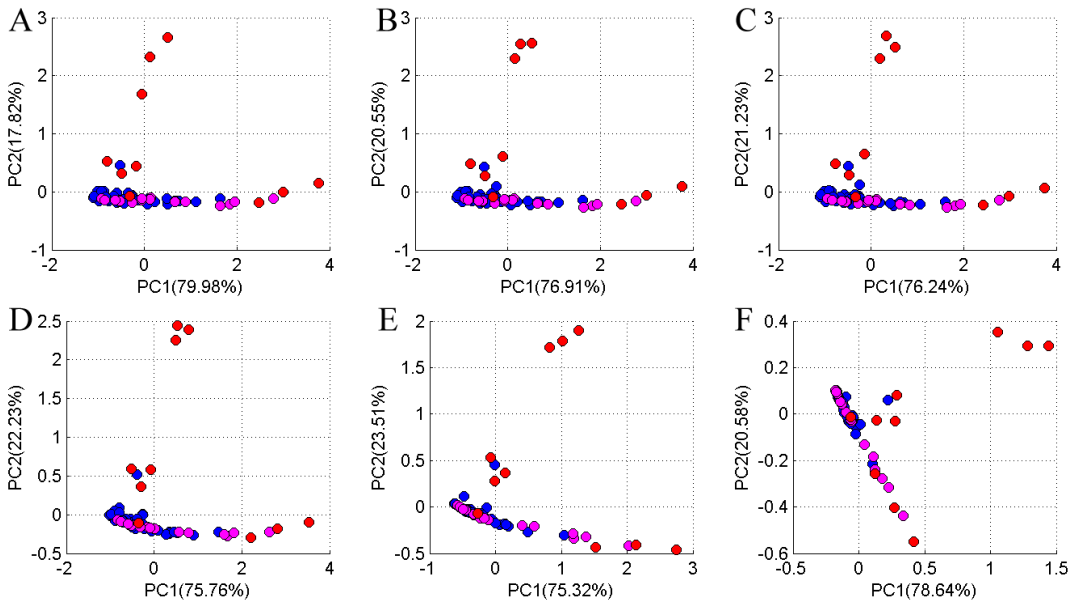


Figure 4.9.4 Score plots for the first two PCs for site 2 when different smoothing kernels were applied. The convolution kernels used for smoothing are; A: original, B: 10 nm, C: 20 nm, D: 50 nm E: 100 nm and F: 200 nm.

When the spectra are smoothed by 200 nm, the D_{arch} value strongly decreases to values of 0.11 or 0.13 (which is around half of that for the original spectra) indicating that the characteristic spectral features of archaeological materials disappear with higher level of smoothing. This can be explained by looking back at Figure 3.3.7 where the first three principal components of natural soils and archaeological materials from Italy are very similar in their shape and differ only in small absorption features within a 100 nm range. For the 200 nm convolution, the D_{ratio} value drops to 1. This is still sufficient enough to detect some archaeological

materials but by far not as good as the D_{ratio} of the original spectra. This indicates that with a low resolution spectrometer, the D calculation becomes less accurate, but still provides useful values to identify archaeological materials.

Figure 4.9.4 shows the PCA results for the different smoothing levels of site 2. Like the result for site 1, the score plots of the first two PCs do not change strongly after broad smoothing kernels are applied. Only the displayed PCA results are orientated in a clockwise direction when higher smoothing kernels are used.

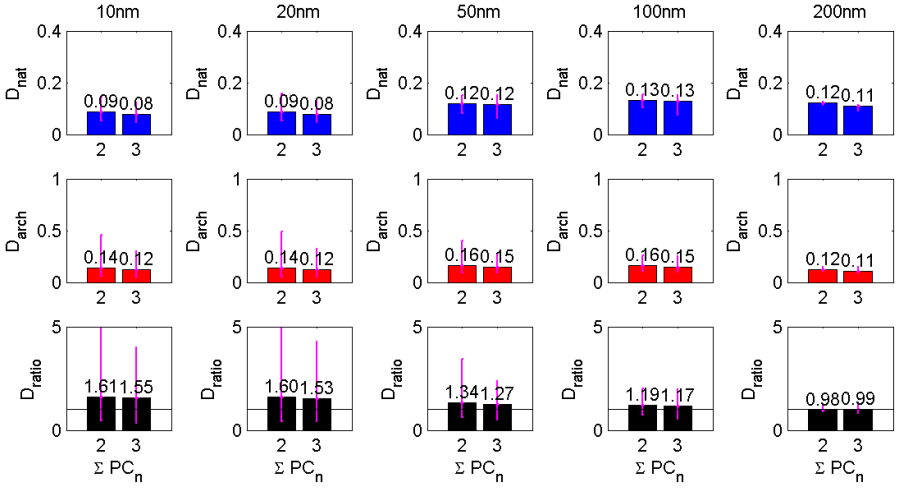


Figure 4.9.5 D values for site 2 where spectra are smoothed between 400 – 1000 nm with different convolution kernels (10 nm, 20 nm, 50 nm, 100 nm and 200 nm). Here, N_{soil} includes natural soils from the ISRIC spectral library.

Figure 4.9.5 shows the D values for the PCA performed in Figure 4.9.4 for site 2. Similar to the results from site 1 (Figure 4.9.3), the D_{nat} and D_{arch} values steadily increase as the spectra are smoothed. Unlike site 1, where there were no changes in the D values until the 50 nm convolution kernel, for site 2, one can observe a slight increase of D_{nat} and D_{arch}. For 10 nm and 20 nm kernels, the D_{ratio} is around 1.6, which is slightly higher than the original D_{ratio} value (1.59, refer to Figure 4.8.10). For smoothing kernels ≥ 50 nm, the D_{ratio} value decreases towards a D_{ratio} value less than 1 for the 200 nm kernel.

The smoothing results in this chapter showed that a degradation of the spectral resolution to values of up to about 50 nm does not lead to worse results compared to the results for the original spectra. This surprising outcome indicates that the spectral features relevant for the separation of archaeological material from natural soils are rather broad.

This is an important result, because it provides the possibility to use spectrometers with much lower spectral resolution than the ASD spectrometer used in this thesis. For such spectrometers either the temporal or spatial resolution (or both) of the measurements can be largely improved. Also, it gives the possibility of applying the method to either airborne or satellite images (if the spatial resolution is sufficient) even if the spectral resolution is not as fine as that of the ground-based spectrometers.

4.10 Summary

In Chapter 4, reflectance spectra of archaeological materials and natural soils were gathered from two archaeological sites in Calabria and analysed using different methods. Both sites contained a buried archaeological remain, which was assumed to be a prehistoric kitchen settlement. The excavated trenches contained kitchen formations of the settlements with soils of reddish colour, assumed to have been influenced by the archaeological materials (various potteries and burned materials). Around the archaeological pit, several natural soil pits were also excavated to obtain spectra of pure natural soils for that site. Spectra of these archaeological materials and natural soils were collected on the site (at the soil profiles) with the ASD FieldSpec Pro spectrometer using an artificial halogen light source.

Such buried archaeological materials or soils influenced by the archaeological activities showed a clear colour difference compared to the natural soils and thus probably represent different chemical compositions. However, the XRF results (Chapter 4.2.1) showed that there were no large differences found between the elemental compositions of natural soils surrounding the archaeological sites and the archaeological remains. The dominating elements (K, Ca, Ti and Fe) are identical for soils and archaeological materials, although there are some elements such as Sr and Ba which showed outstanding values for archaeological materials.

Because the XRF results did not show consistent patterns, the analysis method focused on the spectral analysis methods. Soil spectra contain various kinds of information including mineral content and their characteristic absorption bands. These absorption bands are emphasised by the continuum removal method where the intensity (which is greatly influenced by the moisture level) and small absorption features, which may have little influence on the overall spectra, are removed. Since the archaeological materials or soils were reddish in colour, the redness index (RI) gave high values for archaeological soils and materials (Chapter 4.2.2). The colour index (CI) was also investigated to analyse the wavelength information in a simple and robust way. However, although the CI is limited to only two wavelengths, it showed some interesting results. In particular, the 550 nm range, which is influenced by the haematite content (reddish colour), always showed higher CI values for archaeological materials regardless of what other wavelength range was used together. Also not only the visible range, but also the infrared region contains information where archaeological materials yield a higher CI value (Chapter 4.2.3). Besides of these interesting findings, it was concluded that the calculation of the CI is probably not the best way to quantify the entire spectral information of the recorded reflectance spectra.

Thus, to use the information of the whole wavelength range, the PCA method was applied to the spectra (Chapter 4.3). The results showed a distinct separation between archaeological materials and natural soils. Based on these findings, various wavelength ranges were tested to observe which wavelength range is the most suitable for archaeological feature detection (Chapter 4.4). Especially the infrared region was investigated in detail to determine whether it contains important information. It was, however, found that the 400 – 1000 nm range showed the most useful information, since the higher IR region was mainly dominated by strong water absorption bands. Further studies might investigate whether it is possible to extract useful information in that spectral range in arid regions.

Although the first three PCs contain more than 90% of the variance, score plots of the higher

PCs (up to PC4) were investigated (Chapter 4.5). Based on the score plots of PC2-PC3 or PC2-PC4 for site 1 it was possible to separate spectra of different types of archaeological materials (potteries and burned materials). However, the PC3-PC4 combination did not show any useful information. This indicates that from higher PCs alone a separation between archaeological materials and natural soils is difficult, but together with the first two PCs the use of higher PCs can add useful information.

The investigation of higher order PCs showed a good separation between different types of archaeological materials at site 1 but not at site 2. Therefore, a modified PC algorithm, the so called Base PC, was applied with the idea of using PC spectra from a site where the PCA results showed a good separation between archaeological materials and natural soils, in this case site 1 (Chapter 4.6).

The analysis so far was based on the visual interpretation of the score plots of the PCA, showing a cluster of natural soils and archaeological spectra which were not within this cluster. However, the definition of the boundaries of the cluster of natural soils is based on personal interpretation. Since the aim of the thesis is to develop a quantitative and universal methodology, further investigations were made to numerically express the degree of 'archaeological features' of the reflectance spectra. To do this, the D calculation method was developed based on the PCs of natural soils (Chapter 4.8). PCs of natural soils were used to recalculate spectra with characteristic features of natural soils. By doing so, the difference between the original spectrum and the recalculated spectrum shows some difference D if the original spectrum represents an archaeological feature. Using this D calculation method, average D values for archaeological materials (large) and natural soils (near to 0) were calculated and the ratios between them were investigated. It was expected that spectra of archaeological soils should give D_{ratio} values larger than 1. The analysis tried to define a universal D_{ratio} value which can be applied to any site by changing the two variables which influence the D value, the numbers of PCs added and the N_{soil} spectra used for the *NPC* calculation ($\sum_1^2$ PC and $\sum_1^3$ PC).

It was found that when N_{soil} does not include natural soils measured at the archaeological site, using up to the third PC for the D calculation produces higher D_{ratio} values (allowing a better separation of archaeological materials and natural soils). However, if natural soils from the archaeological sites are included, already $\sum_1^2$ PC is sufficient to achieve a good separation. Various N_{soil} groups were tested, where the main results focused on the local (from the site) and the global (ISRIC spectral library) natural soils. Figure 4.10.1 shows the resulting D values for local (N_{soil} from original site) and global (N_{soil} from ISRIC spectral library) natural soils used for the D calculation for sites 1 and 2, Italy. For both sites, N_{soil} from local soils always produced high D_{ratio} values (around 3 to 10, depending on the site). When N_{soil} is from global soils, lower D_{ratio} values (1.5 to 2) were obtained. Surprisingly, when global N_{soil} is used, the problem of false detection (natural soils having D values larger than 1) is reduced (refer to Appendix 3.2 and 3.3). These results indicate that the D value method using global N_{soil} has the potential to be used universally for any archaeological sites. Here it is interesting to note that the N_{soil} did not contain any natural soil spectra from the site investigated but still gave reasonably high D_{ratio} values of about 1.5 for archaeological materials. However, these results are only found for the sites in Calabria, Italy and more investigations are needed to confirm the applicability of the D value method to other sites.

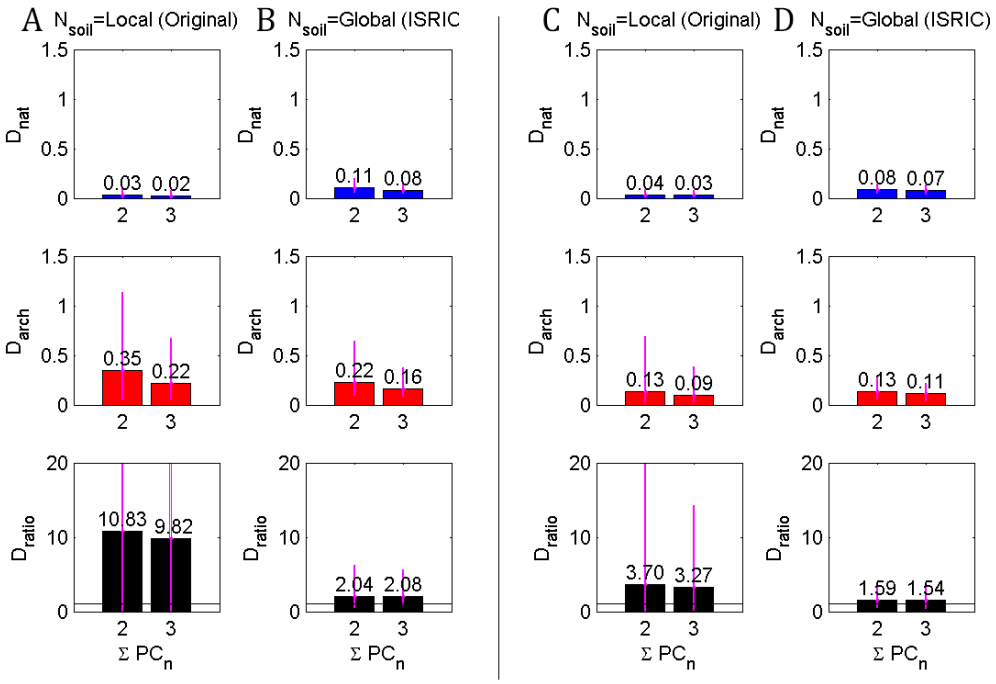


Figure 4.10.1 Final D value results for sites 1 and 2, Italy. Figures A and B show results for site 1, and figures C and D for site 2. Spectra are resampled to a ten-nanometre grid.

This D calculation was repeated with smoothed spectra to test the level of minimum spectral resolution for identifying large enough D values for archaeological remains. Using the Gaussian smoothing kernel, the results showed that the D_{ratio} is higher than the original spectra until the Gaussian kernel of 50 nm. Even beyond this limit, buried remains can still be identified. This indicates that the spectral resolution is not very important in identifying the D values for buried remains since most of the dominating features are broadband features.

This chapter presents the analysis results from the archaeological sites in the Sárvíz valley, Hungary. The geological and topological environment of this site is extremely different from Calabria, Italy. While the data from Italy represents archaeological remains in a complex environment, the data from Hungary represents archaeological remains in a homogenous environment.

In the previous chapter, various methods to separate archaeological features from natural soil were investigated for spectra gathered from Italy. Among these methods, the PCA result gave the most promising results, and a further development of the PCA method provided a quantitative way (D value calculation) of identifying archaeological materials when a set of known natural soil spectra data from the same site exists. In this chapter, the PCA and D-calculation method are applied to spectra recorded in Hungary. The main aim is to verify whether this method also works for buried remains in a totally different environment.

Chapter 5 Sárvíz Valley, Hungary

5.1 Site information

5.1.1 Location and basic properties

The Sárvíz valley is located in the Northern part of the Mezőföld region in the flat Great Hungarian Plain (southwest from Budapest). The valley extends over nearly 100 km from the valley of the Danube with a total area of about 60000 ha (Evelpidou et al., 2010). The valley is located on lowland with small hills formed by loess and wetland environment. The word Sárvíz means “muddy water”, which well describes many parts of the area. Unlike the sites in Calabria, Italy, which had complex soil distributions due to the mountainous landscape and the Mediterranean climate, the sites in the Sárvíz valley are thickly covered by loess (up to several meters in some regions), which is homogenously spread throughout the flat landscape. The valley is predominantly used for agriculture and intensive ploughing occurs throughout the year. This results in large amounts of potteries and archaeological artefacts being transported to the soil surface. Also, due to various water levels, the area has shown diverse settlement remains since the Neolithic, which revealed characteristic spatial patterns of the archaeological sites. The locations of the sites are shown in Figure 5.1.1.

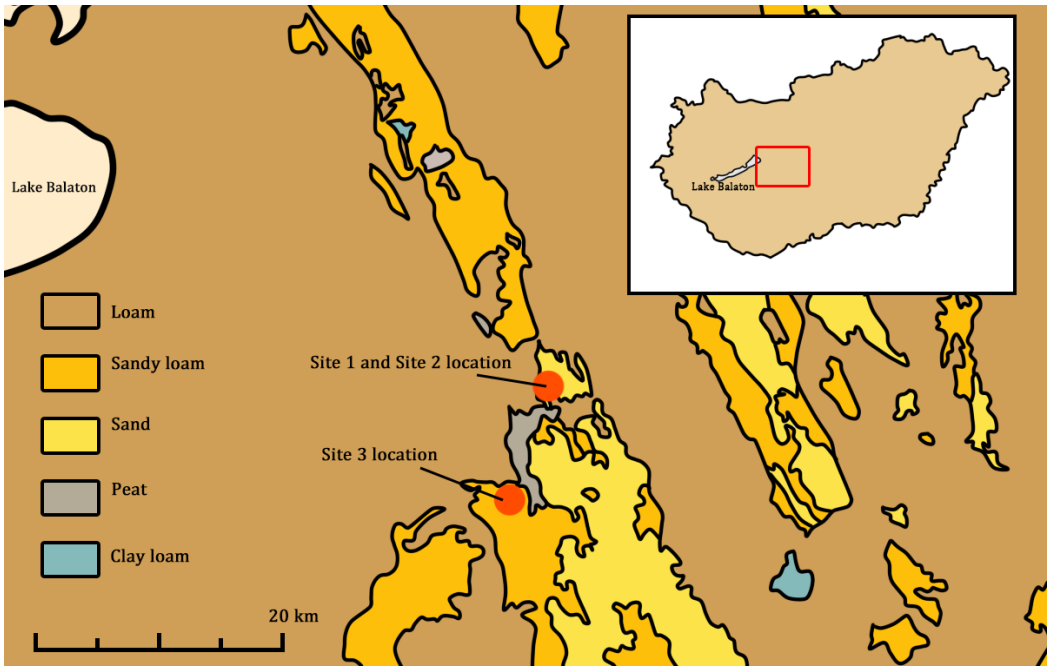


Figure 5.1.1 Map of the study region, Sárvíz valley, Hungary (modified from Várallya, 1989).

The field work took place in April 2013, just after the snow melt and first ploughing period of the region. This resulted in visible soil mark features and archaeological materials on the ground surface, but also an extremely muddy environment. Some parts of the field were still covered by snow due to the cold extensive winter. Such conditions were similar to the site condition of Italy (refer to Chapter 4.1.1) where the field measurements were made after the rain. Within the valley, three specific sites were chosen for soil analyses and spectral measurements. The field work took place together with a soil science team and an archaeological team from the Hungarian National Museum.

Figure 5.1.2 Images of the ground conditions of the sites in Hungary. The ground is extremely muddy due to recent snow melt. The sites are subject to extensive ploughing and therefore large pieces of potteries are scattered on the top soil as shown in the left figure.

5.1.2 Soil profiles

According to Figure 3.1.1, three archaeological sites were excavated in Hungary. Initially, excavations at four archaeological sites were planned, but one site turned out to be a natural soil feature which was misinterpreted as an archaeological feature. Sites 1 and 2 are located on a flat homogeneous farm land and site 3 is located on a dune crest and, therefore, subjected to strong erosion.

Unfortunately, unlike in Italy (Chapter 4.1.2), no additional pits were made for natural soil measurements. Since the measurements were made on agricultural land, only a limited number of soil pits was allowed by the farmers. To obtain as much spectral information of the soils as possible, these pits were dug at the boundaries of the buried archaeological features where part of the soil profile contains the archaeological stratum, and the other part contains the natural soil profile (refer to Figure 5.1.4 to Figure 5.1.6). Luckily, the field had been nicely ploughed before the research team arrived to the site, therefore, clear soil marks were visible on the ground surface (Figure 5.1.3), which made it possible to identify the 'boundaries' of buried archaeological features. For each pit (site), two soil profiles were obtained, one containing the archaeological profile and the other containing the natural soil profile.

Figure 5.1.3 Soil mark of an archaeological pit feature visible on the ground surface due to the colour difference between the archaeological pit and the surrounding natural soil from site 2.

At the Alap region, two archaeological sites (site 1 and 2) were excavated. This region is located in a fairly flat landscape where the soil formation is uniform and homogeneous. Within the same farm yard (Alap region), two archaeological pits were made. Each pit contained one natural soil profile and one archaeological soil profile. These soil profiles are measured from the same pit, but depending on whether the profile contains an archaeological feature or not, the profile was assigned as an archaeological profile or a natural soil profile (refer to Figures 5.1.4 and 5.1.4 for visual explanation of the soil profiles). There is the danger that such measurement might not represent a pure archaeological or natural soil profile, but it was the best way to make optimum use of the measurements with limited number of soil pits.

(1) Site 1

Site 1 contains a ditch feature which is assumed to be filled with fallen wall structures (Figure 5.1.4). This pit contains two profiles, one archaeological profile containing the ditch formation and one natural soil profile located on the other side of the pit (Figure 5.1.4). On the floor of the pit, a clear ditch feature (Figure 5.1.4.B.2) is indicated by the bright coloured patches, which are assumed to be part of the fallen wall structures within the ditch formation. Also, on the side of the soil profile (Figure 5.1.4.B.1), some pieces of pottery were found. The background soil of the ditch formation (floor of the pit) looked similar in colour to the surrounding natural soils, but these background soils are also treated as archaeological soils (AS) since they are part of the ditch formation. A detailed explanation of these 'background soils' within the ditch formation is given in Chapter 5.1.3. Next to the archaeological profile is the natural soil profile without any archaeological remains buried beneath. This natural soil profile showed a similar Ap horizon thickness of around 30 cm depth, and the colour of the soil got darker with depth (Figure 5.1.4.C).

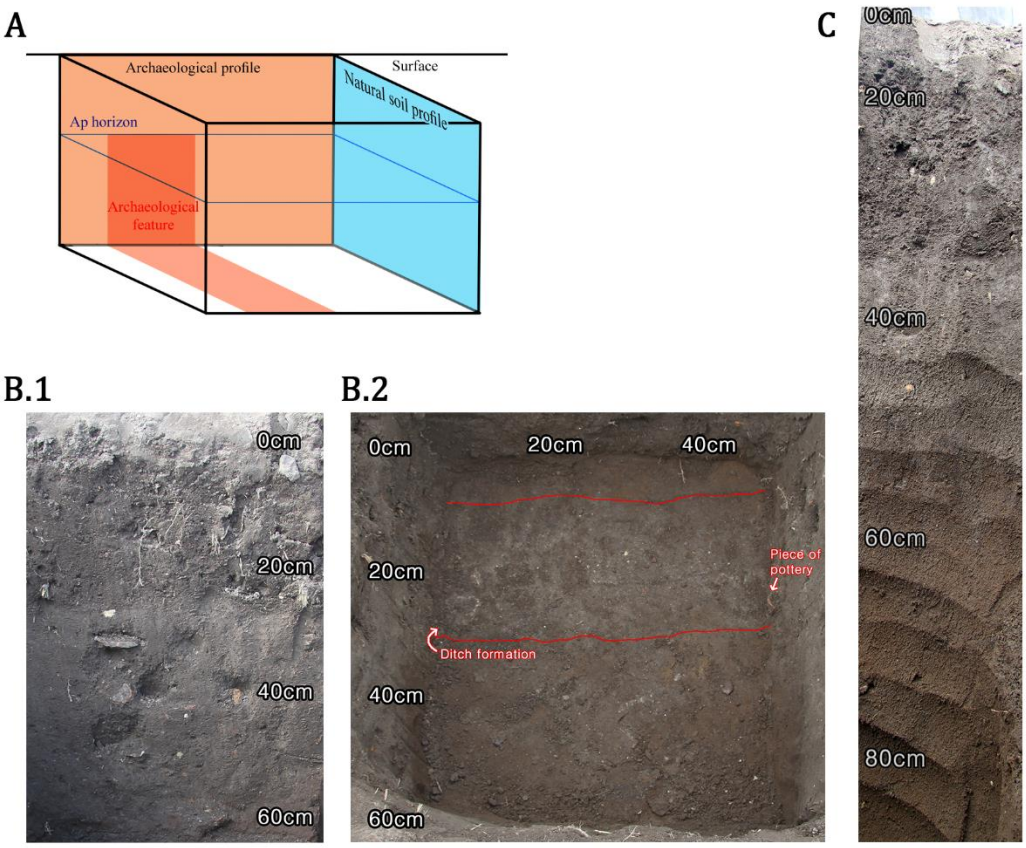


Figure 5.1.4 A: Schematic illustration of the pit at site 1, Hungary and location of the soil profiles. B.1: Archaeological soil profile (ditch formation). B.2: Floor of pit showing the ditch formation. C: Natural soil profile of the pit.

(2) Site 2

Site 2 was excavated at the border of a soil mark shown in Figure 5.1.3, where a clear colour difference between the archaeological feature and the surrounding natural soils can be observed. This site is located around 50 m away from site 1, in the same field. The vertical soil profile in this pit contains both archaeological features and natural soil features (Figure 5.1.5). This pit feature is assumed to be an ancient rubbish dump, which contained a high fraction of organic matters and some wooden materials. The right side of the profile is the archaeological pit feature (with its boundary clearly seen by the darker soil colour) and the left side of the feature is the natural soil part.

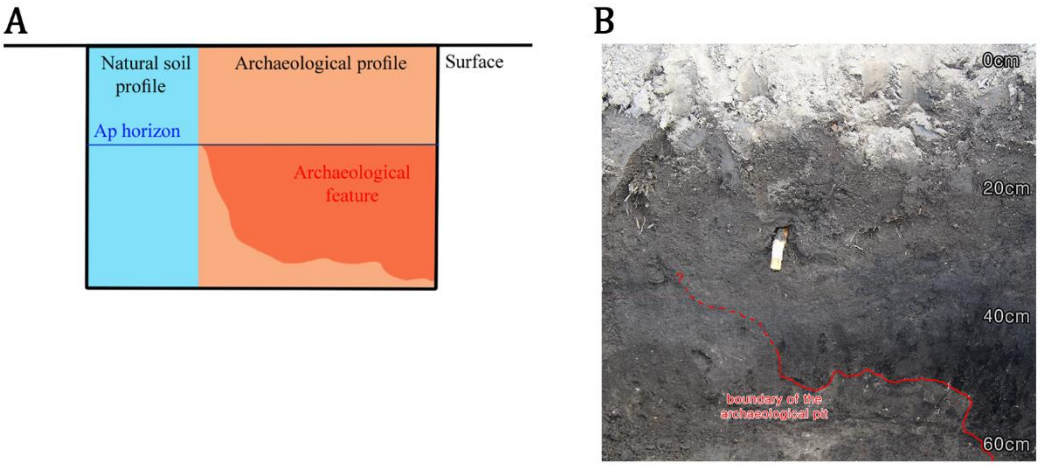


Figure 5.1.5 A: Schematic illustration of the pit at site 2, Hungary and location of the soil profiles. B: Soil profile, where the left side of the image shows the natural soil profile and the right side the ancient pit formation and the archaeological soil profile.

(3) Site 3

In the Cece region, also two 'archaeological sites' were excavated. However, only one site contained an archaeological remain (site 3) and the other site was a natural soil pit. The natural soil pit was intended to be located on an archaeological feature indicated by the soil mark feature on the surface. However, this soil mark was just a geological pattern shown at the ground surface but not a real soil mark (unfortunately, no photo image of this site was taken before the excavation). This is one of the problems discussed in Chapter 2 where trained archaeologists may misinterpret soil marks. Until the soil mark is really excavated to reveal the buried remain, it is difficult to justify whether it is a 'real' soil mark (unless it has a shape of an obvious man made feature, e.g. an exact rectangle). Soil marks can be identified by archaeologist, but this strongly depends on the archaeologist's personal experience as well as the shape. The misinterpretation of the soil mark again emphasises the importance of developing a method using spectroscopy to separate archaeological remains from natural soils in a quantitative way.

Site 3 is located on a hilly site created by ancient sand dunes, where the archaeological site is

located on the crest of the dune and a natural soil pit is located at the trough. Figure 5.1.6 shows the soil profiles of these pits and also the floor of the archaeological pit. Since the site is located on the crest, it is highly eroded due to transportation of material. The archaeological remain contained a ditch formation which was filled with bright coloured soil (beige colour) and, thus, showed a clear boundary between the ditch formation and the surrounding natural soil. However, in this ditch, no archaeological artefacts were found. The background soil of the ditch was beige to brown in colour with some patches of yellow and grey soil within. Here, the top soil (Ap) did not show a clear soil colour difference between the archaeological part and the natural soil part. The natural soil pit (Figure 5.1.6.B) is located on the trough where transported materials accumulate. Similar to the archaeological pit, the Ap horizon extended to 25 cm in depth. Below this horizon was a transition zone until a depth of 70 cm and below the parent soil (C horizon).

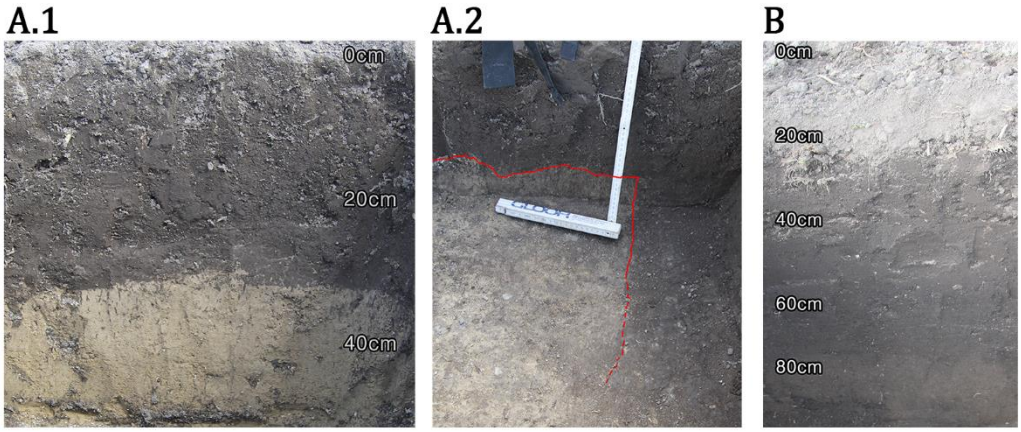


Figure 5.1.6 Profile images of site 3, Hungary. Figure A shows the archaeological soil profile and figure B the natural soil profile. A.1: Soil profile of site which contains an archaeological ditch formation (looking at the uphill direction). A.2: Floor of the pit showing the ditch formation (looking at the downhill direction).

5.1.3 Archaeological materials

The archaeological materials investigated in the Sárvíz valley, Hungary, are different from those at the sites in Calabria, Italy. The remains found at the Calabrian sites are mainly potteries and kitchen formation where the archaeological soils tend to be more reddish in colour and contained some burned materials within. On the other hand, the archaeological formations found in the Sárvíz valley are pit and ditch formations which showed clear boundaries of the formations indicated by soil mark features on the ground surface.

Ancient pit features are areas of land dug by humans in the past for food storage, waste disposal and etc. They tend to have a circular shape, and due to their usage, they contain highly organic material and generally have a dark soil colour. When the pit is undisturbed by humans, the surrounding soils erode and fall into the pit filling it with natural soils, which are gradually influenced by the high organic matters within the pit.

Similarly, ditches were mainly dug outside the walls of fortresses or building structures for defensive usage. Therefore they have an elongated shape and were often filled with water. When a ditch feature is left undisturbed by human activities, the walls break and fall down into the ditch together with the surrounding soils. Therefore, ditch features are filled with the natural soils from the surrounding together with parts of wall structures or pieces of broken potteries. These ditch formations in Hungary normally contain background soils, which are the natural soils transported into the ditch but are assumed to be archaeological soils, filled with pieces of wall structures which are indicated by bright colour patches (also regarded as archaeological soils). Figure 5.1.7 illustrates the formation of a pit or ditch feature. Both features are created in a similar process, but due to their different usage, pit features tend to contain more organic matters.

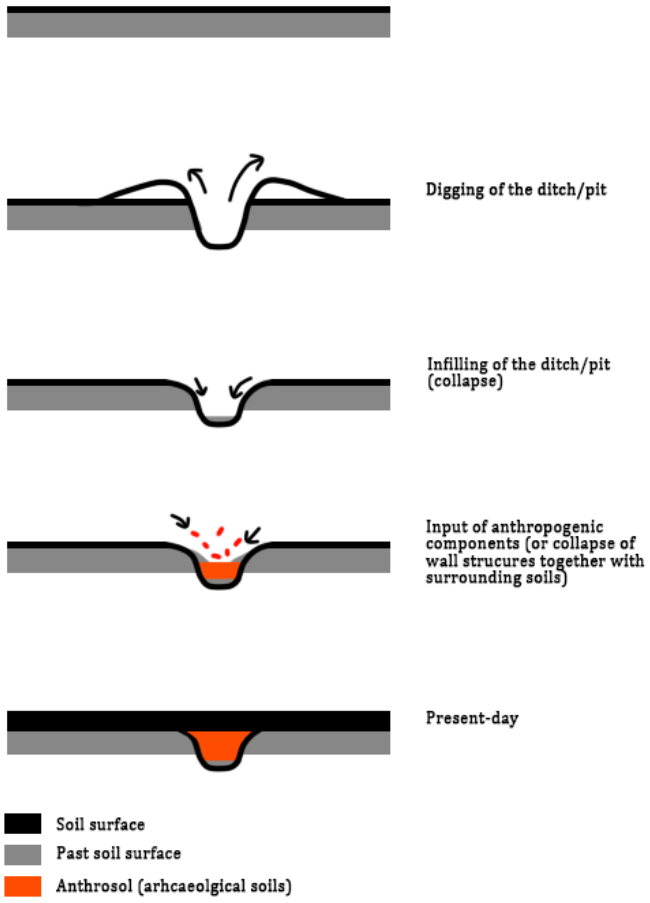


Figure 5.1.7 Schematic formation process of a ditch (pit) feature (modified from Brancier et al., 2014). The ditches (pits) are totally refilled with soil and the feature is buried under the Ap horizon.

5.2 Soil analysis

5.2.1 X-ray Fluorescence (XRF)

Similar to the measurements in Italy, the XRF measurements of the soil samples gathered from the sites in Hungary are investigated (refer to Chapter 3.1.1 for details on the sample and

Chapter 3.4.2 for details on the XRF measurements). The results of the XRF measurements of these soils were averaged for each horizon. These measurements provide information on whether the compositions of the elements change between the natural soils and the archaeological material. Also, from the comparison of the specific composition of the elements, it might be possible to observe which element is dominant for the archaeological materials at these sites. However, since the soils in this region are known to be fairly homogenous, the XRF values for the natural soils might be similar to the archaeological materials.

Similar to the soils in Italy, the dominant elements in Hungary are potassium (K), calcium (Ca), titanium (Ti), manganese (Mn) and iron (Fe), refer to Appendix 4 and 5 for figures illustrating the dominant elements and the exact XRF values.

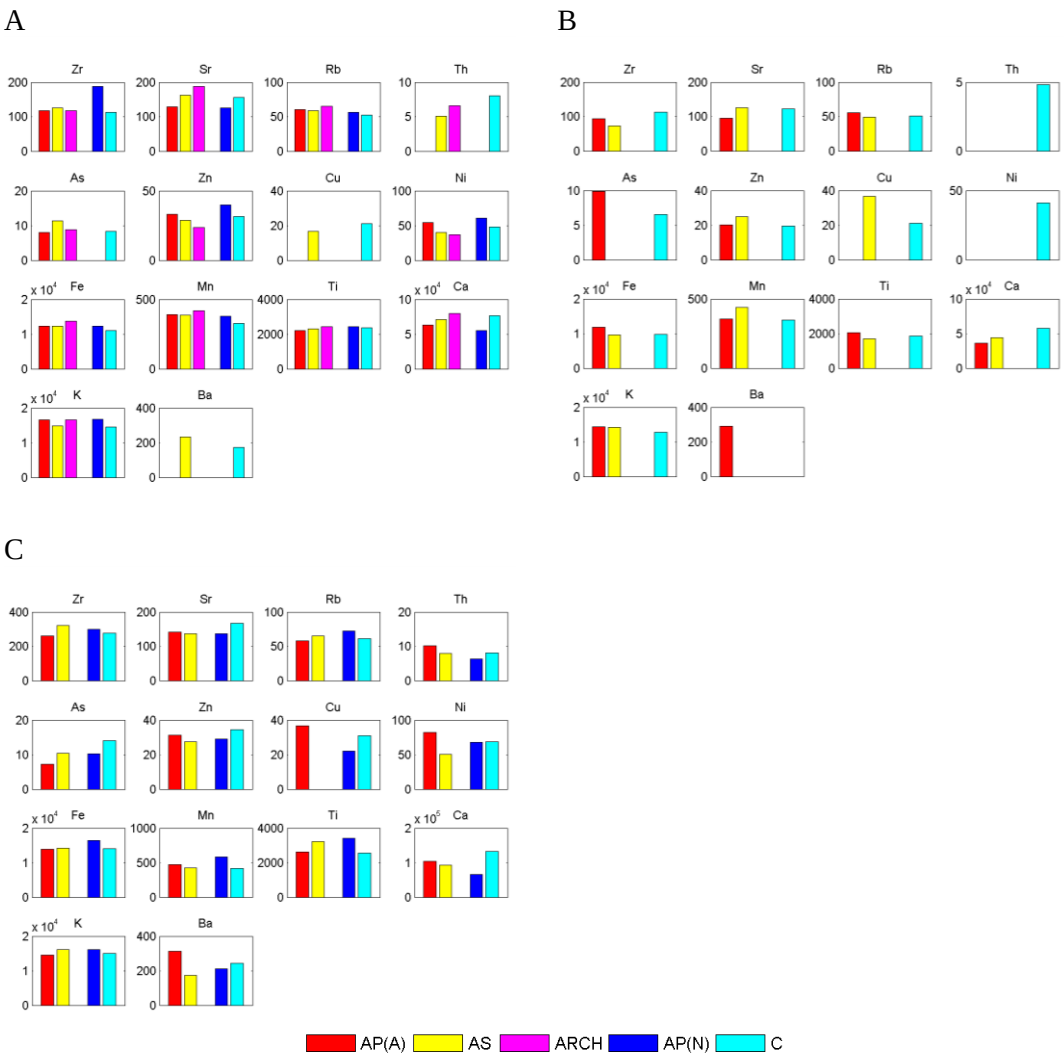


Figure 5.2.1 XRF results for (A) site 1, (B) site 2 and (C) site 3. Bar plots show the mixing ratios of the different elements (Zr, Sr, Rb, Th, As, Zn, Cu, Ni, Fe, Mn, Ti, Ca, K and Ba in parts per million) in soil features from the Ap horizon above the archaeological remains (red), averaged archaeological soils (yellow), archaeological remains (magenta), Ap of natural soils (blue) and averages of natural soils (cyan). For the exact XRF values refer to Appendix 5.

Figure 5.2.1 show the XRF results for sites 1, 2 and 3. Note that at these sites, clear colour differences between the Ap soils, the buried archaeological remains and natural soils were observed by bare-eye (Figures 5.1.4 to 5.1.6). However, these differences are not clearly observed in the XRF results. The dominant calcium (Ca) amount found in the archaeological soils in Italy (Figure 4.2.2) is not even detectable in the Hungarian soils.

Figure 5.2.1.A shows the XRF results for site 1 where each bar plot represents different soil horizons including natural and archaeological soils. These results clearly indicate that there are no significant differences between the elemental compositions of archaeological and natural soils.

Figure 5.2.1.B shows the XRF results for site 2. Here the natural soils seem to have dominant thorium (Th) and nickel (Ni) amounts. However, this value is similar to the Th and Ni content for the archaeological materials at site 1 and, therefore, does not represent a characteristic element of natural soils in Hungary. Figure 5.2.1.C represents the XRF results for site 3. Similarly, at this site, no significant difference was observed between archaeological and natural soils.

The overall XRF results for the archaeological sites in Hungary show that the soils are extremely homogenous, even with the archaeological soils.

5.3 Principal Component Analysis (PCA)

The principal component analysis (PCA) is applied to the spectra gathered in Hungary in order to separate archaeological features from those the surrounding natural soils. The archaeological features found at these sites are mainly pit or ditch formations which show characteristic soil colours with only few artefacts (ARCH). Since archaeological soils (AS) are the main archaeological materials at the sites in Hungary, the PCA is focused on separating archaeological ‘soils’ from a homogeneous background natural soil.

Figure 5.3.1 shows the score plots of the first two PCs for the sites in Hungary. The PCA was applied over the whole spectrometer range (400 – 2400 nm). Different soil horizons are represented in different colours (refer to Chapter 3.1.1). Note that there are no archaeological artefacts (burned materials or potteries) found at site 3 (Chapter 5.1.2). All three sites contain archaeological features, and, depending on what had fallen into these formations, the type of ‘archaeological soil’ varies. All three sites show that the first two PCs already account for more than 90% of the variation of the reflectance spectra, indicating that the soil spectra gathered are rather homogeneous.

The score plot of site 1 (Figure 5.3.1.A) shows that the archaeological soils are clustered according to the soil colours. Site 1 is a ditch formation filled with fallen wall pieces and therefore has a darker (brownish) background archaeological soil and patches of beige coloured soils. These beige coloured archaeological soil spectra are grouped in the right corner of the plot (the yellow marker points within the red circle), which have the highest contribution of the first PC (PC1 value of 5). The archaeological artefacts (such as potteries) showed PC1 values around 3 and are separated from the natural soil cluster. However, the burned materials are mixed in the natural soil cluster. This is different from the results at site 1, Calabria, where

the burned materials were strongly isolated from the natural soil cluster (Figure 4.3.1). Most of the natural soils (from both the C and Ap horizons) are located around 0 for PC1 (although their PC2 value varies) and some AS spectra are included within this cluster.

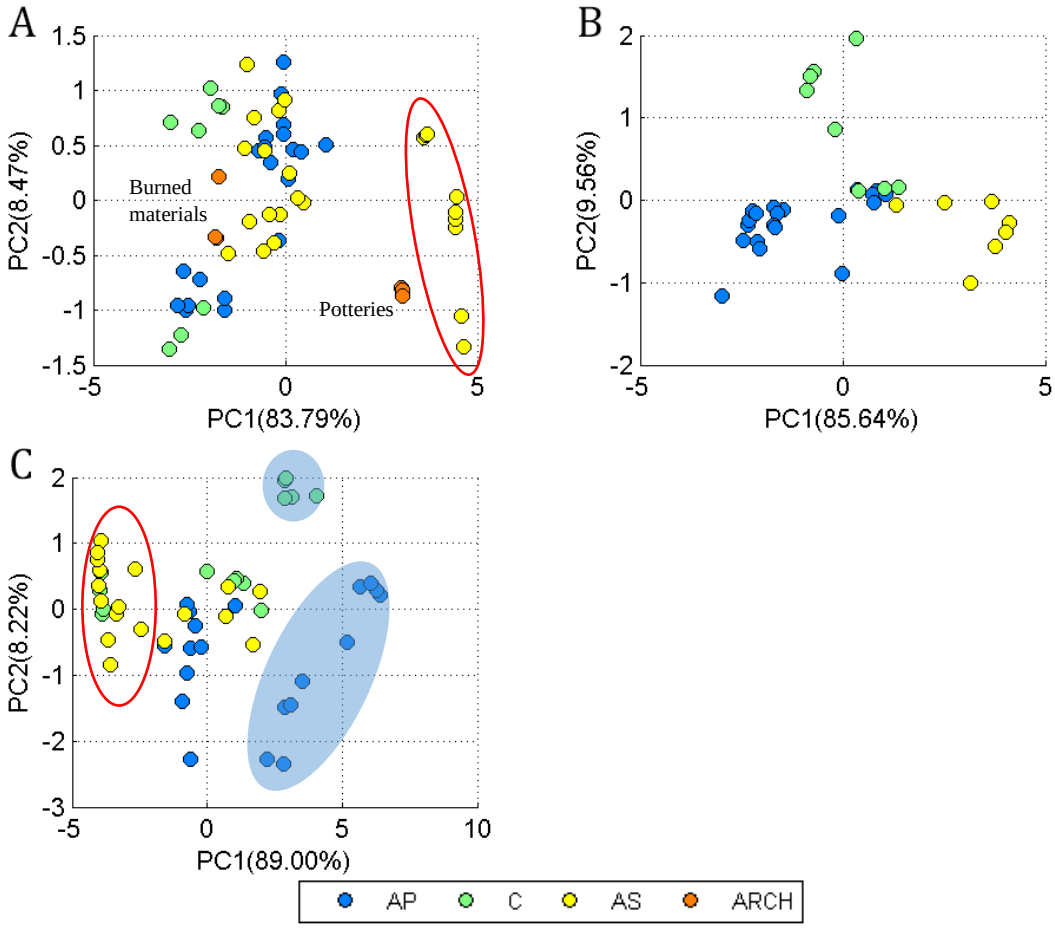


Figure 5.3.1 Results of the PCA analysis (score plot of the PC1 against PC2) obtained for the wavelength range 400 – 2400 nm for site 1 (A), site 2 (B) and site 3 (C). At site 1 (A), points within the red circle represent the archaeological soils (AS) which show a bright beige colour compared to other AS.
At site 2 (B), the results show a clear difference between different soil horizons.
At site 3 (C), the points within the blue circles represent the archaeological soils (AS) which shows yellow to grey colour.

For site 2 (Figure 5.3.1.B), the PCA result shows that the soil spectra are grouped according to their soil horizons. In this site, one could observe the soil mark feature on the ground surface which showed that the Ap soils above the archaeological remain has different colours compared to the surrounding Ap soils. However, in Figure 5.3.1.B, there are no big differences between the Ap soils (they are all gathered together into the natural soil cluster) and thus the soil colour difference is not the main factor determining the PCA result.

Figure 5.3.1.C shows the score plot of the first two PCs for site 3 for a wavelength range 400 – 2400 nm. The spectra gathered from the natural soil pit are within the blue circle, and these spectra have PC1 values of more than 3 while the other natural soils (from the archaeological

profile) have PC1 values around 0. The natural soil profile spectra are centred on PC1 values of 0 and some spectra of the background soil of the ditch (dark soil coloured AS) are within this cluster. Figure 5.1.6 shows the image of the profile from two different viewing directions: one side of the profile shows clear beige coloured archaeological soil and the other side of the profile shows beige archaeological soil mixed together with darker soils. Therefore, the background soil colour of the ditch varies from beige to darker brown. This difference is also represented in the PCA score plot where the bright AS is in the left side of the plot and the dark AS is near the natural soil cluster (in other words, brighter soils have a larger negative PC1 and for darker soil colours the PC1 value decreases towards 0). This can be related to the analysis made in Chapter 4.2.5 where the mixing of natural soil with archaeological material influenced the overall spectra indicating that a high coverage of archaeology is needed to show its spectral characteristic.

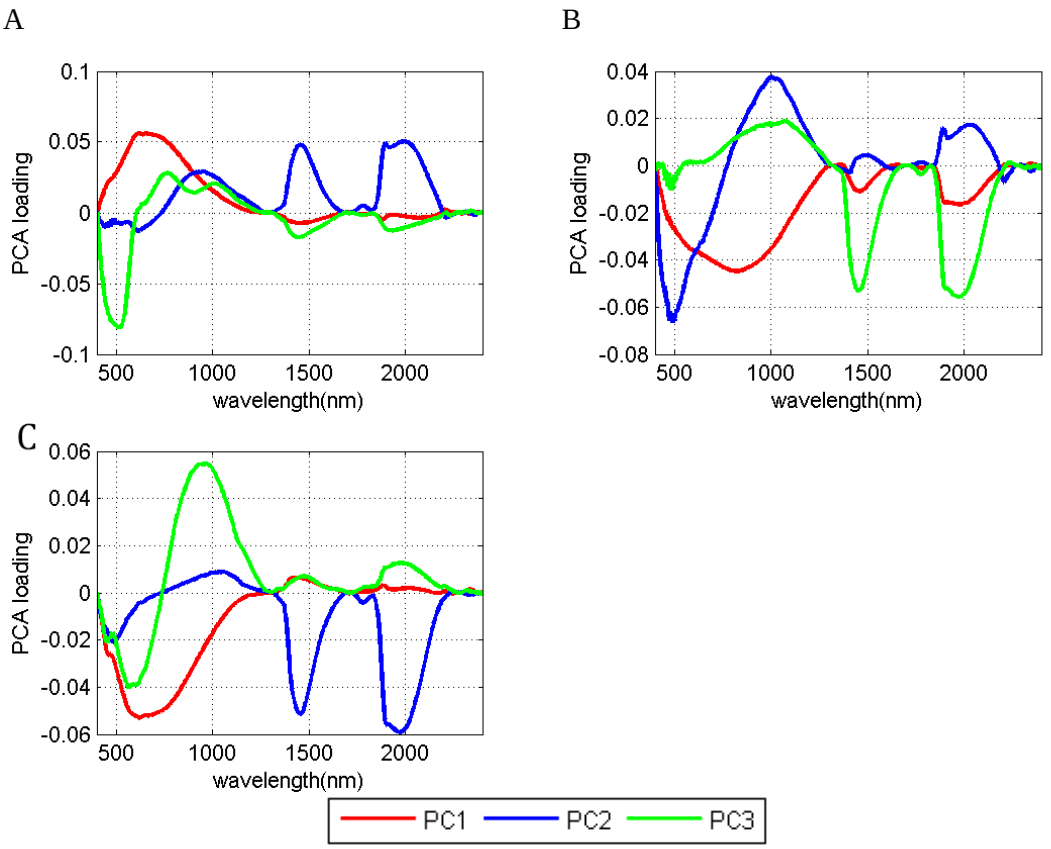


Figure 5.3.2 400 to 2400 nm spectra of the first, second and third principal component (PC1, PC2 and PC3) for the sites 1 (A), 2 (B) and 3 (C).

Figure 5.3.2 shows the spectra for the first three PCs. For all three sites, the spectral features in the NIR region are highly dominated by the water absorption bands. In the visible range (for all three sites), the PC1 is dominated by a large broadband feature. This feature is similar to the PC1 feature of site 1, Italy (Figure 4.3.2). Also, the strong absorption peak at 550 nm is observed in the PC2 or PC3 of all three sites.

At the Calabrian sites, Italy (sites 1 and 2), no clear difference of the PCA results for the natural soil horizons was found (Figure 4.4.1.A and 4.4.2.A). However, in the Sárvíz valley, Hun-

gary, the soil horizons and pits were all well separated. This is an unexpected result because the Hungarian soils were assumed to be rather uniform compared to the Italian soils.

5.4 PCA results for different wavelength ranges

The results in Chapter 4.4 showed that 400 – 1000 nm is the best suited wavelength range for separating archaeological materials from natural soils for the archaeological sites in Italy. The NIR region (beyond 1000 nm) is dominated by strong water absorption bands making it difficult to observe other spectral features. The Italian results also indicated that the colour of the soil is not the main factor that determines the PCA of soil spectra. Figures 5.4.1 to 5.4.3 show the PCA score plots of the first two PCs for various wavelength ranges for sites 1 to 3. For a better comparison, only results for NAT (including all natural soils), ARCH (archaeological artefacts) and AS (archaeological soils) are shown instead of the individual natural soil horizons. Based on the results from Italy (Chapter 4.4), the PCA is applied for a wavelength range between 400 – 1000 nm and then extended to further wavelength ranges according to the findings in Chapter 4.4.

Figure 5.4.1 shows PCA results for various wavelength ranges for site 1. The first impression of the results is that some AS and ARCH are separated, but most of these archaeological spectra are mixed within the natural soil cluster. The AS spectra which are within the natural soil cluster are mainly background soils of the ditch formation. Data points for potteries (ARCH) are mostly isolated from the other soil spectra for all wavelength ranges (below 2000 nm) indicating that the spectral features of pottery are systematically different from soil spectra gathered in Hungary (note that the pottery spectra in Italy were not always separated, refer to Chapter 4.4). Here, the data points of the light coloured archaeological soils are shifted to the left side with PC1 values between -3 and -4, but the burned materials and AS (background of the ditch formation) are located within the natural soil cluster.

The PCA results for the 400 – 1000 nm wavelength range are similar to the PCA results for 400 – 2400 nm (the whole spectral range). This indicates that, most of the spectral information which separates the archaeology from natural soils is contained within the first 1000 nm. However, to make sure that one does not lose important spectral information, also PCA results for other wavelength ranges are presented (Figure 5.4.1.B to 5.4.1.I). For the 400 – 1300 nm range (Figure 5.4.1.B to 5.4.1.D), only the bright AS and potteries are separated from the natural soil cluster. PCA results for 1300 – 2200 nm (Figure 5.4.1.E to 5.4.1.I) show mixtures between spectra of archaeological soils and natural soils.

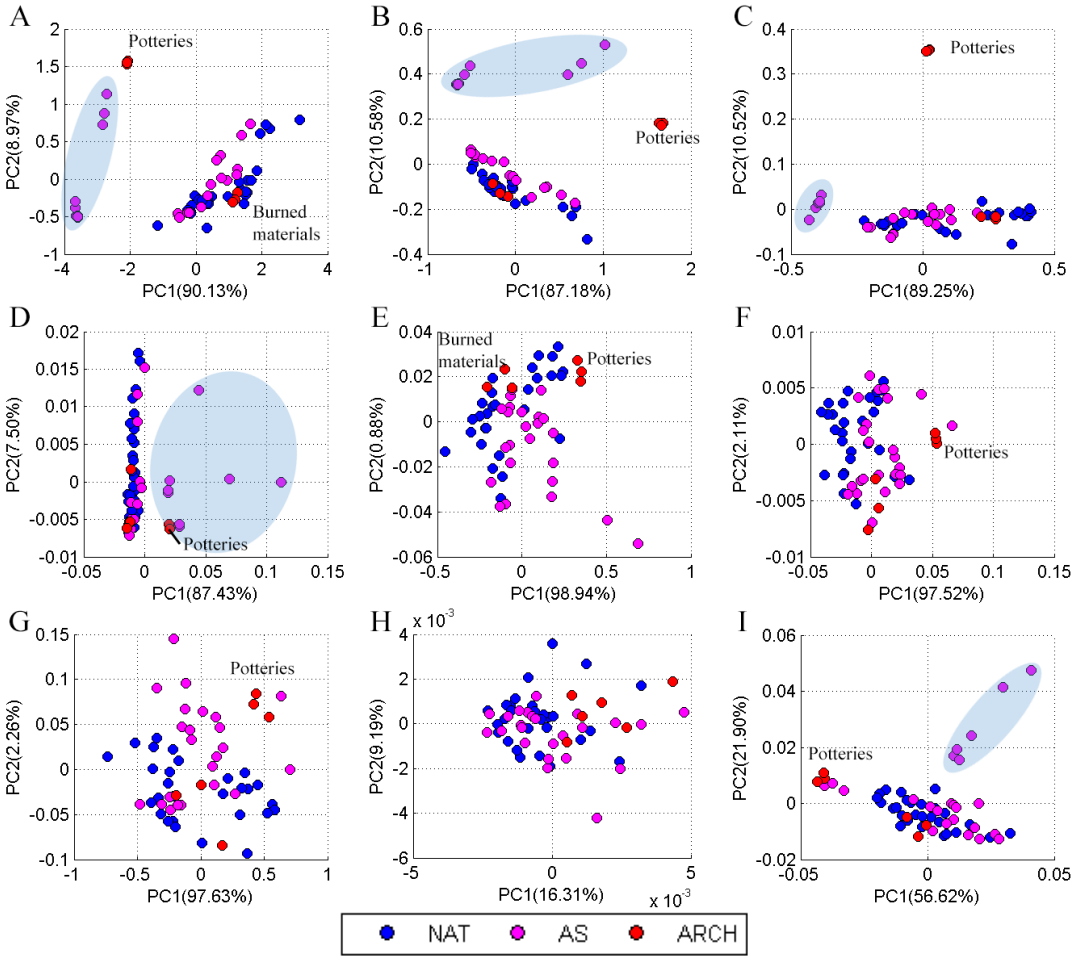


Figure 5.4.1 PC1-PC2 score plots for site 1 in Hungary for different wavelength ranges. A: 400 – 1000 nm, B: 400 – 700 nm (visible range), C: 700 – 1000 nm, D: 1000 – 1300 nm, E: 1300 – 1500 nm, F: 1500 – 1830 nm, G: 1830 – 2000 nm, H: 2000 – 2200 nm and I: 2200 – 2400 nm. NAT represents the spectra of natural soils (including Ap, Bt and C), AS the spectra of archaeological soils and ARCH the spectra of archaeological artefacts. Data points within the blue circles represent beige coloured AS (wall materials fallen into the ditch formation).

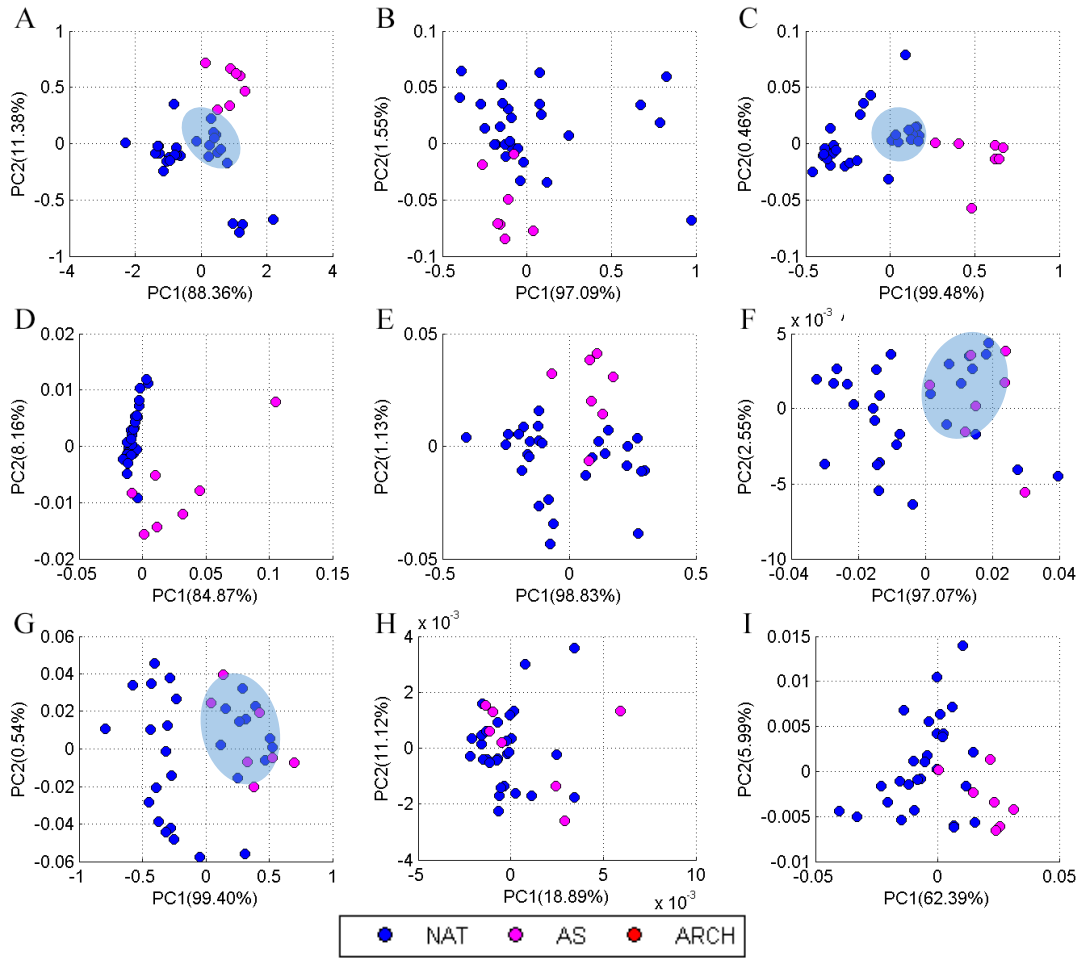


Figure 5.4.2 PC1-PC2 score plots for site 2 in Hungary for different wavelength ranges. A: 400 – 1000 nm, B: 400 – 700 nm (visible range), C: 700 – 1000 nm, D: 1000 – 1300 nm, E: 1300 – 1500 nm, F: 1500 – 1830 nm, G: 1830 – 2000 nm, H: 2000 – 2200 nm and I: 2200 – 2400 nm. NAT represents the spectra of natural soils (including Ap, Bt and C), AS represents the spectra of archaeological soils and ARCH the spectra of archaeological artefacts. Natural soil spectra within the blue circles are recorded below the archaeological pit.

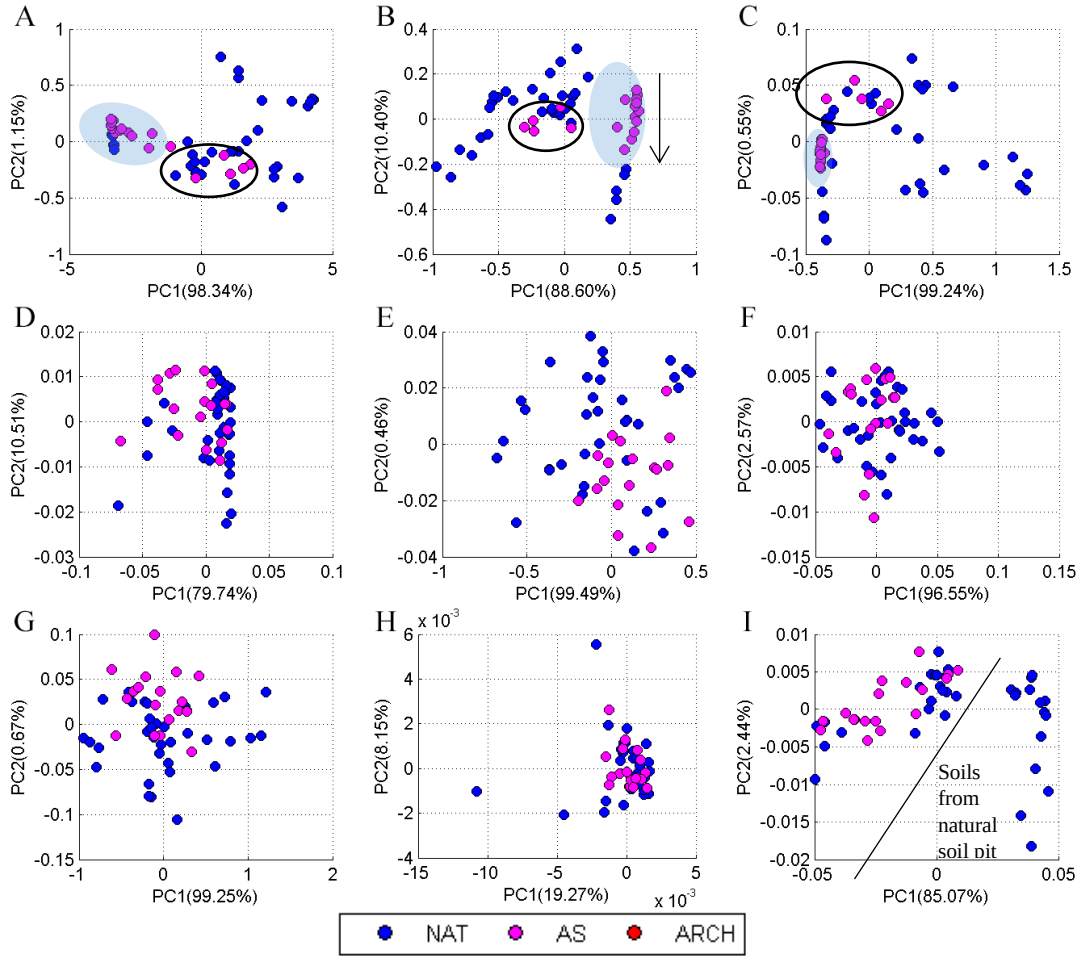


Figure 5.4.3 PC1-PC2 score plots for site 3 in Hungary for different wavelength ranges. A: 400 – 1000 nm, B: 400 – 700 nm (visible range), C: 700 – 1000 nm, D: 1000 – 1300 nm, E: 1300 – 1500 nm, F: 1500 – 1830 nm, G: 1830 – 2000 nm, H: 2000 – 2200 nm and I: 2200 – 2400 nm. NAT represents the spectra of natural soils (including Ap, Bt and C), AS represents the spectra of archaeological soils and ARCH the spectra of archaeological artefacts. Data points of AS within the blue circles represent beige coloured AS. Data points of AS, within the black circles, represent dark archaeological soils (see Figure 5.1.6.A.2).

The PCA results for site 2 (Figure 5.4.2) are similar to those for site 1. The archaeological pit (AS) soils are well separated from the natural soil cluster for spectral ranges up to 1300 nm, but beyond this range the archaeological soils are mixed into the natural soil spectra. One interesting finding at this site is that the natural soils, which were measured just below the archaeological pit (Figure 5.1.5), are sometimes mixed within the AS spectra (points inside the blue circle) in high wavelength regions (Figures 5.4.2.F and 5.4.2.I). This indicates that some minerals within the archaeological pits were probably transported down to the natural soil.

Figure 5.4.3 shows the results of the PCA for site 3. Note that no archaeological artefacts were found at this site and, therefore, only AS measurements were made. For The wavelength range of 400 – 1000 nm, the light coloured archaeological soils are located at the left side with PC1 values around -4 to -2. Other archaeological soils are mixed together with the natural soils. PCA results for the visible range (400 – 700 nm) also do not improve the result dra-

matically compared to the 400 – 2400 nm (Figure 5.3.1) or 400 – 1000 nm region (Figure 5.4.3.A). Here, light coloured AS have PC1 values of around 0.5 (right side), but also the natural soils (from C horizon) have PC1 values of 0.5. For wavelength ranges above 700 nm the PCA results indicate that it is difficult to distinguish between archaeological spectra and natural soil spectra for site 3.

Based on the PCA results of the three archaeological sites in Hungary, the wavelength of 400 – 1000 nm is also used for the following analyses. One thing to notice is that the PCA results, regardless of the wavelength range, always showed a rather high PC1 variance. Except for some cases, PC1 had a variance of more than 90% indicating that the soil spectra taken in Hungary are highly homogeneous.

5.5 Results for higher order PCs

The PCA analysis for the selected wavelength ranges did not substantially improve the result compared to the full range analysis (400 – 2400 nm), which is different from the results in Italy where different wavelength ranges provided different possibilities (for example the 2000 – 2200 nm region in Figure 4.4.1.G). The results indicate that the Hungarian soils are fairly uniform despite whether they are influenced by archaeological material or not. Therefore, the next step is to investigate whether higher order PCs provide information which can be used to separate archaeological features from natural soils.

Before investigating the score plots for the higher order PCs, the variance of each PC was calculated (Figure 5.5.1). Here, PC1 represents more than 90% of the total spectral variability indicating that the soil is very homogeneous. The variance drops below 1% from the third principal component for all three sites.

The following figures show results for different combinations of higher order PCs from PC1 to PC4. Results beyond PC4 are not presented because they did not show any useful information.

Figure 5.5.2 illustrates results (score plots of various PC combinations from PC1 to PC4) for site 1, Hungary. Score plots for PC1 and other PCs separate the bright archaeological soils (blue circles). Potteries are always isolated from the other spectra and burned materials are always within the soil cluster. However, for site 1, higher order PC score plots did not improve the results much compared to the use of PC1 and PC2. Beyond the use of PC2-PC3, even the beige archaeological soils are mixed inside the natural soil cluster. This indicates that the combination of higher order PCs for this site does not provide any significant additional information for the identification of archaeological features except for some very characteristic spectra (for example pottery) which were already separated based on the first two PCs.

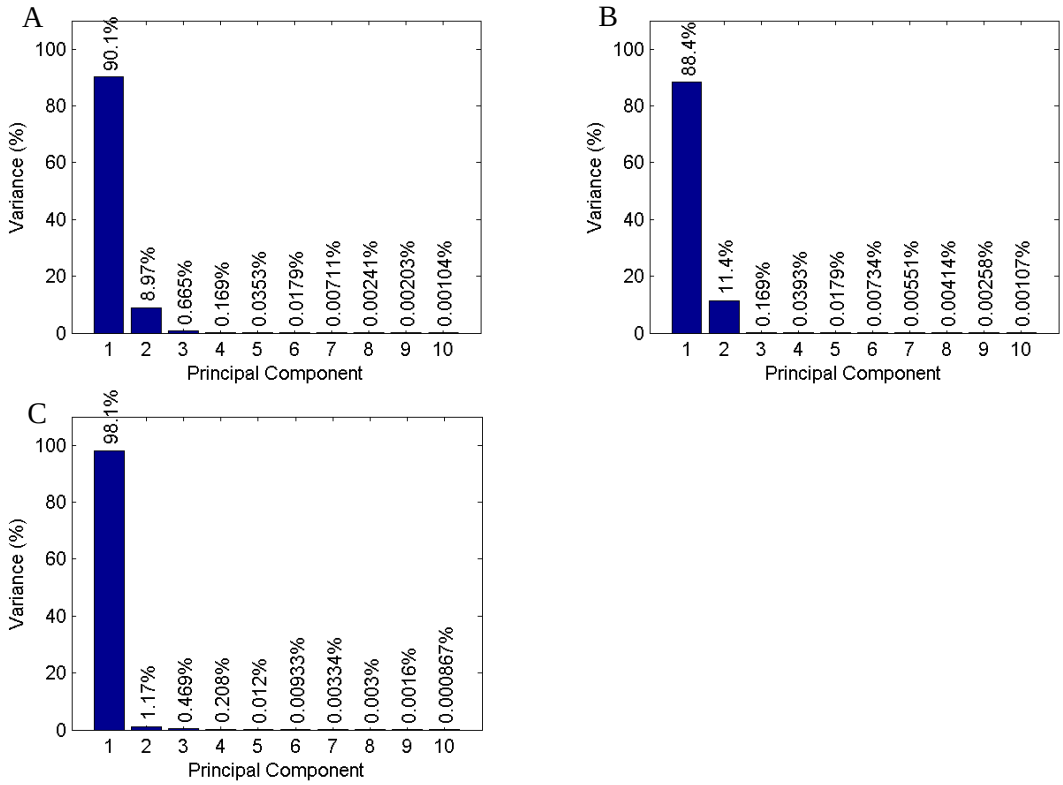


Figure 5.5.1 Contributions of the first ten principal components for the wavelength region 400 – 1000 nm for sites 1 (A), 2 (B) and 3 (C).

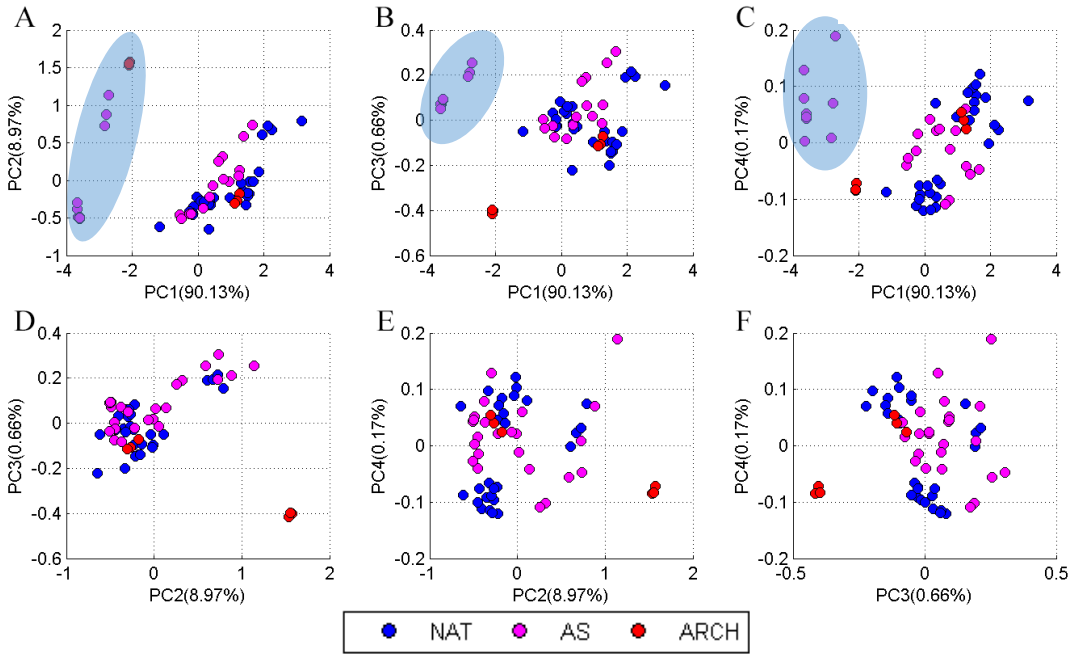


Figure 5.5.2 PCA results (score plots of PC1 to PC4) for site 1(ditch formation). NAT represents natural soils (including Ap, Bt and C), AS represents archaeological soils and ARCH archaeological artefacts. Scatter points within the blue circles represent beige coloured AS (wall materials fallen into the ditch formation).

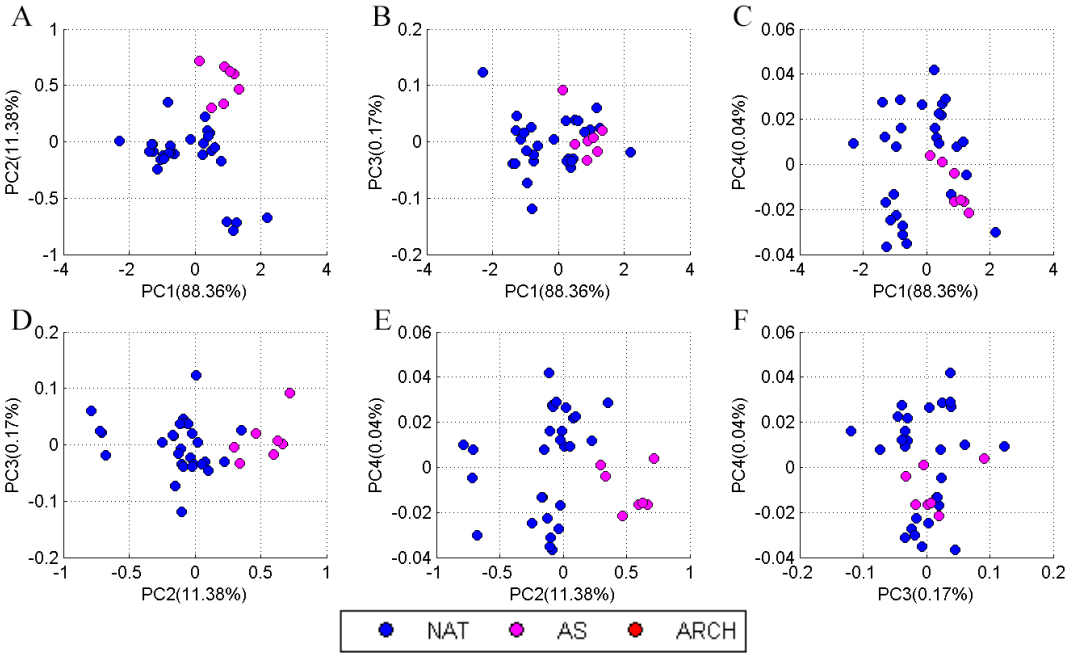


Figure 5.5.3 PCA results (score plots of PC1 to PC4) for site 2 (pit formation). No archaeological artefacts were found in this site.

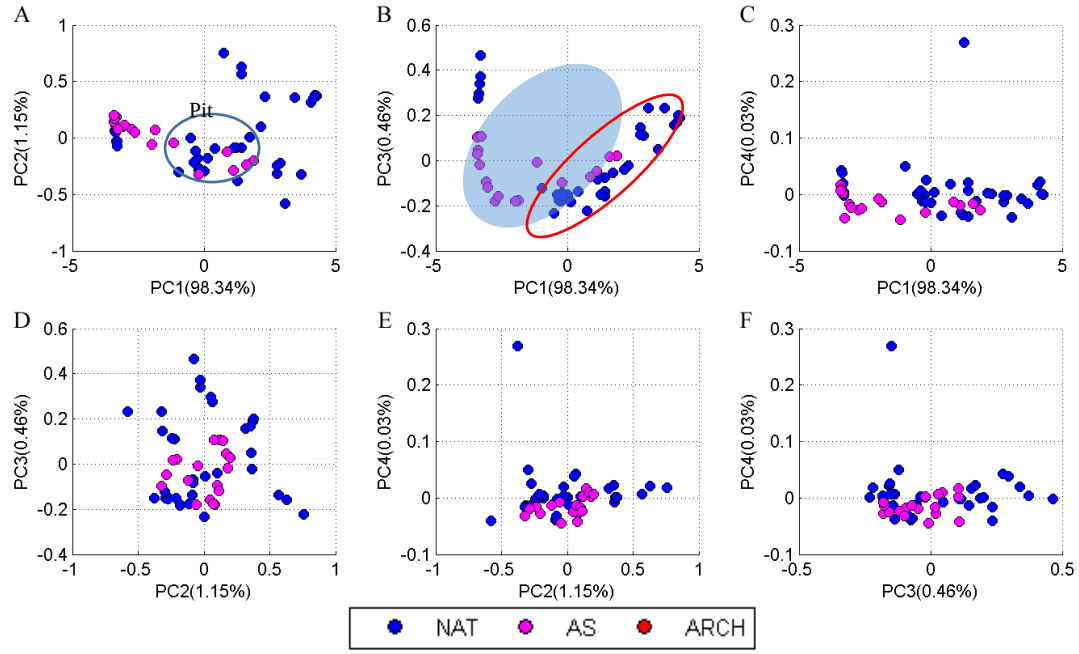


Figure 5.5.4 PCA results (score plots of PC1 to PC4) for site 3 (ditch formation). No archaeological artefacts were found in this site. The marker points within the blue circle indicate the AS spectra and the marker points within the red circle are the AS spectra which showed dark soil colour (the AS in the downhill direction, refer to Figure 5.1.6).

Figure 5.5.3 shows the score plots of different PCs (PC1-PC4) at site 2. Here, except for the combination of PC2-PC3 and PC2-PC4 the score plots of higher order PCs do not provide better results. This shows that at site 2, PC2 is the major principal component which separates archaeological materials from the background natural soils where natural soils have negative PC2 values and archaeological materials have positive PC2 values.

Figure 5.5.4 shows the corresponding results for site 3. Here, the score plot of the first two PCs does not show a clear separation between archaeological materials and natural soils. Some AS spectra are located in the centre of the natural soil cluster (blue circle). Only some of the bright archaeological soils (AS) are shifted to the left side of the plot. On the other hand, the PC1- PC3 score plot provides a comparably better separation between archaeological soils and natural soils. The AS near the soil cluster (indicated by the red circle) represents the spectra of the archaeological soils which showed similar soil colour to the background natural soils (dark archaeological soils in Figure 5.1.6.A.2). Similarly, higher order PC combinations (PC2-PC3, PC2-PC4 and PC3-PC4) do not provide useful additional information to identify archaeological soils among natural soils.

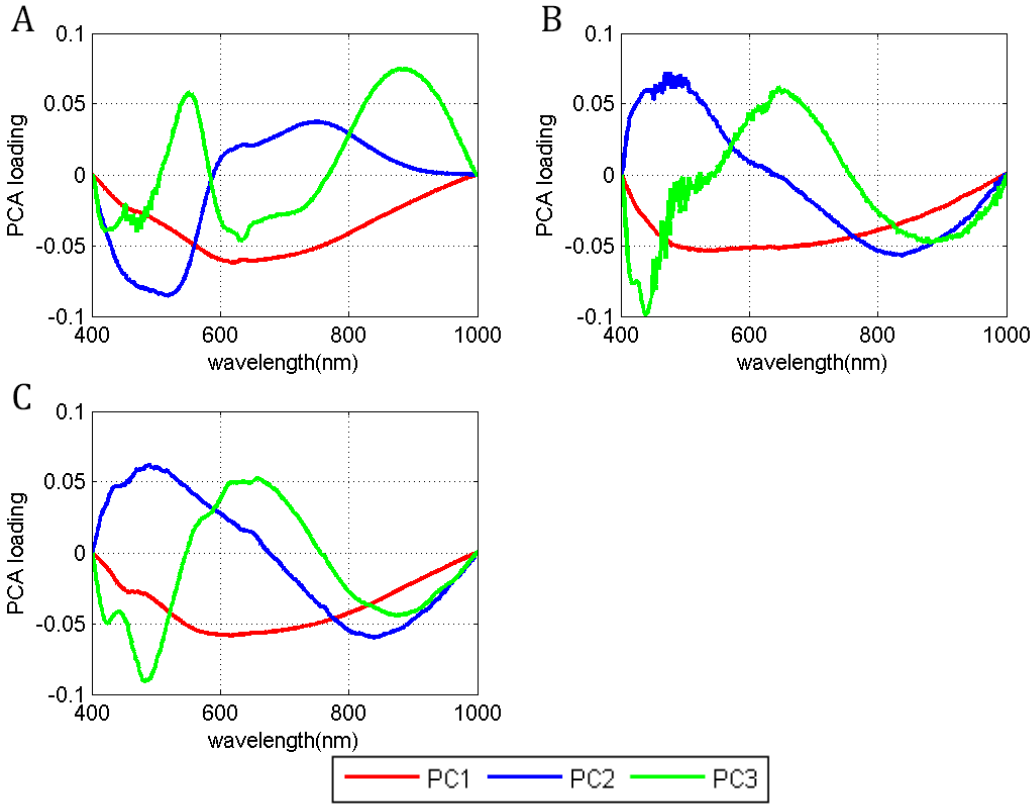


Figure 5.5.5 Spectra of the first three principal components for sites 1 (A), 2 (B) and 3 (C) for the wavelength range 400 – 1000 nm.

Figure 5.5.5 shows PCA spectra for the wavelength range 400 – 1000 nm. The entire sites share a similar PC1 pattern, a broad band spectral feature in the 400 – 1000 nm, indicating the homogeneity of the Hungarian soils. In addition, as mentioned above, the PC2 at site 2 represents a strong separation between archaeological materials and natural soils. PC2 at sites 2 and 3 are extremely similar, but a clear separation between archaeological and natural soils is

not observed at site 3.

Overall, the score plots of the higher order PCs provided some cases where archaeological materials can be separated from natural soils. However, typically the score plots of the first two principal components represent best results.

5.6 D calculation: Calculation of the Euclidean distance

So far, compared to the Italian sites, the Hungarian sites did not provide a clear separation between archaeological and natural soils although clear colour differences were observed (Chapter 5.1.2). As a next step the D calculation method (Chapter 3.3.2) is applied to the Hungarian sites. This method assumes that although the spectral features of natural soils vary depending on the soil type and sampling location, they share a similar spectral pattern compared to the spectral features of archaeological remains. Therefore a difference (D) between the original spectrum and modified spectrum, which represents the principal component (PC) values of natural soils, can be determined. If the difference D between the two spectra is small, then the spectrum is similar to the spectral features of natural soils. If not, it indicates that the spectrum is more likely to represent a non-natural soil, probably an archaeological material.

Refer to Chapter 3.3.2 for detailed explanation of the D calculation method. Also the definition and detailed information on different types of N_{soil} are explained in Chapter 3.3.2.1.

5.6.1 D calculation results for local N_{soil}

Using the N_{soil} from the original site (site 1) and Italy (Chapter 3.3.2.1 and Figure 3.3.6), the D calculation is applied to the archaeological sites 1 to 3 from Sárvíz valley, Hungary.

When the N_{soil} is obtained from the site where the measurements were taken (in this case from site 1), the D_{nat} value varies around 0.03 to 0.05 (this is better than the D_{nat} value of site 1, Italy) and the D_{arch} value varies around 0.28 to 0.44, depending on the number of PCs added. Here, when $\sum_1^2 PC$ is applied, the D_{ratio} is slightly smaller than for $\sum_1^3 PC$, but this difference is fairly small. The D_{ratio} value for N_{soil} from the local site is around 9.50, which is rather large (similar to the D_{ratio} value of site 1, Italy which was 10) and therefore indicates that the spectra of archaeological materials clearly stand out among natural soils although the score plots of the PCA results did not indicate a clear separation.

When N_{soil} is obtained from natural soil spectra gathered at a totally different site (IT), the results are reversed. Large D_{nat} values (indicating that the natural soils of Italy and Hungary are extremely different) and comparably small D_{arch} values are found, leading to D_{ratio} values lower than 1. These results indicate that the natural soils from Italy are spectrally more similar to the archaeological soils in site 1.

Figure 5.6.1 shows the D values for site 1 when N_{soil} is used from the original site (left) and from another country, Italy (right).

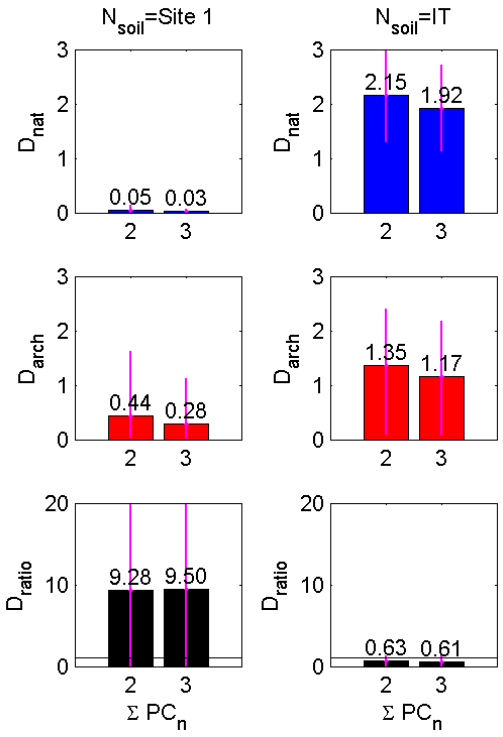


Figure 5.6.1 D_{nat}, D_{arch} and D_{ratio} values for site 1, Hungary for different sets of N_{soil} (either from the original site (site 1) or Italy (IT)). The error bars are calculated by the maximum and minimum D values using the error propagation method. The x-axis represents the numbers of PCs used for the D calculation, here the results from only $\sum_1^2 PC$ and $\sum_1^3 PC$ are shown. Overall, D_{nat} values are much smaller than D_{arch} values causing D_{ratio} values larger than 1.

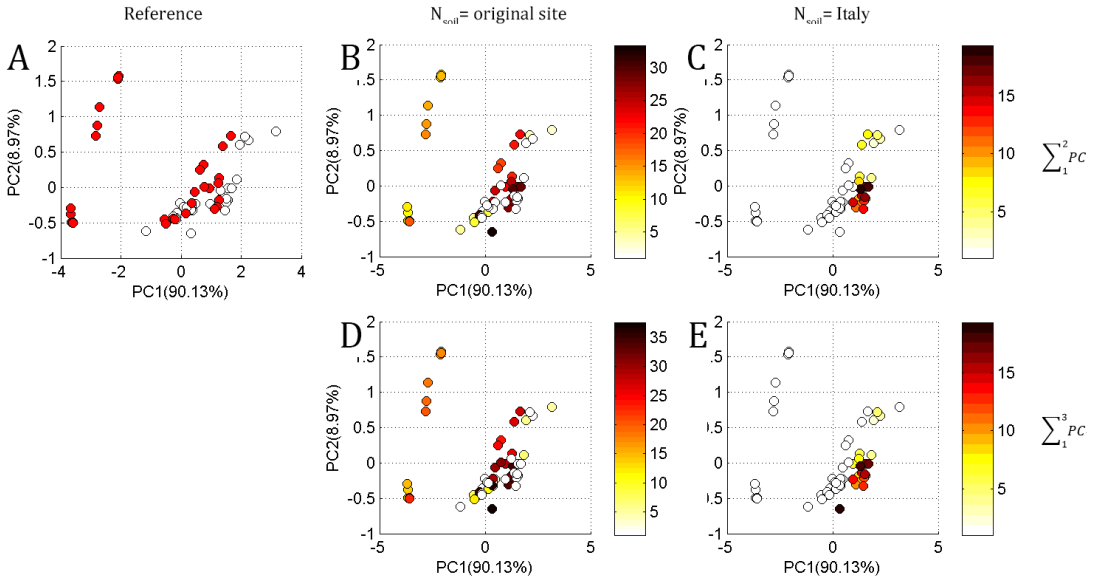


Figure 5.6.2 D_{ratio} values represented as colour on the PC1 against PC2 score plot of site 1. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) Results for N_{soil} used from the original site (site 1) with $\sum_1^2 PC$ (first two principal components used from N_{soil}). C) Results for N_{soil} used from Italy with $\sum_1^2 PC$. D) Results for N_{soil} used from the original site (site 1) with $\sum_1^3 PC$(first three principal components used from N_{soil}). E) Results for N_{soil} used from Italy with $\sum_1^3 PC$.

Figure 5.6.2 shows score plots for site 1 for which the colours indicate the D_{ratio} for the individual spectra. The first impression is that when N_{soil} is used from the original site, except for one AS spectrum all archaeological spectra are identified. However, among 28 natural soil spectra, also 12 spectra showed D_{ratio} values higher than 1, indicating that some false detection has been made. This is more extreme for N_{soil} from Italy. Here less than half of the archaeological spectra gave D_{ratio} larger than 1. Also the beige archaeological soils, which were always outstanding in the left side of the score plots, gave D_{ratio} values less than 1 suggesting that the methodology does not work for N_{soil} from another country for the sites in Hungary.

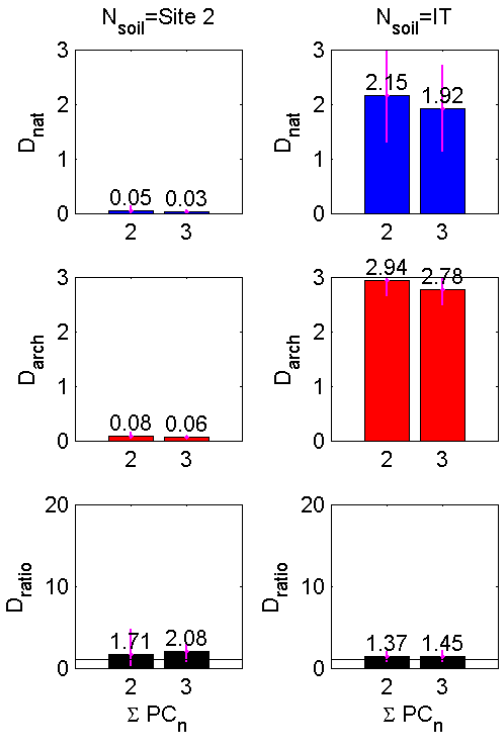


Figure 5.6.3 D_{nat}, D_{arch} and D_{ratio} values for site 2, Hungary for different sets of N_{soil} (either from the original site (site 2) or Italy (IT)). The x-axis represents the numbers of PCs used for the D calculation, here the results from only $\sum_1^2 PC$ and $\sum_1^3 PC$ are shown. Overall, D_{nat} values are much smaller than D_{arch} values causing D_{ratio} values larger than 1.

Figure 5.6.3 shows the D values for site 2 with N_{soil} from the original site (site 2) and from Italy. The D_{ratio} values at site 2 are much lower than those at site 1, but the overall result is much better. When N_{soil} is obtained from the original site (site 2), the D_{nat} value is around 0.03 to 0.05, which is similar to the D_{nat} value for site 1 (see above). However, the D_{arch} value is around 0.06 to 0.08, which is much smaller than for site 1, which was around 0.28 to 0.44. This results in D_{ratio} values around 1.71 to 2.08 which are still large enough to detect archaeological features.

For site 1, when N_{soil} obtained from Italy was used for the D calculation the method did not work well and it was even found that the archaeological soils in Hungary are similar to the natural soils in Italy. However, for site 2, although the D_{nat} values are rather large (around 1.92 to 2.15) also high D_{arch} values of around 2.78 to 2.94 are obtained. This leads to D_{ratio} values greater than 1, indicating that the methodology works well for site 2. It is interesting to note that, sites 1 and 2 share similar natural soil spectra (since they are located in the same field, refer to Chapter 5.1.2) but provide totally different results due to different D_{arch} values. The buried remain at site 1 is a ditch formation with fallen wall structures whereas at site 2 is an

ancient rubbish pit with a high amount of organic matter. This might indicate that pit formations, especially the ones with high amounts of organic matter, contain more complex mineral compositions than natural soils (Courchesne et al., 2015), and are thus easier to detect than brick fractures.

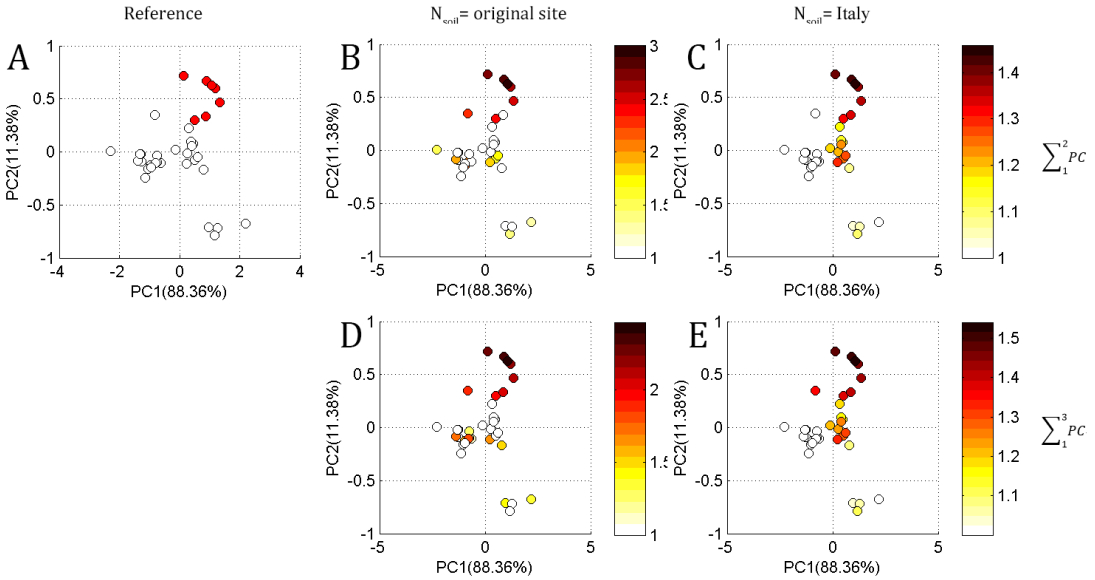


Figure 5.6.4 D_{ratio} values represented as colour on the PC1 against PC2 score plot of site 2. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) Results for N_{soil} used from the original site (site 2) with $\sum_1^2 PC$ (first two principal components used from N_{soil}). C) Results for N_{soil} used from Italy with $\sum_1^2 PC$. D) Results for N_{soil} used from the original site (site 2) with $\sum_1^3 PC$(first three principal components used from N_{soil}). E) Results for N_{soil} used from Italy with $\sum_1^3 PC$.

Figure 5.6.4 shows score plots for site 2 for which the colours indicate the D_{ratio} for the individual spectra. Similar to site 1, all archaeological spectra are identified when N_{soil} from the original site is used, but with some false detections for the natural soils (11 natural soil spectra gave D_{ratio} larger than 1). When N_{soil} from Italy is used, all archaeological spectra have D_{ratio} larger than 1. However there are a large numbers of false detections of natural soils (12 natural soils with D_{ratio} values larger than 1).

For site 3 (Figure 5.6.5), the overall result is less good compared to sites 1 and 2. Here, when the N_{soil} is obtained from the site where the measurements were taken (in this case from site 3), the D_{nat} value varies around 0.05 to 0.17 and the D_{arch} value varies around 0.70 to 1.13, depending on the number of PCs added. The difference between the D_{nat} and D_{arch} values is small and, therefore, D_{ratio} values around 0.67 to 1.43 are obtained. When $\sum_1^2 PC$ is applied, the D_{ratio} value is less than 1 (D_{ratio} of 0.67) and when $\sum_1^3 PC$ is applied the D_{ratio} improves slightly. This suggests that the method does not even work when N_{soil} from the original site is applied. The calculation gives better results with summations of higher number of PCs for the D calculation. When N_{soil} is calculated from natural soil spectra gathered at a totally different site (IT), the D_{ratio} values are even smaller than 1. Site 3 is a ditch feature (similar to site 1) but showed clearer colour boundaries between archaeological and natural soils (Figure 5.1.6)

than site 1. This result once again indicates that the colour of the soil is not the main factor the methodology takes into account.

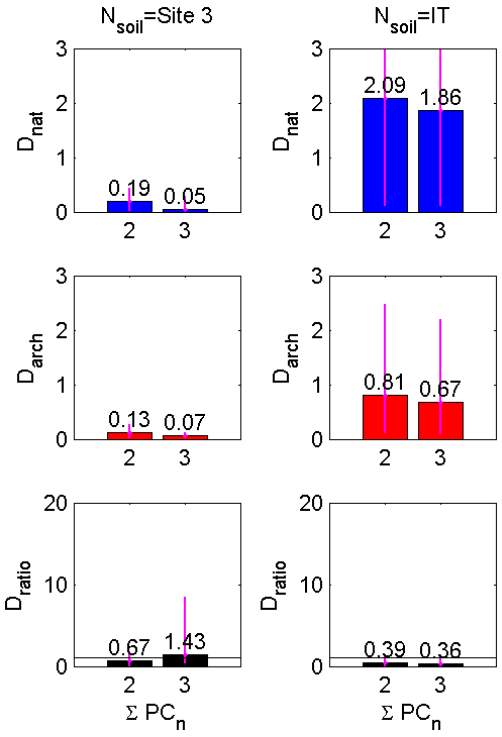


Figure 5.6.5 D_{nat}, D_{arch} and D_{ratio} values for site 3, Hungary for different sets of N_{soil} (either from the original site (site 3) or Italy (IT)). The x-axis represents the numbers of PCs used for the D calculation, here the results from only $\sum_1^2 PC$ and $\sum_1^3 PC$ are shown. Overall, D_{nat} values are much smaller than D_{arch} values causing D_{ratio} values larger than 1.

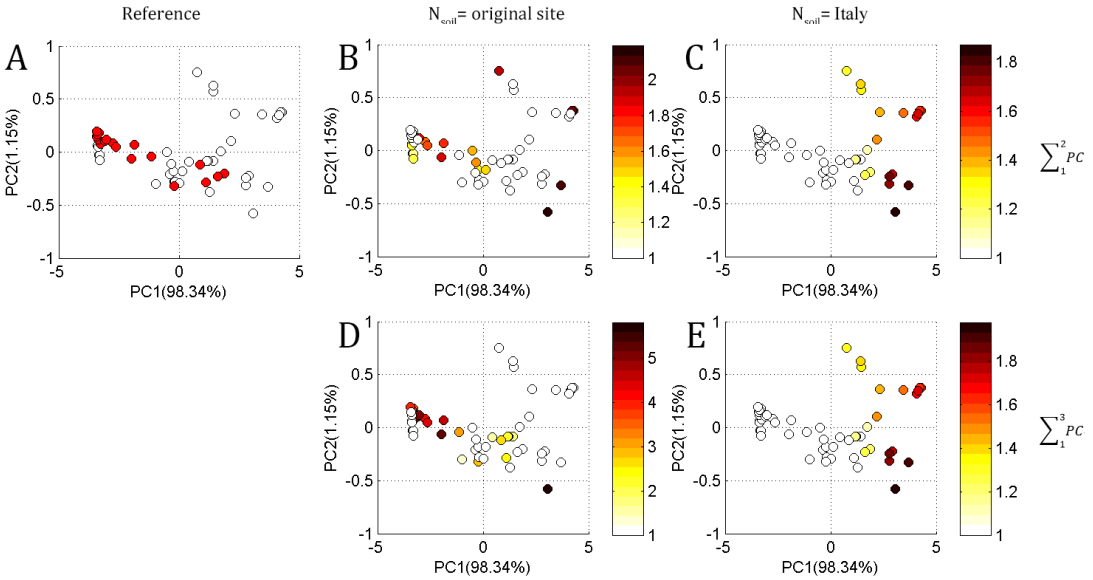


Figure 5.6.6 D_{ratio} values represented as colour on the PC1 against PC2 score plot of site 3. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) Results for N_{soil} used from the original site (site 3) with $\sum_1^2 PC$ (first two principal components used from N_{soil}). C) Results for N_{soil} used from Italy with $\sum_1^2 PC$. D) Results for N_{soil} used from the original site (site 3) with $\sum_1^3 PC$(first three principal components used from N_{soil}). E) Results for N_{soil} used from Italy with $\sum_1^3 PC$.

Figure 5.6.7 shows score plots for site 3 for which the colours indicate the D_{ratio} for the individual spectra. The first impression is that when N_{soil} from the original site is used, only six archaeological spectra among 18 spectra and 15 natural soil spectra showed D_{ratio} values larger than 1. The results are even worse for N_{soil} from Italy. Here, only two archaeological spectra show D_{ratio} larger than 1 and the rest are false detections for the natural soils.

Overall, the D calculation results suggest that site 2 contains the most distinctive archaeological remains, whereas the archaeological remains at site 3 are difficult to identify. One interesting finding is that, although site 1 and site 2 share similar natural soils, since they are located in the same field, the D calculation varied strongly (due to the different types of buried archaeology). This finding indicates that the D calculation for different types of archaeological remains should be investigated in more detail in future studies.

5.6.2 D calculation results for global N_{soil}

The previous section showed D results when N_{soil} were either collected from the local site or a completely different site using a limited number of natural soil spectra. In this chapter, the calculations are repeated for a 'universal' N_{soil} dataset. Two types of global datasets are used: 1) natural soils gathered from both Italy and Hungary (IT+HUN) and 2) natural soils from the ISRIC spectral library. As mentioned in Chapter 3.3.2.1, the ISRIC spectra are sampled at low resolution (every tenth nanometre). Therefore, first the natural soils from Italy and Hungary are used to investigate the effect of different spectral resolutions on the D results.

Figure 5.6.7 shows the D values for sites 1, 2 and 3 when N_{soil} is used from both Italy and Hungary. The overall results look much better than the result from Chapter 5.6.1 when the local N_{soil} was used. The D_{nat} values, which vary between 0.07 and 0.14, are identical for sites 1 and 2, confirming that they share the same natural soil spectra. Sites 1 and 3 show similar D_{arch} values of 0.10 to 0.22 while for site 2 a comparably higher value of around 0.11 to 0.46 is found. The D_{ratio} of site 3 is still below 1 when $\sum_1^2 PC$ is applied (D_{ratio} of 0.79) but improves when PC3 is added ($\sum_1^3 PC$).

Results from Chapter 4 (Italy) showed that using only the first two principal components for the D calculation provides better result than using up to the third principal component since the contribution of PC3 is fairly low compared to PC1 and PC2 (refer to Figure 3.3.6 and Figure 3.3.8). However, some results of the Hungarian sites suggest that sometimes adding up to the third PC ($\sum_1^3 PC$) can improve the D_{ratio} value (e.g. sites 1 and 3, Hungary). Therefore it will be useful to consider both $\sum_1^2 PC$ and $\sum_1^3 PC$ for investigating D_{ratio} results.

The D calculation method is will also be applied with N_{soil} gathered from the ISRIC spectral library. According to Chapter 3.3.2.1, the soil spectra from the ISRIC spectral library are resampled at every tenth-nanometre. Since this resampling process represents an undersampling of the data, one might expect an effect of this undersampling on the D values. Therefore, before the ISRIC library is applied, a comparison is made between the D results for the original spectra (from Italy and Hungary) and the corresponding resampled spectra (Figure 5.6.8).

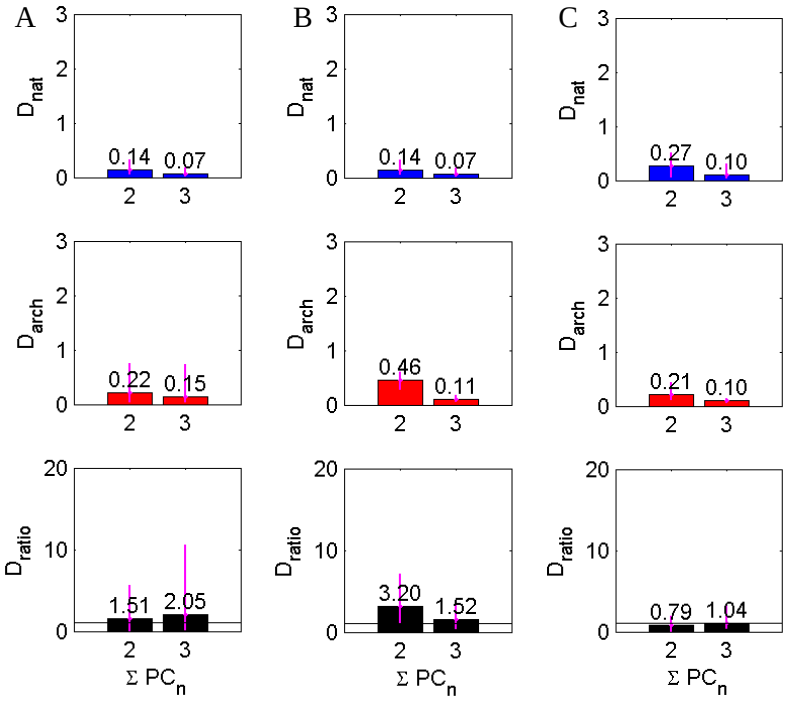


Figure 5.6.7 D calculation results for site 1 (A), 2 (B) and 3 (C) when N_{soil} from Italy and Hungary (IT+HUN) is used for $\sum_1^2 PC$ and $\sum_1^3 PC$.

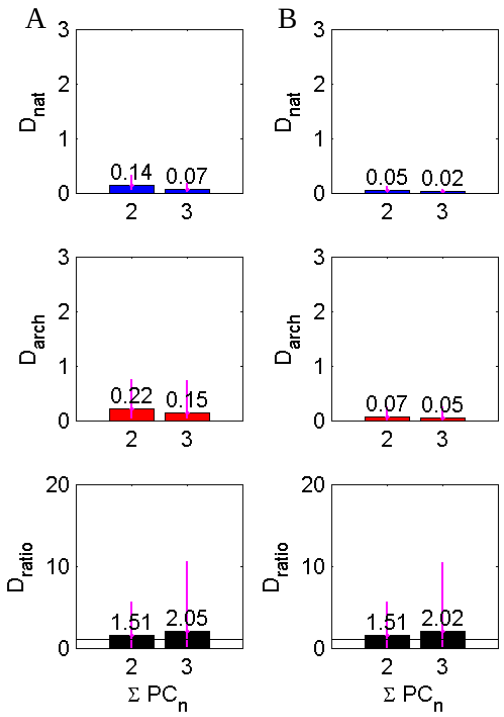


Figure 5.6.8 D_{nat}, D_{arch} and D_{ratio} values for site 1, Hungary when N_{soil} is from both Italy and Hungary (IT+HUN). A: Original spectra. B: Spectra resampled at every tenth-nanometre. Refer to Appendix 6 for the results for sites 2 and 3, Hungary.

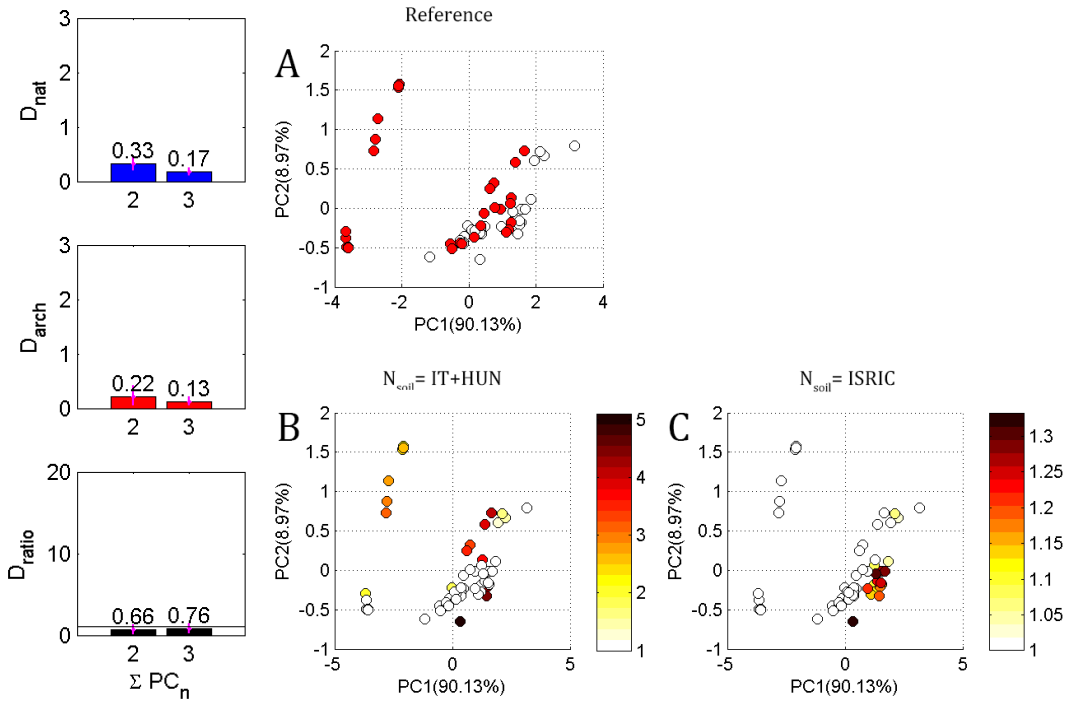


Figure 5.6.9 D calculation results for site 1 when a combined N_{soil} from ISRIC spectral library is used. The bar chart on the left side shows the D_{nat}, D_{arch} and D_{ratio} values for $\sum_1^2 PC$ and $\sum_1^3 PC$. In the PCA score plots on the right sides the D_{ratio} values are represented as colour, where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. Figure A) is a reference score plot where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. B) Results for N_{soil} used from both Italy and Hungary with $\sum_1^2 PC$ (first two principal components used from N_{soil}). C) Results for N_{soil} used from ISRIC with $\sum_1^2 PC$. Spectra used in these figures are resampled on a ten-nanometre grid.

The D_{nat} and D_{arch} values for the resampled dataset are around a factor of three smaller than the original D values (Figure 5.6.8). In addition, the D_{arch} values (0.05 to 0.07) of the resampled spectra are around three times larger than the D_{nat} values (0.15 to 0.22), which is in similar proportion to the D_{arch} and D_{nat} values of the original data. This leads to similar D_{ratio} values of around 1.5 to 2 for the original and resampled spectra. Similar results are also obtained for sites 2 and 3 (refer to Appendix 6).

Figure 5.6.9 shows the D calculation results for site 1 when the N_{soil} is used from ISRIC. Here, the D_{nat} value is around 0.33 to 0.17 and the D_{arch} value is around 0.13 to 0.22. This results in D_{ratio} values lower than 1 showing that the method does not work well when N_{soil} from ISRIC is used. When the individual D_{ratio} values (for N_{soil} from ISRIC) are observed (right figures in Figure 5.6.9), all characteristic archaeological soils (the beige colour AS) have D_{ratio} values lower than 1. Only 4 archaeological spectra show D_{ratio} values larger than 1 and there is a large number false detections of natural soils as archaeological materials. Unlike Italy, where the D calculation method still worked nicely for the ISRIC spectral library, the method does not work well at this site.

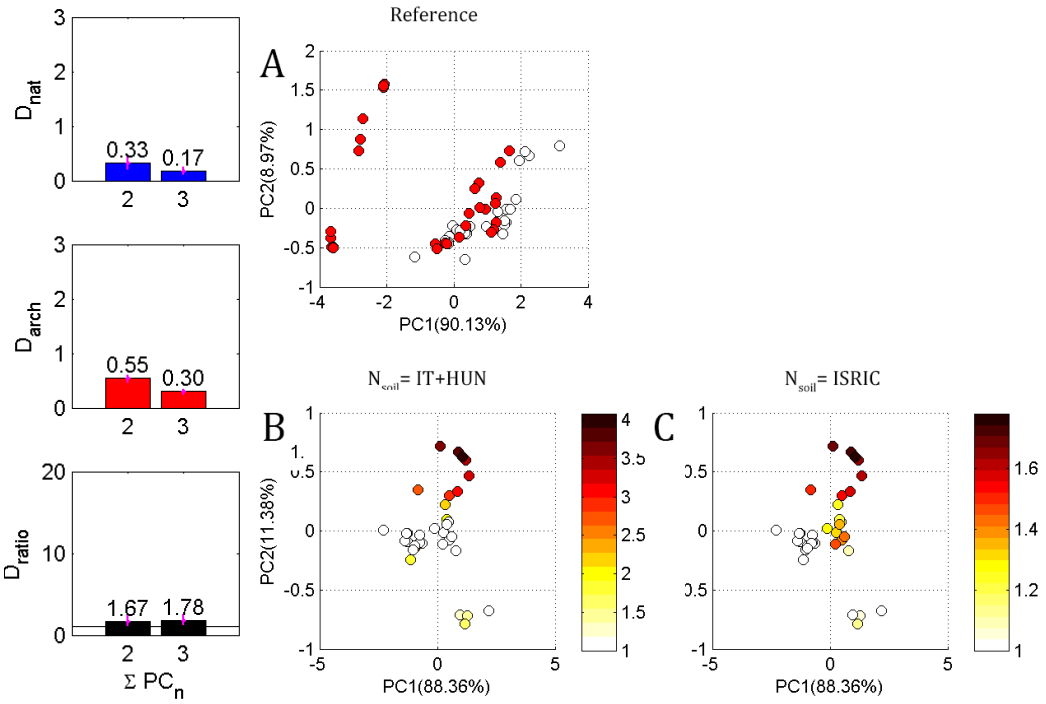


Figure 5.6.10 D calculation results for site 2 when a combined N_{soil} from ISRIC spectral library is used. The bar chart on the left side shows the D_{nat}, D_{arch} and D_{ratio} values for $\sum_1^2 PC$ and $\sum_1^3 PC$. In the PCA score plots on the right sides the D_{ratio} values are represented as colour, where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. Figure A) is a reference score plot where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. B) Results for N_{soil} used from both Italy and Hungary with $\sum_1^2 PC$ (first two principal components used from N_{soil}). C) Results for N_{soil} used from ISRIC with $\sum_1^2 PC$. Spectra used in these figures are resampled on a ten-nanometre grid.

Figure 5.6.10 shows the D values for site 2 when N_{soil} is used from the ISRIC spectral library. Here, the D_{nat} values are around 0.17 to 0.33 and D_{nat} value around 0.3 to 0.55, which is around two times larger than the D_{nat} values. Therefore the D_{ratio} values are around 1.67 to 1.78. $\sum_1^3 PC$ gives slightly higher D_{ratio} values than $\sum_1^2 PC$. Among the Hungarian sites, site 2 always provides positive results which are also indicated by the PCA score plots (Figure 5.6.10). Here, the archaeological soils (pit feature) are identified with data pints with high D_{ratio} values (dark red markers). The D_{arch} values of this site are always more than two times larger than those of sites 1 and 3. Depending on the type of N_{soil} used, the number of false detections of natural soils varies, which also changes the overall D_{ratio} values.

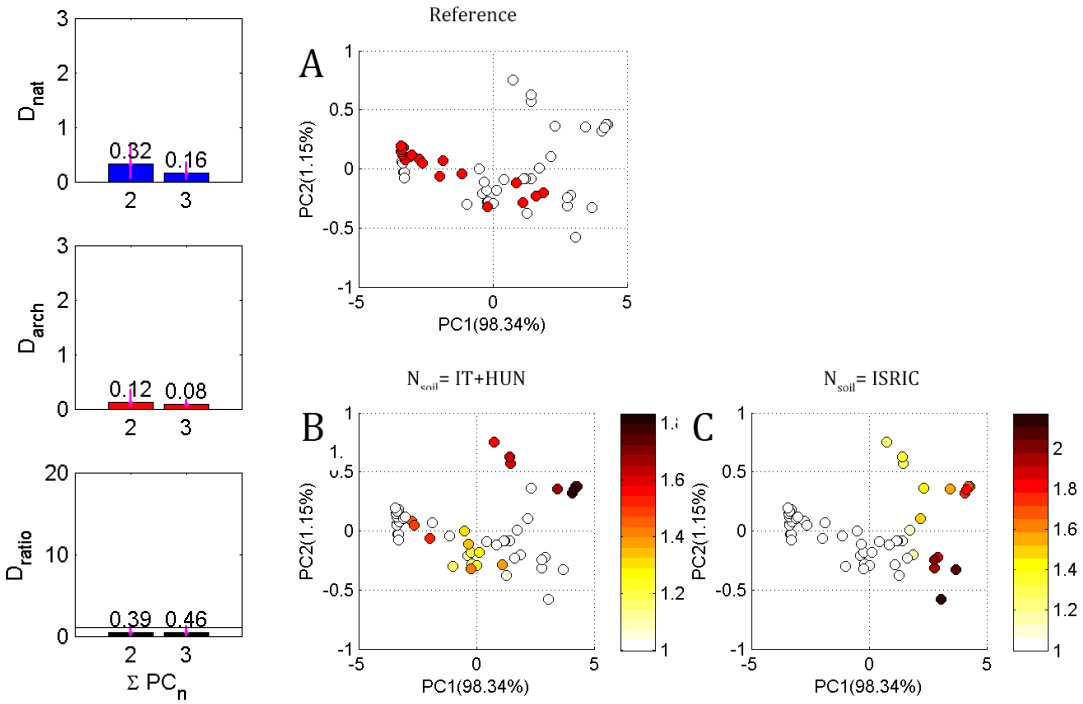


Figure 5.6.11 D calculation results for site 3 when a combined N_{soil} from ISRIC spectral library is used. The bar chart on the left side shows the D_{nat}, D_{arch} and D_{ratio} values for $\sum_1^2 PC$ and $\sum_1^3 PC$. In the PCA score plots on the right sides the D_{ratio} values are represented as colour, where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. Figure A) is a reference score plot where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. B) Results for N_{soil} used from both Italy and Hungary with $\sum_1^2 PC$ (first two principal components used from N_{soil}). C) Results for N_{soil} used from ISRIC with $\sum_1^2 PC$. Spectra used in these figures are resampled on a ten-nanometre grid.

Figure 5.6.11 shows the D_{ratio} values for site 3 when the global natural soil (ISRIC) is used. The D_{ratio} values are significantly below 1 indicating that when the ISRIC natural soils are used the method does not work for site 3, Hungary. The problem is that the archaeological soils show similar spectral features as the natural soils of the ISRIC library leading to very small D_{arch} values. The corresponding score plots (Figures 5.6.12.B and 5.6.12.C) also support this finding. When natural soils from Italy and Hungary are used for N_{soil}, 23 spectra showed D_{ratio} values larger than 1 and only five of them were archaeological soils. When N_{soil} from ISRIC is used, among 17 spectra with high D_{ratio} values only one spectrum represents archaeological soil but the others are false detections.

Just by looking at the D calculation results of the Hungarian sites, it seems that the method does not work well for ditch formations. This might vary according to what anthropogenic materials have fallen into the ditch since site 1 gave slightly better result than site 3. Also although archaeological strata always showed clear colour differences between natural and archaeological soils, the colours of the soil are not the dominating factors for the PC values and, thus, does not always present good D results. Therefore the D calculation method should be applied to more various archaeological sites to test its effectiveness.

5.7 Dependence of the results for smoothing levels

Similar to Chapter 4, different levels of smoothing are applied to the spectra to test the minimum resolution of the spectrometer needed to separate archaeological remains from natural soils through the PCA method developed in this thesis. In the previous section, Chapter 5.6.2, D values indicated that when the spectra were resampled at every tenth-nanometre, the results do not substantially change compared to the D values for the original spectra. Hence, a 'high resolution' spectrometer might not be needed for the identification of archaeological features.

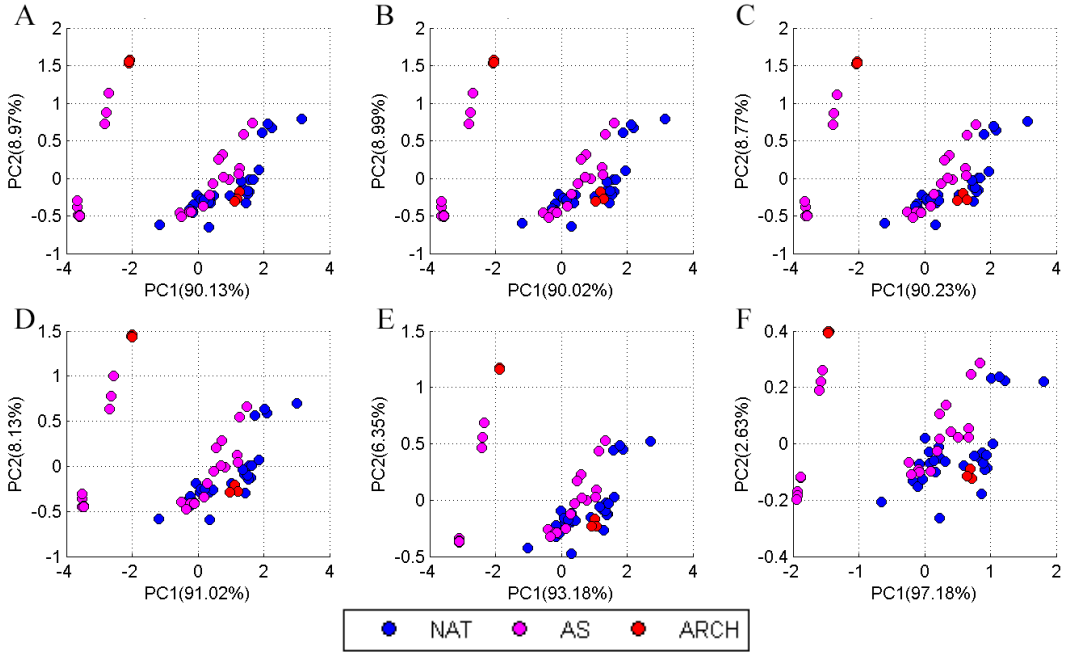


Figure 5.7.1 Score plots for the first two PCs for site 1 when different smoothing kernels were applied. The convolution kernels used for smoothing are; A: original, B: 10 nm, C: 20 nm, D: 50 nm E: 100 nm and F: 200 nm.

Figure 5.7.1 shows various PCA results when different convolution kernels were applied to spectra. Similar to the results for Italy, the PCA score plot seems to be identical for different smoothing kernels. For higher smoothing kernel, the PC2 values decrease: for example, for potteries (the red markers at the top left corner of the figures) PC2 values around 1.5 are found for smoothing kernels up to 50 nm kernels, but the PC2 values decrease to about 0.4 for a smoothing kernel of about 200 nm. A similar pattern was also found for PC1. Overall, the score plots (PC1 against PC2) did not change strongly by applying the smoothing kernel.

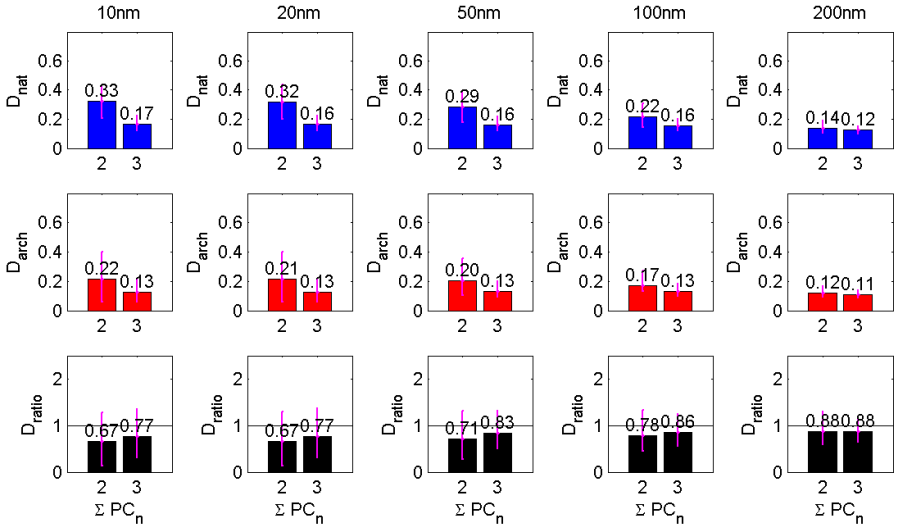


Figure 5.7.2 D values for site 1 where spectra are smoothed with different convolution kernels (10 nm, 20 nm, 50 nm, 100 nm and 200 nm). Here, N_{soil} from the ISRIC spectral library is used.

Figure 5.7.2 shows the D values for the corresponding PCA results from Figure 5.7.1. Here, N_{soil} from the universal dataset is used and only results for $\sum_1^2 PC$ and $\sum_1^3 PC$ are shown. Notice that, similar to the results shown in Figure 5.7.1, the D values depend on the smoothing kernel. Especially, when a 200 nm kernel is used, the D values of natural soils and archaeological soils become similar and therefore yield to D_{ratio} values close to 1. The decreased D_{arch} value indicates that when a 200 nm kernel is applied, the characteristic archaeological features are smoothed out and the archaeological spectra behave more similar to natural soil spectra.

One interesting thing is that, although the D calculation method did not work nicely at this site (site 1), the overall D_{ratio} increases as more smoothing is applied (D_{ratio} of 0.88 for 200 nm kernel and D_{ratio} of 0.67 for 10 nm kernel). This is contradictory to the statement above where archaeological soils lose their characteristics by high levels of smoothing. However, this result (Figure 5.7.2) once again indicates that the global N_{soil} (ISRIC spectral library) is not very successful at site 1 and smoothing does not improve this.

Figure 5.7.3 shows various PCA results when different convolution kernels were applied for smoothing the spectra of site 2. This figure also shows that the PCA score plot are very similar although different smoothing kernels were applied. But again, with increasing smoothing kernel the PC1 and PC2 values systematically decrease.

Figure 5.7.4 shows the D values for the corresponding PCA results from Figure 5.7.3. Here for increasing smoothing kernels both the D_{nat} and D_{arch} values decrease, but the D_{ratio} values are almost unchanged with highest values for the 100 nm kernel.

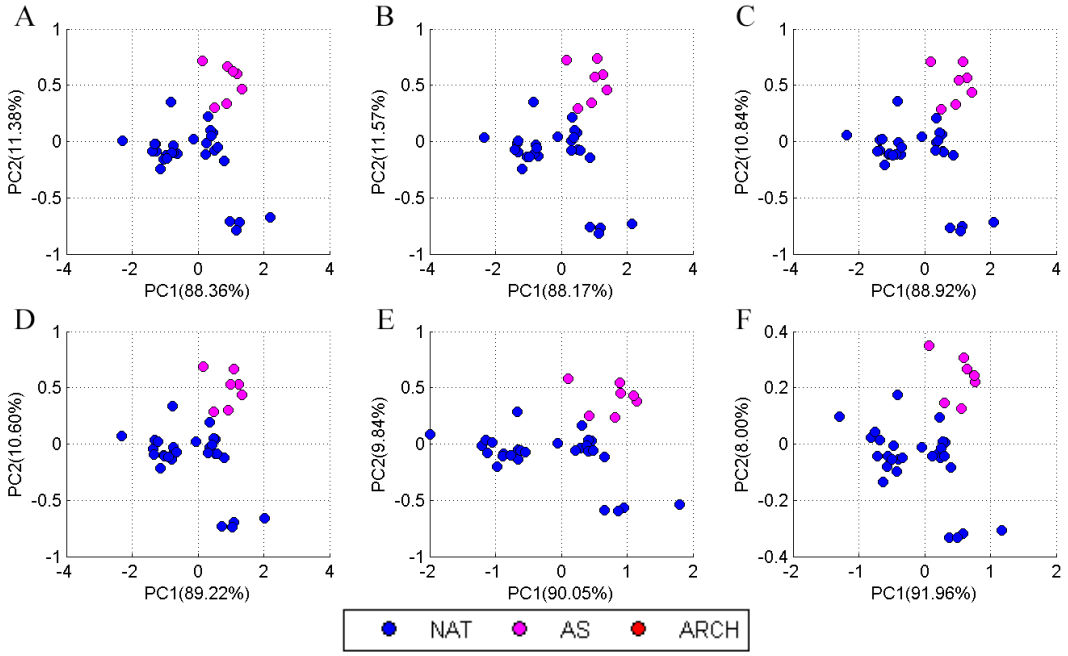


Figure 5.7.3 Score plots for the first two PCs for site 2 when different smoothing kernels were applied. The convolution kernels used for smoothing are; A: original, B: 10 nm, C: 20 nm, D: 50 nm E: 100 nm and F: 200 nm

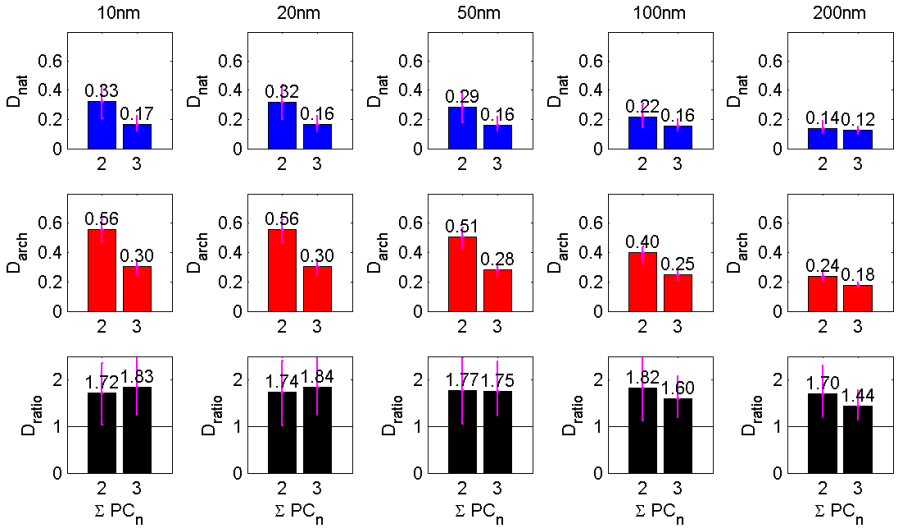


Figure 5.7.4 D values for site 2 where spectra are smoothed by different convolution kernel (10 nm, 20 nm, 50 nm, 100 nm and 200 nm). Here, N_{soil} from the ISRIC spectral library is used.

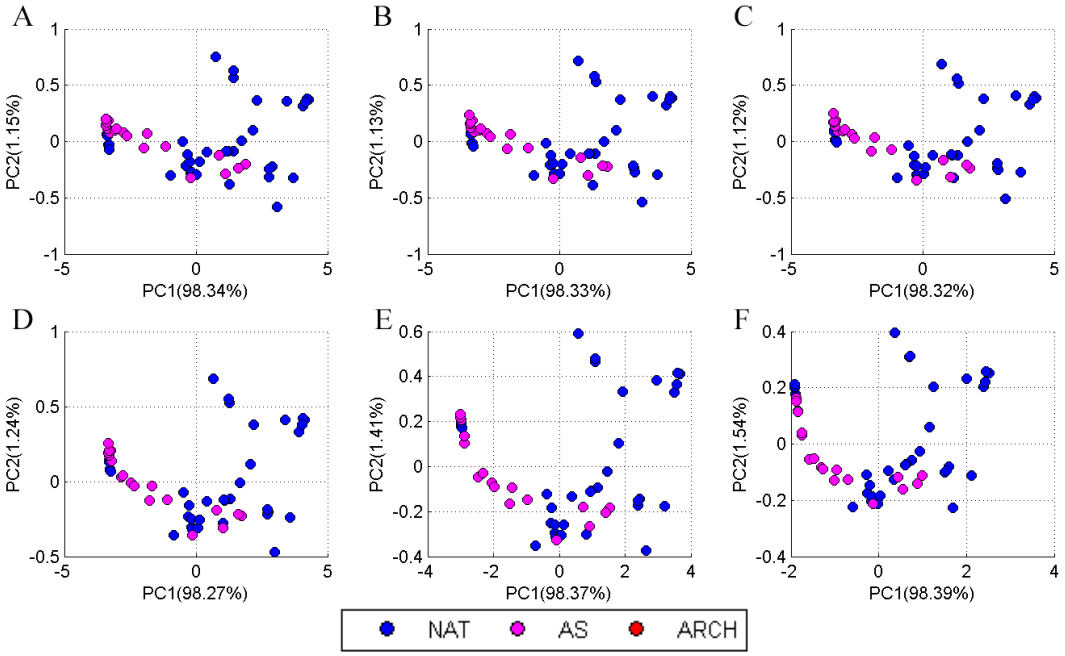


Figure 5.7.5 Score plots for the first two PCs for site 3 when different smoothing kernels were applied. The convolution kernels used for smoothing are; A: original, B: 10 nm, C: 20 nm, D: 50 nm E: 100 nm and F: 200 nm.

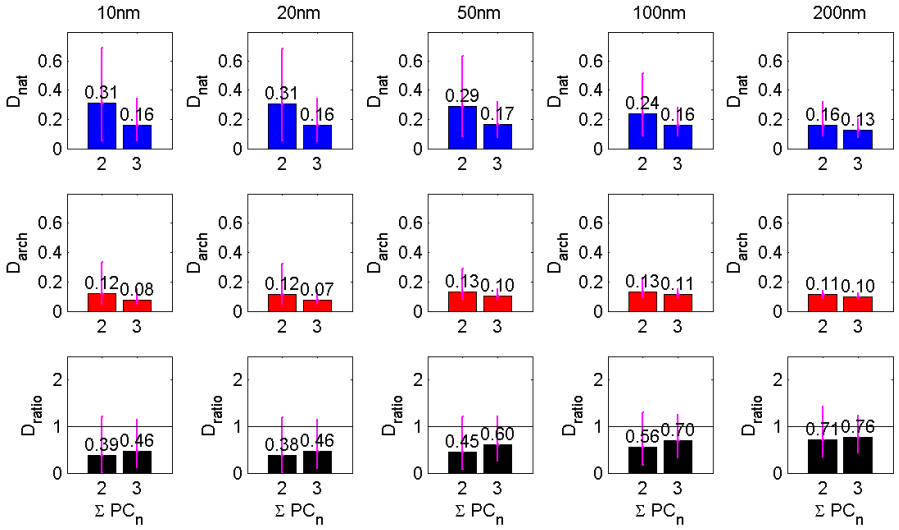


Figure 5.7.6 D values for site 3 where spectra are smoothed between 400 – 1000 nm at different convolution kernel (10 nm, 20 nm, 50 nm, 100 nm and 200 nm). Here, N_{soil} is from the natural soils of ISRIC spectral library.

Figure 5.7.5 shows the PCA score plots for the different smoothing kernels applied to the spectra from site 3. Similar to the result for sites 1 and 2, as the smoothing kernel gets higher, the PC values decrease. However, as shown in Figure 5.7.6 for this site the D values even increase with increasing smoothing kernels. But similar to site 1, the D_{ratio} values stay below 1.

This chapter indicated that spectrometers with a high spectral resolution may be nice, but it is probably not essential to have such high spectral resolution. However, since the D calculation method did not work well for most of the sites in Hungary and, therefore, it is difficult to define an exact spectral resolution value.

Site 2 was the only "well-defined" site in Hungary and based on the smoothing results of site 2 (Figure 5.7.3 and Figure 5.7.4), perhaps a spectral resolution of about 100 nm might be sufficient to identify archaeological features.

5.8 D values of soil mark features and geological feature

So far, the D calculation method was applied to *in situ* spectral measurements of archaeological remains and natural soils directly from the soil profile. This chapter applies the method to a soil mark features (from site 2) and a geological feature which was misinterpreted as a soil mark feature (from site 3). To do this, soil surface spectra (gathered from top 10 cm within the Ap horizons) and natural soils from the C horizons of each pit are selected from the original dataset. For soil mark features, three spectral measurements from the Ap horizon are selected for each site (site 2 and site 3). Note that all measurements are made using an artificial light source with the ASD spectrometer (refer to Chapter 3.1). The obtained results might provide a guideline of which D_{ratio} values might be expected when the method is applied to airborne or satellite images. Also it shows whether geological features, which look similar to soil mark features, can be identified as non-archaeological features using the D calculation method. In the following, for soil mark calculation, the D_{arch} stands for the D value of the soil mark features and for geological feature calculation, the D_{arch} stands for the D value of the geological features.

5.8.1 Soil mark

Figure 5.1.3 showed a clear soil mark feature visible at site 2, which is an ancient rubbish pit feature with a high amount of organic matter and thus a darker soil colour compared to the surrounding natural soils.

Figure 5.8.1 shows the D calculation results for the soil mark feature when N_{soil} is from (1) the original site (site 2), (2) the natural soils gathered from Italy and Hungary and (3) the ISRIC spectral library. Overall, the N_{soil} from the original site and from IT+HUN gives D_{ratio} values slightly higher than 1 indicating that the soil mark feature is detectable. However, when N_{soil} from ISRIC soils is used, the D_{ratio} values drop slightly below 1. The low D_{arch} values may also be related to the weathering of the surface soils and transportation of soils due to physical or chemical environmental changes. Note that soil mark features are most distinctive just after ploughing (Taylor, 1979) and the distinctive soil mark colours fade away with time.

These factors might at least partly be responsible for the low D_{ratio} value of the soil mark feature at site 2.

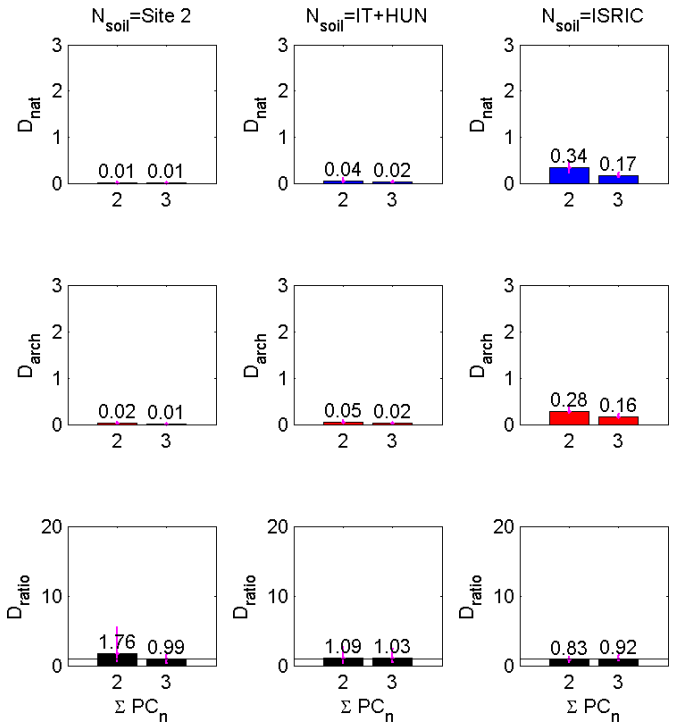


Figure 5.8.1 D value (D_{nat}, D_{arch} and D_{ratio}) results for the soil mark feature at site 2, Hungary, for various N_{soil} combinations. The spectra are resampled at every tenth-nanometre. Here D_{arch} represents the D values of the archaeological soil mark feature.

5.8.2 Geological surface feature

At site 3, a geological soil colour difference was misinterpreted as a soil mark feature (Chapter 5.1.2). This feature had a comparably brighter soil colour than the surround natural soils and had a circular shape. However, when the site was excavated, no traces of archaeological activities were observed. Therefore, for this feature a low D_{ratio} value (smaller than 1) is expected.

Figure 5.8.2 shows the D values for this geological feature when various N_{soil} groups were applied. The D_{ratio} value is less than 1 when N_{soil} from the original site is used, indicating that this is not an archaeological feature. However, when N_{soil} is used from either IT+HUN or ISRIC, the D_{ratio} value exceeds 1 indicating that the feature is showing a non-natural spectral signature. Compared to the soil mark result (Chapter 5.8.1), the D_{ratio} value of the geological feature looks more like an archaeological feature.

This shows that for surface soil spectra, using N_{soil} from the original site provides more realistic results.

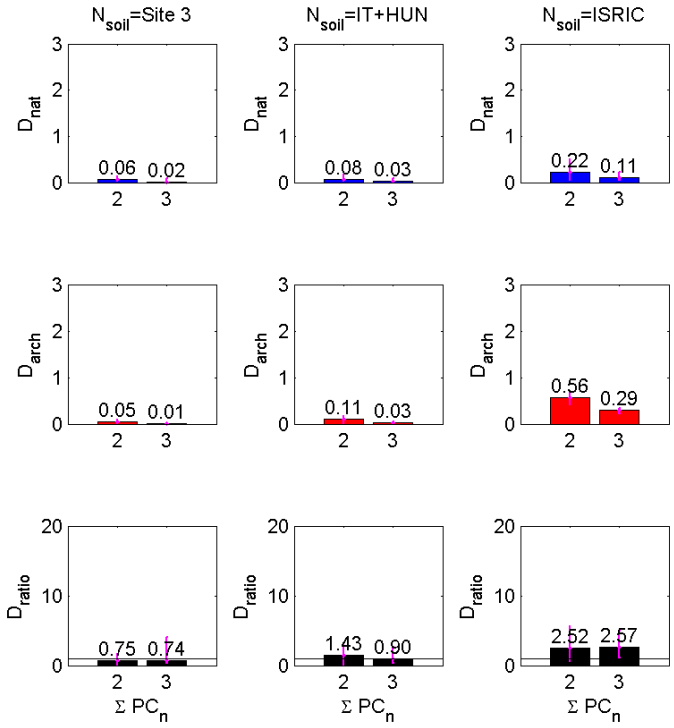


Figure 5.8.2 D value (D_{nat}, D_{arch} and D_{ratio}) results for the geological feature at site 3, Hungary with various N_{soil} combinations. The spectra are resampled at every tenth-nanometre. Here D_{arch} represents the D values of the geological feature which was misinterpreted as a soil mark feature.

5.9 Summary

To identify spectral features of buried archaeological remains, a modified quantitative PCA method (so called D values) was developed in this thesis. The application of the D calculation method showed that it is possible to identify buried archaeological materials at the sites in Italy (Chapter 4). Therefore, in Chapter 5, the PCA and D calculation methods were also applied to the archaeological sites in Hungary with great expectation due to simple background natural soil. The sites in Hungary are ditch and pit formations where a clear boundary between the archaeological formations and surrounding natural soils was observed by eye.

Similar to the results in Italy, the XRF analysis showed that the soil samples gathered in Hungary were mainly K, Ca, Ti and Fe. The element Ba, which showed outstanding values for archaeological materials at the sites in Italy, was also high in the Ap soils above archaeological remains in Hungary. This indicates that barium (Ba) may give a first glance of whether the soil investigated is influenced by archaeological or human activities. However, overall, no outstanding or dominant elements were observed for archaeological materials.

The PCA results for the sites in Hungary were not as successful as the results in Italy. Only very bright and light coloured archaeological soils were separated from the natural soil cluster. Burned materials, which were nicely separated in the spectra from Italy, were mixed within the natural soil cluster (Chapter 5.3). Applying the PCA to various wavelength regions showed that in the infrared region, the archaeological materials are mixed within the natural soil clusters and no significant patterns are observed (Chapter 5.4). Therefore, as suggested in Chapter 4.4, the wavelength range 400 – 1000 nm was also used for the Hungarian sites. In the Hungarian sites the first two PCs account for more than 90% of the total variance indicating that the soil spectra are extremely homogenous. The score plots of higher order PCs showed that they usually do not provide additional information to separate archaeological materials from natural soils.

Since the aim of the thesis is to develop a quantitative methodology, further investigations were made to numerically express the degree of 'archaeological features' of the reflectance spectra. To do this, the D calculation method was developed based on the PC of natural soils (Chapter 4.8). PC of a group of natural soils was used to recalculate a spectrum with characteristic features of natural soil. By doing so, the difference between the original spectrum and the recalculated spectrum showed different D values. Using this D calculation, average D values for archaeological materials (D_{arch}) and natural soils (D_{nat}) were calculated and the ratio (D_{ratio}) between them was investigated. It was expected that spectra of archaeological soils should give D_{ratio} values larger than 1. The analysis tried to define a universal D_{ratio} value which can be applied to any site, but did not work as well as the Italian sites.

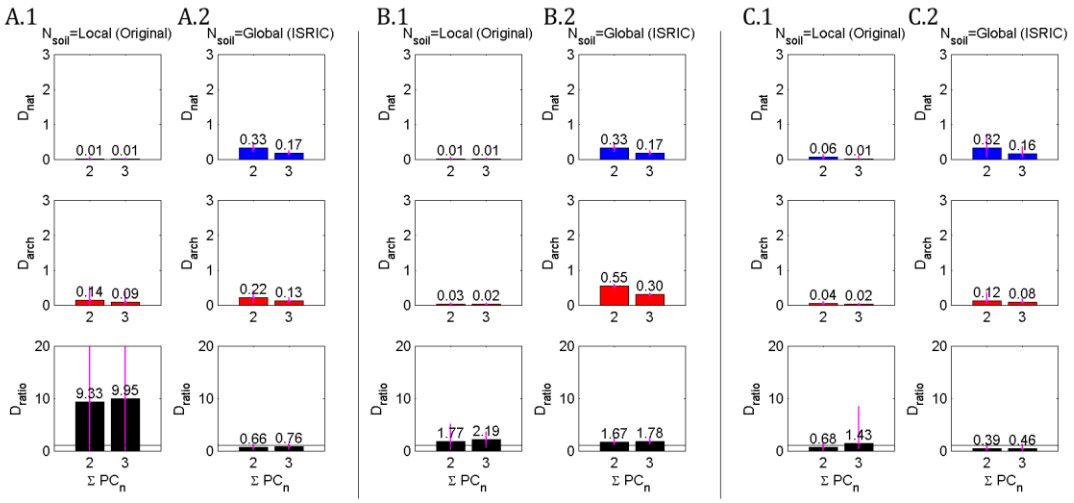


Figure 5.9.1 Final D value (D_{nat}, D_{arch} and D_{ratio}) results for sites 1 (A), 2 (B) and 3 (C), Hungary. Spectra are resampled by every tenth-nanometre. Refer to Appendix 7 for score plots with D_{ratio} values.

Similar to the results in Italy, for the sites in Hungary, the D_{ratio} values are the largest when N_{soil} is used from only the original site (Figure 5.9.1). Site 1 shows highest D_{ratio} value (around 9) when N_{soil} is used from the original site. However, when N_{soil} is used from the spectra collected from the global natural soils (ISRIC spectral library), the D_{ratio} drops below 1 indicating that the archaeological soils in this site is spectrally more similar to the global natural soils. Site 2 provides a constant D_{ratio} value of around 1.6 to 2 regardless what N_{soil} group is used.

Sites 1 and 2 share the same natural soils (notice that the D_{nat} values of the two sites are identical), but the types of archaeological remains make a great difference for the D calculation. D_{ratio} for site 3 is only larger than 1 when $\sum_1^3 PC$ is used for local N_{soil}. The overall results suggest that for the Hungarian sites, adding up to the third PC provides better D_{ratio} results although PC3 accounts for less than 1% of the total spectral variability (Chapter 5.5). However, the Hungarian sites showed promising results for applying the D-calculation method to distinguish the origin of soil colour changes (whether it is naturally occurring soil colour difference or not). Here, real soil mark feature gave a D_{ratio} value larger than 1 when local N_{soil} was used. The naturally occurred soil colour change, which the archaeologists thought it was a soil mark feature, showed D_{ratio} values smaller than 1 indicating that there are no anthropogenic activities involved in this soil colour difference.

Overall, results indicate that the D calculation method may not work in some buried archaeological remains (especially ditch formations). This suggests that further investigation of the D calculation is needed, applying the method to various archaeological structures at diverse environmental conditions.

Chapter 6 Conclusions and outlook

6.1 Overall summary and conclusion

This thesis aims to build up a methodology which can identify spectral signatures of archaeological remains among natural soils. The method should also identify archaeological soil horizons or soil marks which may not be visible in colour.

Among various archaeological prospection methods, soil colour is one of the most important and effective aspects that indicate traces of anthropogenic activities (Aston et al., 1998; Canti and Linford, 2000; James, 1999; Jones and MacGregor, 2002) without applying complicated analysis. Soil colour is influenced not only by the mineral content but also by chemical processes (Clark, 1999; Donoghue and Shennan, 1988; Viscarra Rossel et al., 2006a,b,c), such as the chemical state of iron, and therefore can become more distinct in archaeological features. However, sometimes such colour difference is either not distinctive or it is unclear whether it is due to anthropogenic activities. It is known that archaeological remains have different, but not unique, spectral features to the natural soils (Cavalli et al., 2009) which are influenced by the soil minerals. Nevertheless, there is only a limited understanding of the links between the chemical composition of soils and archaeological materials (Oonk et al., 2009a). In this thesis, it was difficult to distinguish archaeological remains from natural soils by investigating the elemental composition using XRF (Chapter 4.2.1 and Chapter 5.2.1). Using a CI (colour index) or RI (Redness index) allowed a good identification of archaeological soils in Italy due to their reddish colours, but the methods only consider one or two wavelengths (Chapter 4.2.2 and Chapter 4.2.3). Therefore this thesis tried to analyse the soil spectra statistically using PCA. PCA is known to be a powerful statistical tool in soil science (Linker et al., 2005; Singh et al., 2011; Viscarra Rossel, 2008a), therefore it will be further developed and adapted for the distinction of archaeological material from different soil types.

For this study, five buried archaeological sites were investigated from two environmentally different countries: Italy and Hungary (refer to Figure 3.1.1). The sites in Italy contained complex soil compositions and the sites in Hungary had fairly homogenous soils (Hungarian loess). Thus, it was assumed that the sites in Italy would provide more difficulties in identify-

ing and classifying archaeological features than the Hungarian sites. For these sites, *in situ* soil spectra measurements were recorded with the ASD FieldSpec Pro FR spectrometer (350 – 2500 nm) with artificial halogen lighting (Chapter 3.1.2).

This thesis introduces a modified PCA method which spectroscopically separates archaeological features among natural soils. The initial application of the standard PCA to the collected soil spectra from Italy and Hungary gave promising results. Therefore, several attempts were made to further improve the PCA and to represent and analyse the PCA results in a quantitative way. However, not all attempts worked as well as expected (base PC, regression line, etc., refer to Chapter 3.3.4). In the most promising method, a difference value (D value) between the original spectrum S and a recalculated spectrum S' using the principal components of natural soils provided very useful results. The principal components of natural soils (N_{soil}) were gathered from a various group of natural soils ranging from local to global soils gathered from the ISRIC online spectral library (ICRAF and ISRIC, 2010). When ISRIC spectra are used for N_{soil}, all spectra are resampled at every tenth-nanometre (refer to Chapter 3.3.2.1 for the detailed reasoning behind this). This method assumes that although the spectral features of natural soils vary depending on the soil type and location, they are spectrally more similar to each other than to the spectra of archaeological remains. Therefore, D_{nat}, the D value for natural soils, should be close to 0, while D_{arch}, the D value for archaeological remains should be larger than the D_{nat} values. Therefore, if the spectrum measured shows a D_{ratio} ($D_{ratio} = \frac{D_{arch}}{D_{nat}}$) larger than 1, it probably represents an archaeological material (refer to Chapter 3.3 for detailed calculation).

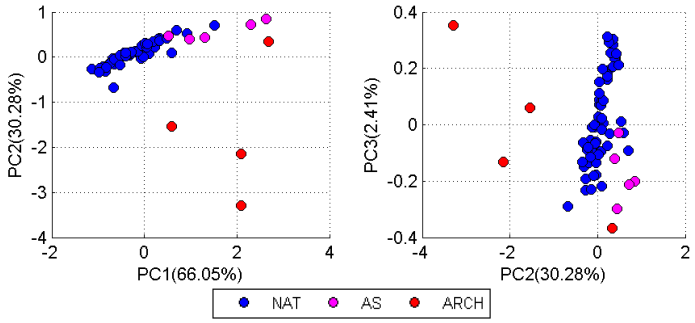


Figure 6.1 PCA score plot for the spectra gathered at site 1, Italy (wavelength of 400 – 1000 nm). In the left figure results forPC1 against PC2 are plotted and in the right figure results for PC2 against PC3 are plotted. For most sites, the PC1-PC2 score plots allowed a clearer distiction between archeological materials and soil spectra than score plots of higher order PCs. However, sometimes PC2-PC3 score plots allow a distinction of different archaeological materials as shown in this figure where burned materials are mostly on the left side of the natural soil cluster and the reddish archaeological stratum (including potteries) is on the right side of the cluster.

As shown in Figure 6.1, most of the score plots of the first two principal components (PC1 against PC2) present a clear difference between natural soil spectra and archaeological spectra. However, since water absorption bands dominate the near infrared region (1400, 1900, 2200 nm) of soil spectra, wavelengths beyond 1000 nm are not considered in this analysis. In some cases, spectral characteristics of archaeological remains could be detected in this region (as shown in Figure 4.4.1.G 2000 – 2200 nm, site 1 of Italy) but for now, it remained difficult to

extract useful information in that spectral range. Therefore, the wavelength range of 400 – 1000 nm was considered for the D calculation analysis. One interesting and important finding is that the PCA of the gathered soil spectra does not strongly depend on the soil colour (Chapter 4.4). This indicates that from the PCA additional information, compared to the simple use of 'colour', can be obtained.

The D calculation only uses up to the third principal components since the first three principal components contain more than 90% of the variance. Table 6.1 summaries the main D_{ratio} values for the five archaeological sites investigated in this thesis.

Table 6.1 Average D_{ratio} values for all the archaeological sites investigated in this thesis. The table shows results for different N_{soil} (local, another country, both countries and the ISRIC spectral library) and for different numbers of PCs considered (ΣPC).

	N_{soil}							
	Local (original site)		**Another Country (IT or HUN)**		**both countries together**		**ISRIC spectral library**	
	$\sum_1^2 PC$	$\sum_1^3 PC$	$\sum_1^2 PC$	$\sum_1^3 PC$	$\sum_1^2 PC$	$\sum_1^3 PC$	$\sum_1^2 PC$	$\sum_1^3 PC$
Italy								
Site1	10.66	9.74	1.68	2.23	3.40	3.14	2.04	2.08
Site2	3.68	3.30	1.60	2.01	1.90	1.89	1.59	1.54
Hungary								
Site1	9.28	9.50	0.63	0.61	1.51	2.05	0.66	0.76
Site2	1.71	2.08	1.37	1.45	3.20	1.52	1.67	1.78
Site3	0.67	1.43	0.39	0.36	0.79	1.04	0.39	0.46

Simple description of the archaeological sites.

- *Site1 (IT): Kitchen formation with red archaeological stratum in the soil profile*
- *Site2 (IT): Kitchen formation where potteries and reddish archaeological soils found at the floor of the excavated pit.*
- *Site1 (HUN): Ditch formation of fallen wall structures.*
- *Site2 (HUN): Rubbish pit formation with visible soil mark feature.*
- *Site3 (HUN): Ditch formation (strong yellowish colour) located on top of ancient dune feature.*

The average D_{ratio} results suggest that it is better to use both $\sum_1^2 PC$ and $\sum_1^3 PC$ for the D value calculation since they often show different results (sometimes $\sum_1^3 PC$ gives better results and vice versa). In particular, when N_{soil} does not contain any soils from the local site (ex. N_{soil} form another country or ISRIC), $\sum_1^3 PC$ tends to give higher D_{ratio} values than $\sum_1^2 PC$. This suggests that when natural soil spectra were not gathered around the archaeological site, adding up to the third PC is.

Overall, D_{ratio} values are highest when N_{soil} is collected around the archaeological site (local).

Results for N_{soil} from two countries or the global ISRIC data base also provide satisfactory D_{ratio} values in many cases, but the success of the method depends on the type of the archaeological materials. The method works especially well with materials from kitchen formations, probably because of the exposure of the materials to fire (Arcenegui et al., 2008). In particular, the red soil colour of the archaeological stratum in the investigated kitchen formation might be the result of fire (Matneya et al., 2014). Also pit fillings (site 2 of Hungary) produce high D_{ratio} values probably since they contain high amounts of organic carbon, nitrogen, or phosphorus compared to the adjacent natural soils (Lauer et al., 2014; Slager and Van de Wetering, 1977). However, for ditch features, depending on what was fallen inside the ditch (can be infilled with collapsed surrounding soil or walls which are still 'natural' soils) the identification might be difficult using the D calculation methods.

The following specific D_{ratio} values are obtained for three different archaeological remains:

- With N_{soil}=original
    - Kitchen formations: D_{ratio} value larger than 3,
    - Pit features: D_{ratio} around 1.7
    - Ditch features: D_{ratio} larger than 1.
- With N_{soil}= ISRIC
    - Kitchen formations: D_{ratio} value around 2
    - Pit features: D_{ratio} around 1.7
    - Ditch features: D_{ratio} varies around 0.4 to 0.7

In the beginning, it was assumed that for the archaeological sites in Italy it would be more difficult to identify archaeological materials due to the complex background natural soil composition. However, the results derived in this thesis suggest that the degree of variability of the natural background soils is not a big problem as long as the collected natural soil spectra cover this variability. If no natural soil spectra from the local site are available, soil spectra N_{soil} from the ISRIC spectral library can be used. However, in this case the D_{ratio} values will be lower than for the local soil samples. Based on the results presented above, the more important factor is what kind of buried remains one prospected.

In this thesis, the method also successfully identified surface features (soil marks) and distinguished soil colour changes due to natural and anthropogenic activities (Chapter 5.8). One limitation of the modified PCA method is that high D_{ratio} can also be obtained for soils, which were influenced by non-archaeological soil activities, e.g. recent anthropogenic activities. But this limitation might also be turned into an advantage: it leads to the possibility to apply the method to identify soil contaminations and other environmental factors influencing the soil.

One important thing to notice is that the PCA score plots and D_{ratio} values did not show a strong difference when the spectra were smoothed by convolution kernels of 10 to 200 nm FWHM, which indicates that broadband features are the main characteristic difference between natural and non-natural (in this case archaeological) soils. This might suggest that 'high'-resolution spectrometer (finer than 3 nm resolution) might eventually not be needed and, therefore, could lead to simpler instrumentation and shorter measurement times.

Overall, the modified PCA method using the D value calculation worked well for the archaeological sites investigated in this thesis. It is of course unclear how well this method will work

at other archaeological sites. However, according to the results of this thesis, which are based on different natural soil samples, one can expect promising results for other archaeological sites.

6.2 Outlook

6.1.1 Investigate spectra beyond 1000 nm

In the near-infrared region (beyond 1000 nm), soil spectra are dominated by water absorption bands (refer to Chapter 3.2.2). However, this thesis suggested that there might be some features within these bands which can still be used to identify archaeological remains (Figure 4.4.1). To further test this possibility, archaeological sites in arid regions, which contain less soil moisture, should be investigated.

6.1.2 Application to other archaeological sites

The D calculation method worked well when N_{soil} is used from the natural soils gathered around the archaeological site. When N_{soil} is from the ISRIC spectral library, among five archaeological sites investigated, still three sites provided D_{ratio} value larger than 1. This is an interesting finding since it suggests that natural soils from the spectral library have common spectral features to the natural soils from Italy and Hungary (however, they did not show spectral similarities in their first three principal components, refer to Figure 3.3.6). The method should be tested for other archaeological sites to further verify it. Also, it will be useful to apply the method to archaeological strata or soil marks which are not visible by eye. By making such investigations, it will in particular be possible to refine the thresholds of the D_{ratio} values for different archaeological materials and different locations and environmental conditions.

6.1.3 Instrumental improvements

As shown in Chapter 4.9 and Chapter 5.7, the developed method works well even if the original spectral resolution is smoothed by different kernels (FWHM of 10 to 200 nm). This finding has important consequences or the instrumental properties. For the existing instrumental set up the acquisition time for a single spectrum is about 1 to 3 s in order to achieve a sufficient signal to noise ratio. The results of this thesis indicate that for archaeological applications the spectral resolution can be reduced by a factor 10 to 20 without a significant loss of information. Consequently, the integration time can be reduced by the same factor. This can even lead to the possibility of performing continuous 1D or 2D imaging measurements with such reduced acquisition times. Besides the reduction of the measurement time, instruments with reduced spectral resolution will also be smaller and cheaper.

6.1.4 Application to airborne or satellite images

The method was so far applied to '*in situ'*, that means at the sides of excavated pits, measurements but it can also be modified and applied to airborne and satellite images. Such applications would in particular help to locate potential buried archaeological remains without time-consuming manual inspection of entire airborne or satellite images. However, it has to be noted that the application to such airborne or satellite images is restricted to certain periods depending on the vegetation coverage and ploughing season. Also, more investigations are needed to quantify the effect of using sun light instead of an artificial light source.

6.1.5 Investigate D values for soil mark features

Chapter 5.7 provides D values for a soil mark feature observed in the Hungarian site. The results show that the D calculation method can only identify this feature if N_{soil} is gathered from the local site. However, since only one soil mark feature was measured, it is not clear how representative the results are for other soil marks. Soil marks depend in particular on the ploughing conditions which transfers archaeological material to the top surface. In addition, depending on the measurement period, the soil marks might become vague due to weathering effects. Thus, more investigation is needed to check the validity of the method to soil mark features.

6.1.6 Application to other environmental issues

The underlying assumption of the D calculation method is that spectral features different to natural soils can be identified. Thus, this method can also be applied to identify soils which were modified by other environmental processes such as soil contaminations.

Acronyms

Ap – Top soil horizon influenced by constant ploughing.

ARCH – Archaeological materials such as potteries.

AS – Archaeological soils.

Bt – Subsoil horizon with high clay content.

C – Parent material (bed rock).

D – Difference value calculated by the modified PCA method developed in this thesis (Chapter 3.3.2)

D_{arch} – D value of archaeological spectra (both AS and ARCH).

D_{nat} – D value of natural soil spectra.

D_{ratio} – Ratio between D_{arch} and D_{nat}.

HUN – Soil spectra collected from the Hungarian sites.

ISRIC – Soil spectra from ICRAF and ISRIC (2010) soil spectra library.

IT – Soil spectra collected from the Italian sites.

IT+HUN – Soil spectra collected from Italy and Hungary.

NAT – Natural soils.

NIR – Near-infrared.

NPC – Principal components of a group of natural soils (N_{soil}).

N_{soil} – A group of natural soil used for the D-calculation method (Chapter 3.3.2.2).

PC – Principal component.

PCA – Principal component analysis.

VIS – Visible spectra.

XRF – X-ray fluorescence.

Acknowledgements

Appendix

Appendix 1

XRF results for sites 1 and 2, Italy.

Pit	Horizon	Zr	Zr Error	Sr	Sr Error	Rb	Rb Error	Th	Th Error	As	As Error	Zn	Zn Error	Cu	Cu Error	Ni	Ni Error
ARCH	Ap	232.74	10.21	91.26	6.69	116.49	9.10	9.44	7.37	16.54	7.70	54.34	16.96	36.18	27.40	77.90	45.81
ARCH	AS	237.22	9.21	103.52	6.33	143.16	8.94	9.20	6.62	22.21	7.25	59.76	15.63	44.70	20.23	122.57	57.26
ARCH	ARCH	228.66	10.36	123.18	7.70	151.40	10.28	<LOD		22.00	7.50	57.88	17.54	55.84	23.69	76.48	58.14
ARCH	C1	270.33	7.20	101.97	4.65	169.69	7.10	10.04	4.30	22.43	5.38	67.08	11.84	34.44	14.57	96.22	31.38
Pit1	AP	173.82	8.03	83.85	5.68	129.39	8.33	7.72	6.77	17.68	6.51	61.66	15.09	37.80	22.85	<LOD	
Pit1	C	235.15	6.49	108.20	4.54	161.02	6.60	<LOD		18.81	4.79	53.11	10.48	29.35	13.40	105.39	30.07
Pit2	AP	237.70	7.92	93.80	5.20	154.97	7.79	<LOD		19.19	6.11	60.54	13.36	42.10	19.02	85.36	36.09
Pit2	BT	259.23	8.12	88.77	5.05	167.28	8.15	10.96	5.53	21.86	5.95	55.48	12.97	39.91	16.82	81.66	35.67
Pit2	C	269.21	8.26	96.18	5.22	168.29	8.23	8.42	6.59	23.92	6.16	62.40	13.53	34.73	16.65	122.55	48.77
Pit3	AP	246.53	6.73	93.74	4.35	117.03	5.86	7.16	3.81	16.54	5.26	76.78	11.79	43.35	14.33	63.80	34.10
Pit3	BT	263.08	9.48	98.98	6.12	144.81	8.89	9.69	7.41	27.35	6.76	55.06	14.86	38.97	19.52	85.26	49.20
Pit3	C1	278.98	9.73	107.14	6.33	155.68	9.11	12.55	6.52	23.34	7.43	56.08	14.99	51.23	20.32	118.43	42.76
Pit3	C2	280.78	9.80	84.53	5.80	154.74	9.22	9.94	6.70	28.34	7.33	69.49	16.27	53.12	20.51	102.07	56.75
Pit4	AP	237.57	7.57	81.82	4.69	110.15	6.60	8.46	5.77	19.89	5.79	46.92	11.78	40.50	16.06	72.81	40.54
Pit4	C	282.20	7.13	89.74	4.30	138.33	6.39	10.43	4.48	23.45	5.10	58.91	11.22	51.36	14.78	71.02	33.32
Pit5	AP	228.92	7.51	89.07	4.92	132.09	7.14	6.94	5.96	19.92	5.63	58.75	12.73	42.34	16.34	84.41	45.16
Pit5	BT	261.46	6.75	96.26	4.31	151.41	6.46	10.40	4.02	21.76	5.20	51.31	10.48	49.36	14.20	75.06	29.19

Pit	Horizon	Co	Co Error	Fe	Fe Error	Mn	Mn Error	V	V Error	Ti	Ti Error	Ca	Ca Error	K	K Error	Ba	Ba Error	Cs	Cs Error
ARCH	Ap	<LOD		39385.54	554.14	1285.68	134.05	114.18	97.91	5376.54	272.03	6296.25	331.24	17862.01	668.91	334.39	99.00	<LOD	
ARCH	AS	<LOD		46559.92	536.99	1074.75	113.24	<LOD		5410.46	289.93	8449.71	386.80	21135.05	748.36	319.29	114.30	<LOD	
ARCH	ARCH	<LOD		42941.83	585.66	1013.85	125.47	<LOD		5027.49	280.60	20503.69	564.33	20433.95	730.50	354.83	118.44	<LOD	
ARCH	C1	<LOD		47156.17	400.39	949.12	80.67	<LOD		5893.13	304.16	7107.17	365.51	20624.11	748.96	273.92	97.85	<LOD	
Pit1	AP	<LOD		39206.52	486.98	711.54	94.64	123.51	87.89	3744.76	228.15	16866.47	484.49	16368.36	625.76	449.60	103.72	88.70	50.67
Pit1	C	<LOD		43835.25	368.47	810.04	72.69	132.97	79.06	5258.52	271.86	11476.53	438.81	20935.15	737.43	329.32	94.43	<LOD	
Pit2	AP	<LOD		45649.00	454.25	995.17	95.31	169.23	100.86	4844.14	263.09	6987.53	349.27	20195.85	709.55	335.81	98.70	69.73	55.32
Pit2	BT	<LOD		46480.54	457.63	873.32	90.03	127.97	103.25	4989.99	267.66	6501.32	338.64	19781.06	704.89	335.30	96.72	<LOD	
Pit2	C	367.21	215.88	49509.39	473.05	1427.71	103.87	144.15	94.63	5401.72	281.34	6386.96	346.62	21379.87	747.23	304.43	96.34	<LOD	
Pit3	AP	213.06	163.90	41056.55	364.05	1302.77	87.06	<LOD		5235.47	270.86	5375.28	309.15	16905.20	649.56	309.12	98.03	<LOD	
Pit3	BT	<LOD		44979.65	522.19	1133.36	113.10	<LOD		5706.27	279.96	5114.14	311.70	20824.28	720.61	314.24	97.54	<LOD	
Pit3	C1	<LOD		42770.70	510.35	1894.62	139.79	<LOD		6141.46	294.33	5099.27	314.46	21748.51	736.56	309.38	96.65	<LOD	
Pit3	C2	<LOD		50278.49	558.43	1331.00	118.55	126.14	108.97	5170.90	276.37	4657.34	304.27	20091.37	719.10	357.47	98.33	<LOD	
Pit4	AP	188.18	184.95	38872.81	406.07	1289.04	99.23	144.89	103.60	5067.30	270.69	4835.01	303.78	18281.93	681.82	219.90	107.53	<LOD	
Pit4	C	<LOD		44553.58	380.82	1770.77	98.78	142.16	112.10	5646.69	283.97	3844.47	283.41	19488.94	707.20	372.37	99.40	<LOD	
Pit5	AP	<LOD		41393.57	421.90	1041.44	92.77	165.13	111.93	5632.89	289.58	7292.66	366.18	20074.87	728.89	312.56	94.66	<LOD	
Pit5	BT	198.55	154.69	44255.35	369.23	1051.13	79.05	151.63	112.21	6463.28	308.55	6413.05	351.38	20752.51	746.67	304.65	93.13	<LOD	

Appendix 1.1 XRF results for site 1, Italy. LOD stands for limit of detection, and therefore, <LOD indicates that the elements were very small or does not exits.

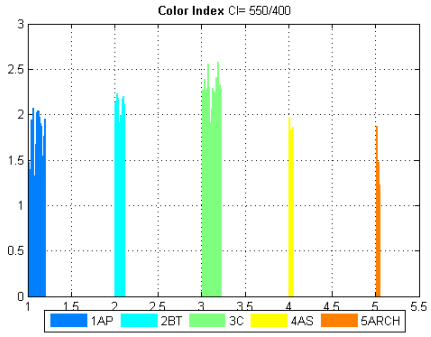
Pit	Horizon	Zr	Zr Error	Sr	Sr Error	Rb	Rb Error	Th	Th Error	As	As Error	Zn	Zn Error	Cu	Cu Error	Ni	Ni Error
ARCH	AP	158.81	9.16	119.6	7.74	113.31	9.24	<LOD		13.76	7.44	67.54	17.97	56.81	24.41	94.07	47.88
ARCH	ARCH	164.41	9.378	122.948	7.894	102.034	8.92	<LOD		20.008	7.532	61.582	17.992	46.094	23.988	72.28	66.064
NAT	AP	148.46	5.97	109.64	4.97	134.8	6.68	9.64	4.11	18.95	5.02	67.97	12.21	38.88	15.51	64.95	31.13
NAT	bare soil	184.61	6.4	96.91	4.7	146.83	6.81	<LOD		18.67	5.08	57.76	11.41	<LOD		59.34	31.17

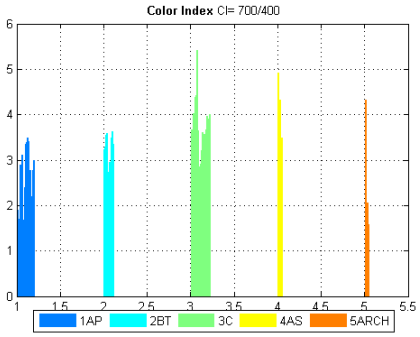
Pit	Horizon	Co	Co Error	Fe	Fe Error	Mn	Mn Error	V	V Error	Ti	Ti Error	Sc	Sc Error	Ca	Ca Error	K	K Error
ARCH	AP	<LOD		33587.06	526.78	551.63	100.75	<LOD		3952.68	248.16	<LOD		102129.2	1205.79	22995.11	790.79
ARCH	ARCH	370.13	252.664	34713.89	539.48	669.568	108.842	129.31	85.29	3532.4	247.756	49.31333	33.266	107835.3	1251.206	21041.99	775.596
NAT	AP	<LOD		35383.64	360.84	481.02	65.21	<LOD		3930.94	253.85	<LOD		91743.92	1172.91	27943.64	877.74
NAT	bare soil	<LOD		34084.67	352.68	447.49	63.5	<LOD		4043.88	253.13	38.25	24.5	84863.61	1114.59	29701.23	887.51

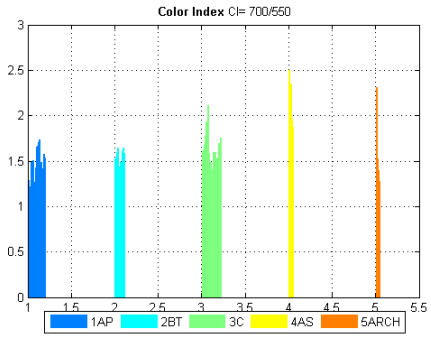
Appendix 1.2 XRF results for site 2, Italy. LOD stands for limit of detection.

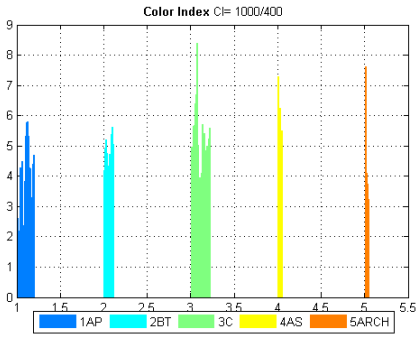
Appendix 2

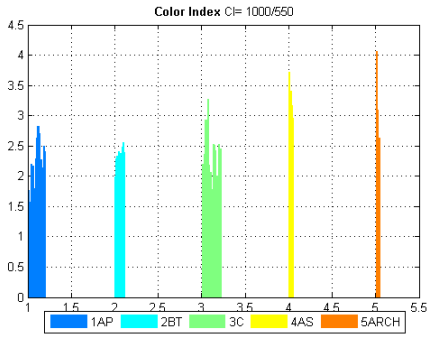
Colour indices of site 1 and site 2, Italy. Various wavelength combinations (λ_a and λ_b) are presented.

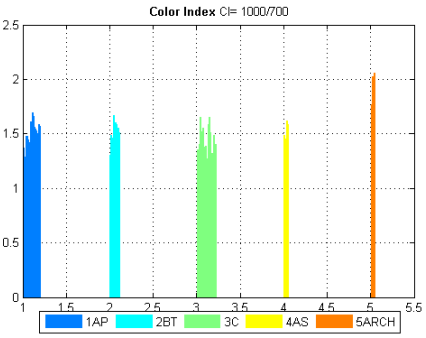


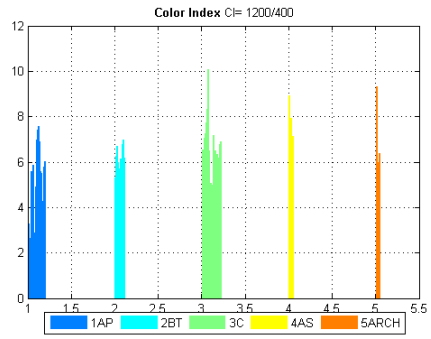


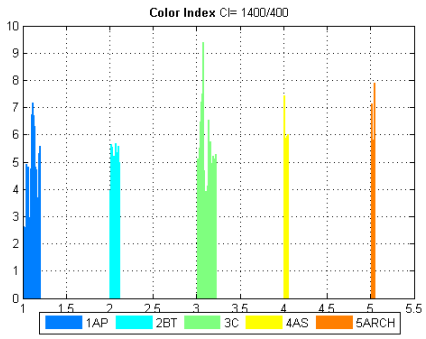


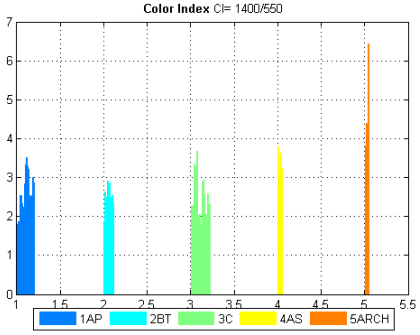


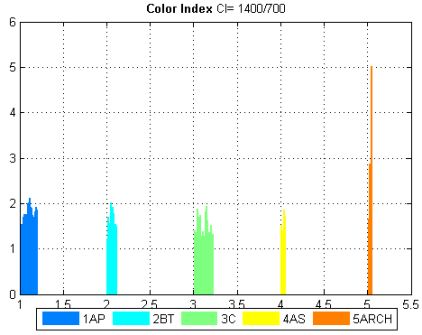


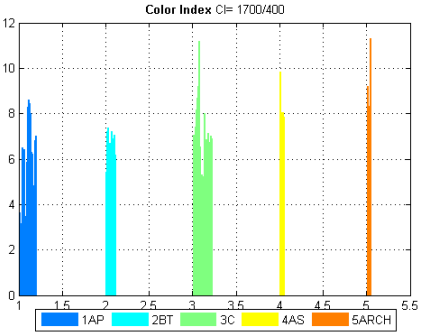


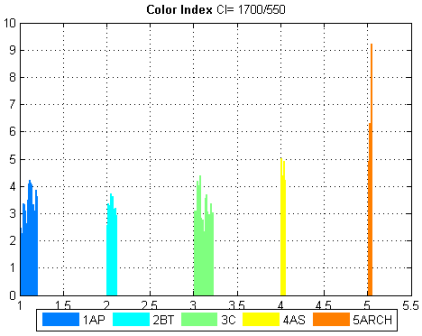


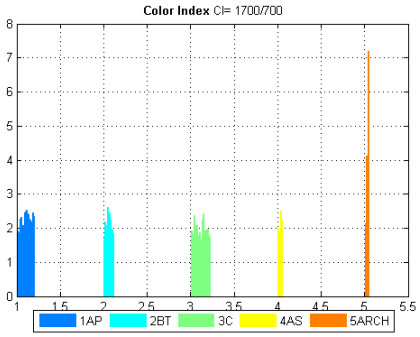


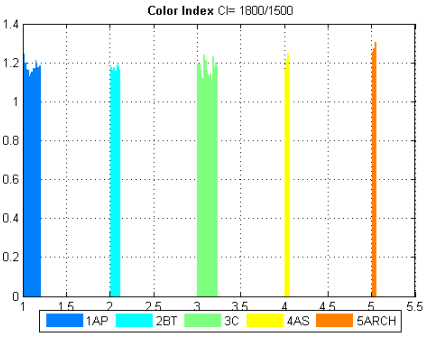

Color Index CI= 1400/550
1AP
2BT
3C
4AS
5ARCH

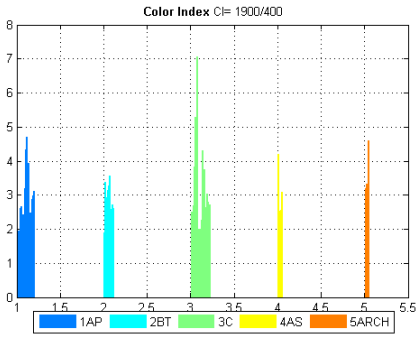

Color Index CI= 1400/700
1AP
2BT
3C
4AS
5ARCH

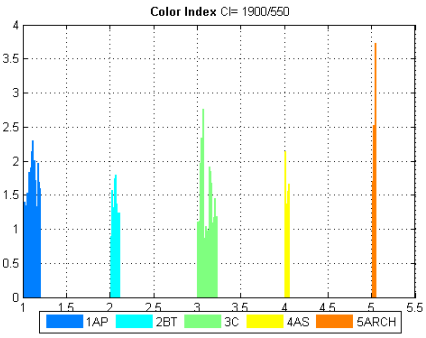

Color Index CI= 1700/400
1AP
2BT
3C
4AS
5ARCH

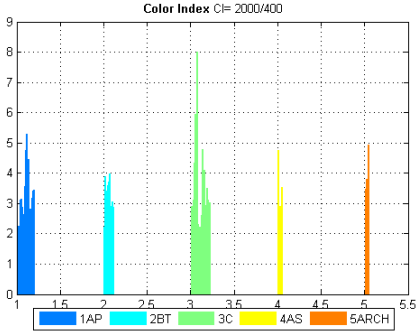

Color Index CI= 1700/550
1AP
2BT
3C
4AS
5ARCH

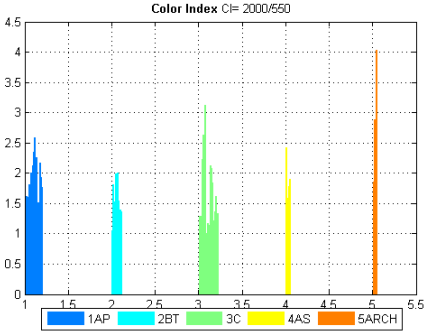

Color Index CI= 1700/700
1AP
2BT
3C
4AS
5ARCH

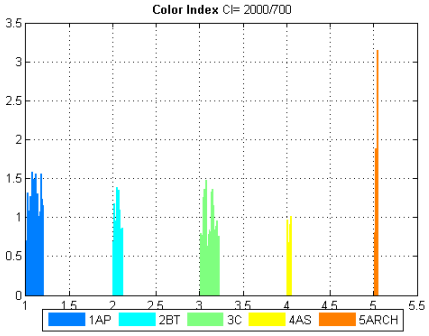

Color Index CI= 1800/1500
1AP
2BT
3C
4AS
5ARCH

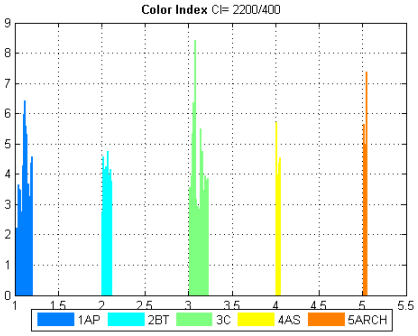

Color Index CI= 1900/400
1AP
2BT
3C
4AS
5ARCH

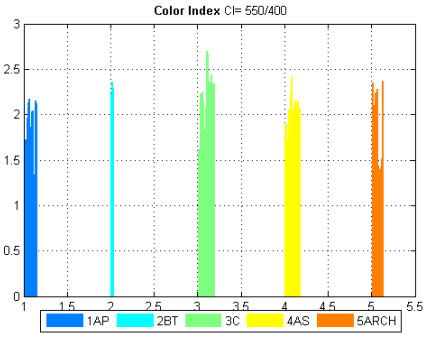

Color Index CI= 1900/550
1AP
2BT
3C
4AS
5ARCH

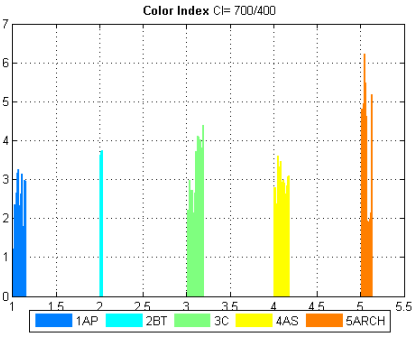

Color Index CI= 2000/400
1AP
2BT
3C
4AS
5ARCH

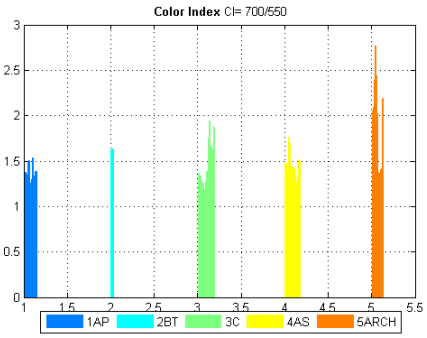

Color Index CI= 2000/550
1AP
2BT
3C
4AS
5ARCH

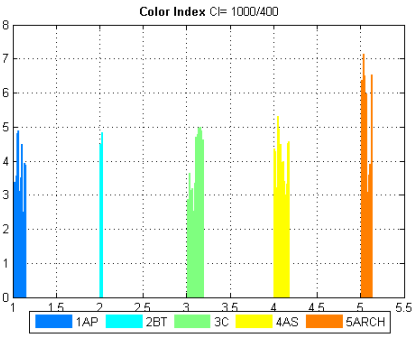

Color Index CI= 2000/700
1AP
2BT
3C
4AS
5ARCH

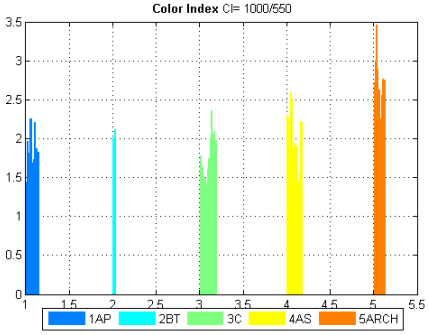

Color Index CI= 2200/400
1AP
2BT
3C
4AS
5ARCH

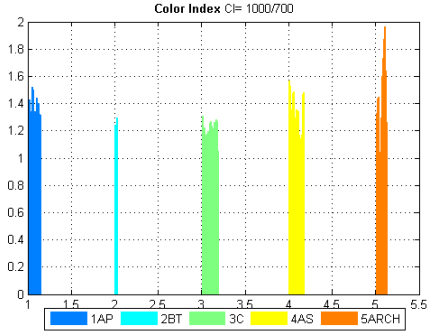

Color Index CI= 550/400
1AP
2BT
3C
4AS
5ARCH

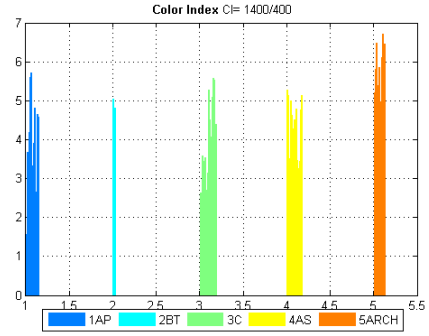

Color Index CI= 700/400
1AP
2BT
3C
4AS
5ARCH

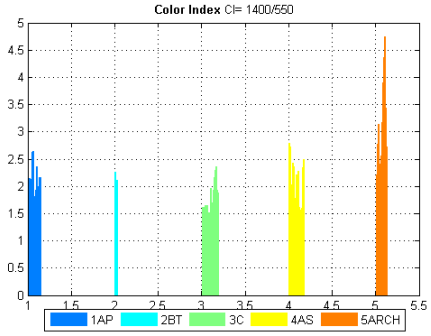

Color Index CI= 700/550
1AP
2BT
3C
4AS
5ARCH

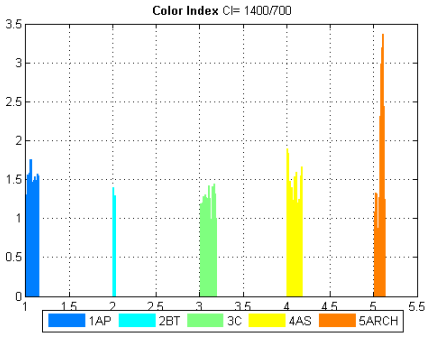

Color Index CI= 1000/400
1AP
2BT
3C
4AS
5ARCH

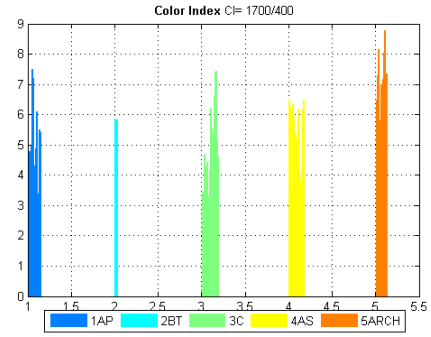

Color Index CI= 1000/550
1AP
2BT
3C
4AS
5ARCH

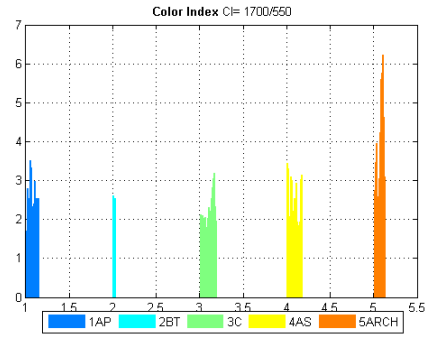

Color Index CI= 1000/700
1AP
2BT
3C
4AS
5ARCH

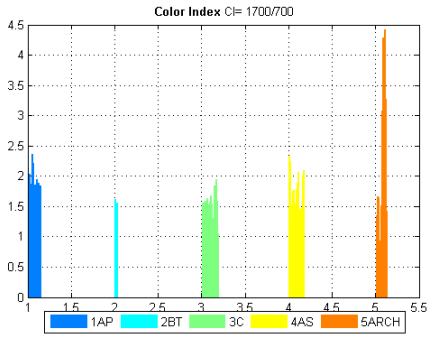

Color Index CI= 1400/400
1AP
2BT
3C
4AS
5ARCH

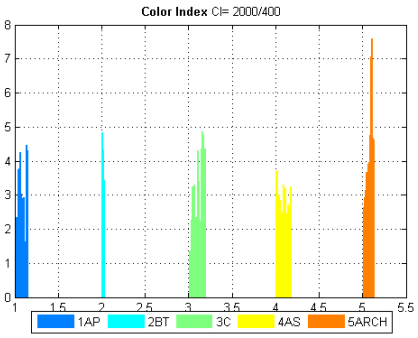

Color Index CI= 1400/550
1AP
2BT
3C
4AS
5ARCH

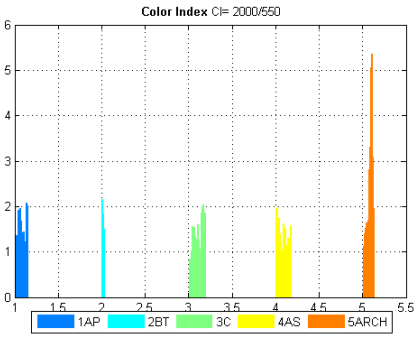

Color Index CI= 1400/700
1AP
2BT
3C
4AS
5ARCH


Color Index CI= 1700/400
1AP
2BT
3C
4AS
5ARCH


Color Index CI= 1700/550
1AP
2BT
3C
4AS
5ARCH

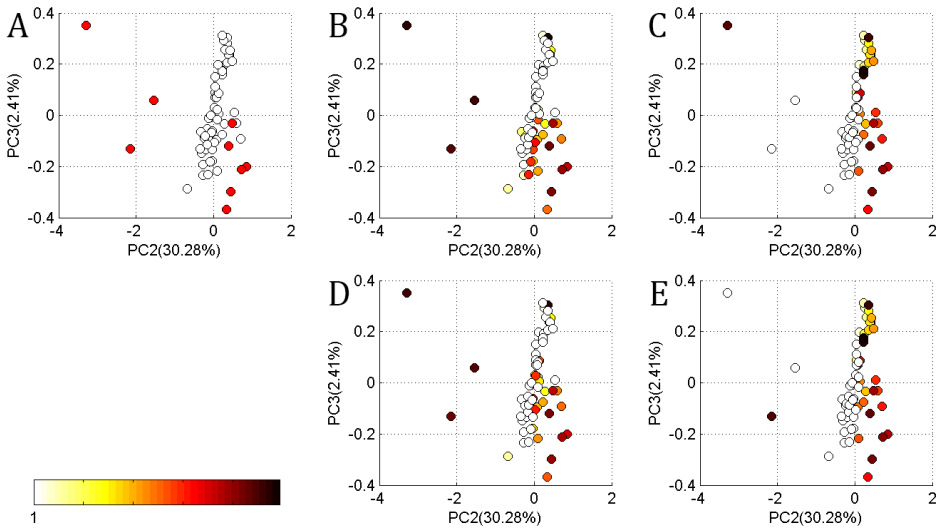

Color Index CI= 1700/700
1AP
2BT
3C
4AS
5ARCH


Color Index CI= 2000/400
1AP
2BT
3C
4AS
5ARCH

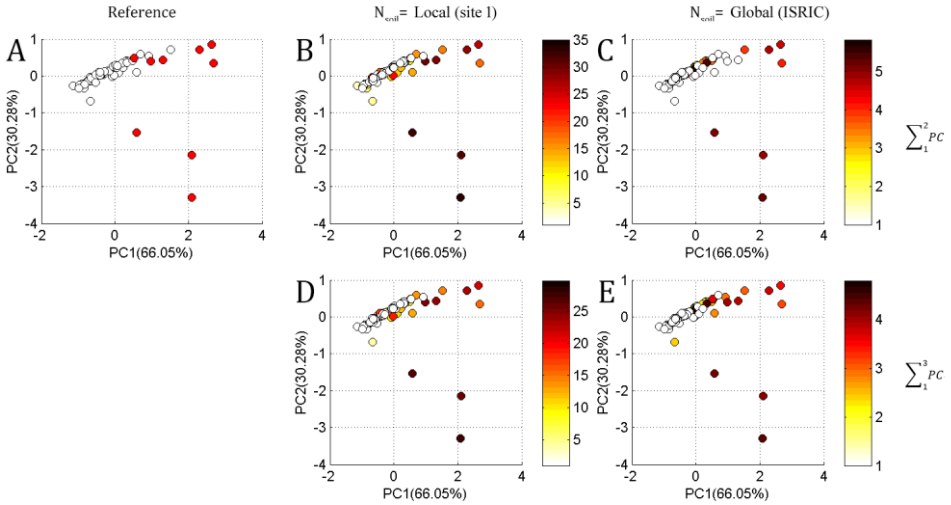

Color Index CI= 2000/550
1AP
2BT
3C
4AS
5ARCH

Appendix 3

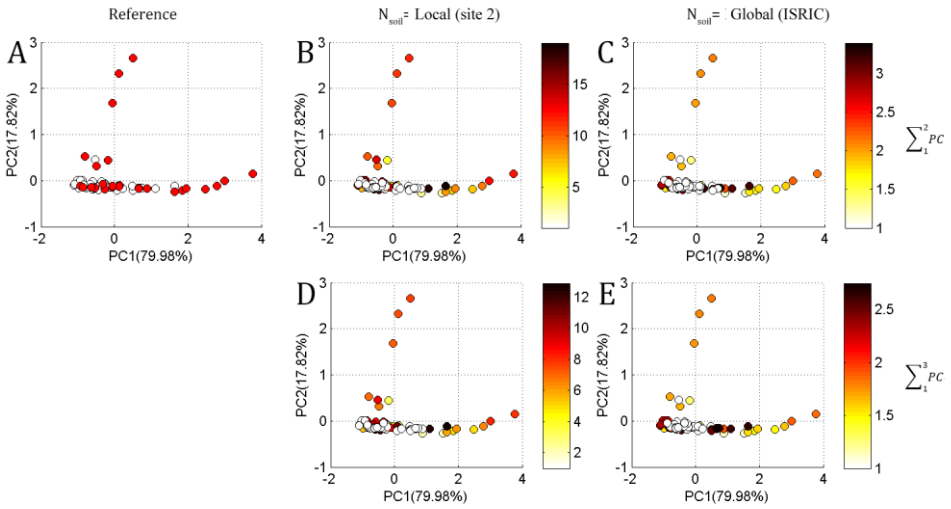
PCA score plot with marker colours indicating D_{ratio} values for the sites in Italy.



Appendix 3.1 D_{ratio} values represented as colour on the PC2-PC3 score plot of site 1. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) Results for N_{soil} used from the original site (site 1) with $\sum_1^2 PC$ (first two principal components used from N_{soil}). C) Results for N_{soil} used from Hungary with $\sum_1^2 PC$. D) Results for N_{soil} used from the original site (site 1) with $\sum_1^3 PC$(first three principal components used from N_{soil}). E) Results for N_{soil} used from Hungary with $\sum_1^3 PC$.



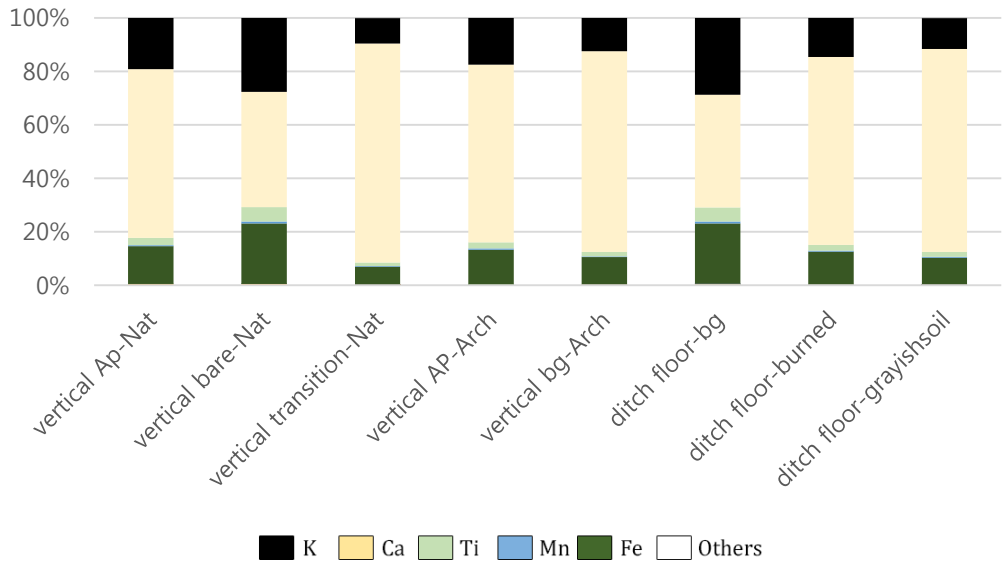
Appendix 3.2 D_{ratio} values represented as colour on the PC1-PC2 score plot of the site 1. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) N_{soil} is used from the original site (site 1) with $\sum_1^2 PC$. C) is the result when N_{soil} is used from ISRIC with $\sum_1^2 PC$. D) N_{soil} is used from the original site (site 1) with $\sum_1^3 PC$ and E) N_{soil} is used from ISRIC with $\sum_1^3 PC$. Spectra used for this plot result are resampled by every tenth-nanometre.



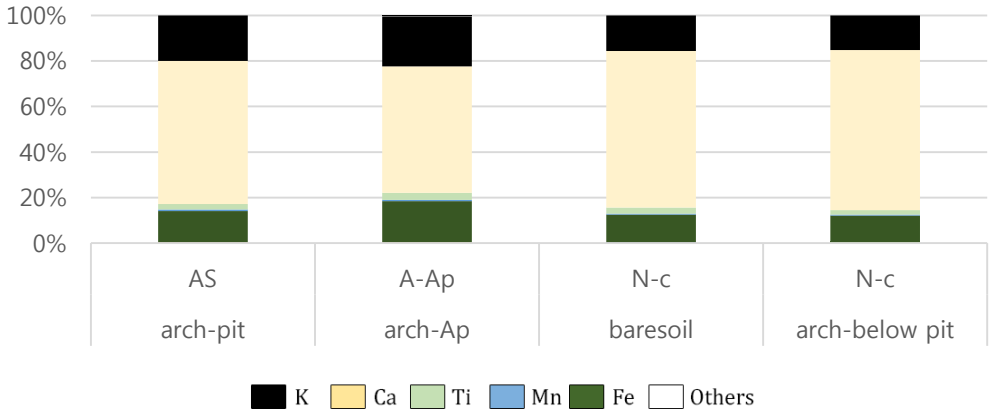
Appendix 3.3 D_{ratio} values represented as colour on the PC1-PC2 score plot of the site 2. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. As a reference, in Figure A), all data points of archaeological materials are marked as red. B) N_{soil} is used from the original site (site 2) with $\sum_1^2 PC$. C) is the result when N_{soil} is used from ISRIC with $\sum_1^2 PC$. D) N_{soil} is used from the original site (site 2) with $\sum_1^3 PC$ and E) N_{soil} is used from ISRIC with $\sum_1^3 PC$. Spectra used for this plot result are resampled by every tenth-nanometre.

Appendix 4

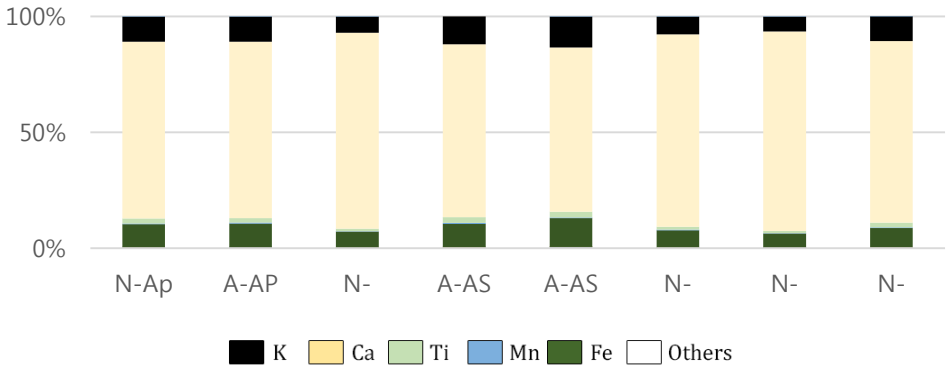
XRF results for the sites in Hungary.



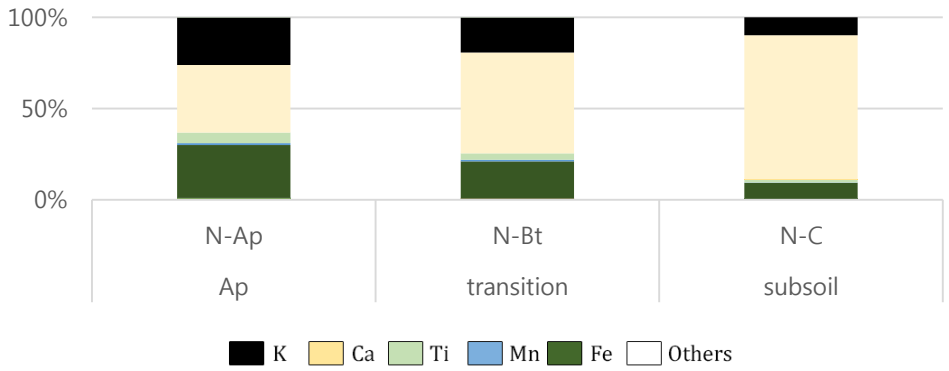
Appendix 4.1 Relative partitioning of the different elements for site 1, Hungary. For the exact XRF values refer to Appendex 5. Dominant elements are potassium (K), calcium (Ca), titanium (Ti), manganese (Mn) and iron (Fe).



Appendix 4.2 Relative partitioning of the different elements for site 2, Hungary. For the exact XRF values refer to Appendex 5. Dominant elements are potassium (K), calcium (Ca), titanium (Ti), manganese (Mn) and iron (Fe).



Appendix 4.3 Relative partitioning of the different elements for the archaeological soil pit in site 3, Hungary. For the exact XRF values refer to Appendex 5. Dominant elements are potassium (K), calcium (Ca), titanium (Ti), manganese (Mn) and iron (Fe).



Appendix 4.4 Relative partitioning of the different elements for the natrual soil pit in site 3, Hungary. For the exact XRF values refer to Appendex 5. Dominant elements are potassium (K), calcium (Ca), titanium (Ti), manganese (Mn) and iron (Fe).

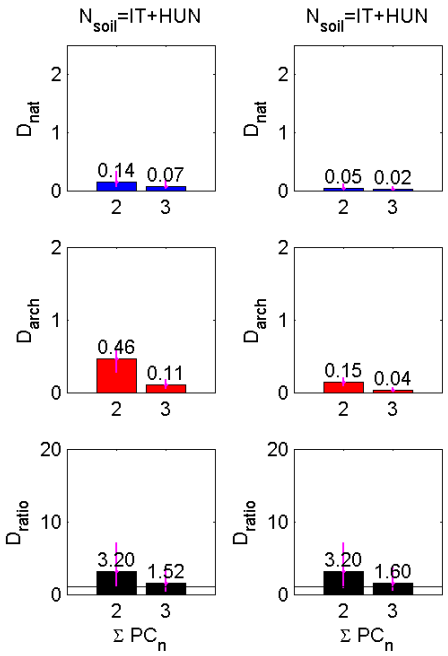
Site	Horizon	A/N	Zr	Zr Error	Sr	Sr Error	Rb	Rb Error	Th	Th Error	As	As Error	Zn	Zn Error	Cu	Cu Error
Site 1	AP	N	187.62	8.5	126.12	6.9	56.19	6.06	<LOD		<LOD		40.05	12.55	<LOD	
Site 1	C	N	113.32	4.725	156.26	4.845	52.605	3.765	8.075	2.745	8.275	3.5	31.675	7.815	21.15	13.655
Site 1	AP	A	117.42	4.6	128.8	4.44	60.55	3.92	<LOD		7.97	3.7	33.23	7.67	<LOD	
Site 1	AS	A	125.99	5.853333	162.5867	6.09	59.17667	4.86	5.115	3.883333	11.405	4.936667	28.79	9.323333	16.63	18.26333
Site 1	ARCH	A	117.54	4.96	188.69	5.52	65.36	4.23	6.62	2.85	8.82	3.8	23.42	7.6	<LOD	
Site 2	AS	A	72.39	6.16	125.57	6.78	49.02	5.62	<LOD		<LOD		24.78	11.33	36.65	18.09
Site 2	AP	A	93.91	6.7	95.71	6.17	55.77	5.91	<LOD		9.85	5.39	20.27	11.21	<LOD	
Site 2	C	N	112.16	4.51	122.635	4.345	50.665	3.565	4.81	2.48	6.55	3.36	19.595	6.96	21.19	11.3
Site 3	Ap	N	298.32	6.93	136.895	4.855	72.28	4.495	6.32	2.985	10.215	4.175	28.89	8.2	21.95	14.94
Site 3	C	N	277.975	8.236667	167.3183	6.326667	61.44333	5.101667	8.1075	4.143333	14.052	5.221667	34.465	10.23667	31.085	19.29667
Site 3	Ap	A	259.28	6.65	140.97	4.98	57.73	4.19	10.15	3.11	7.23	3.78	31.27	8.5	36.77	13.24
Site 3	AS	A	322.78	7.16	136.805	4.855	65.34	4.28	7.93	3.43	10.33	3.855	27.445	8.025	<LOD	0

Site	Horizon	A/N	Ni	Ni Error	Fe	Fe Error	Mn	Mn Error	Ti	Ti Error	Ca	Ca Error	K	K Error	Ba	Ba Error
Site 1	AP	N	60.68	37.81	12358.63	280.12	378.95	72.11	2413.12	176.63	55382.73	832.8	16822.43	623.07	<LOD	
Site 1	C	N	48.39	29.595	10999.18	172.59	327.78	44.27	2368.875	177.75	76687.14	920.085	14526.34	596.695	172.08	108.905
Site 1	AP	A	54.21	23.92	12295.97	178.33	392.64	46.05	2195.46	177.35	63122.83	887.65	16599.52	622.75	<LOD	
Site 1	AS	A	39.91	38.83333	12286.69	220.2533	384.9867	56.62667	2302.65	174.0133	70755.15	908.2967	14917.04	598.7633	234.48	118.11
Site 1	ARCH	A	36.58	24.38	13799.99	197.24	419.85	50.35	2421.6	183.58	79709.92	1000.25	16598.18	633.26	<LOD	
Site 2	AS	A	<LOD		9641.83	244.71	437.77	72.8	1691	155.16	44297.36	726.3	14135.75	556.32	<LOD	
Site 2	AP	A	<LOD		11823.48	277.3	354.25	71.15	2058.39	161.43	36354.42	659.14	14408.31	556.9	289.71	90.43
Site 2	C	N	41.16	23.45	9811.03	160.05	348.635	44.055	1863.4	159.235	57376.47	834.415	12709.56	544.76	<LOD	
Site 3	Ap	N	67.79	32.18	16433.97	218.02	580.6	56.48	3396.845	204.575	65604.52	877.54	16095.07	629.335	211.25	90.265
Site 3	C	N	68.345	38.34333	13981.01	244.8417	413.5	61.24833	2561.74	193.4517	132928.4	1296.45	14954.95	642.9967	242.792	101.2783
Site 3	Ap	A	81.97	27.22	13850.31	203.35	468.71	53.08	2629.9	192.25	103755.5	1154.18	14562.08	618.19	312.59	95.07
Site 3	AS	A	50.605	25.69	14271.59	203.78	427.53	51.095	3220.285	207.28	93133.3	1097.88	16147.87	642.825	173.47	108.6

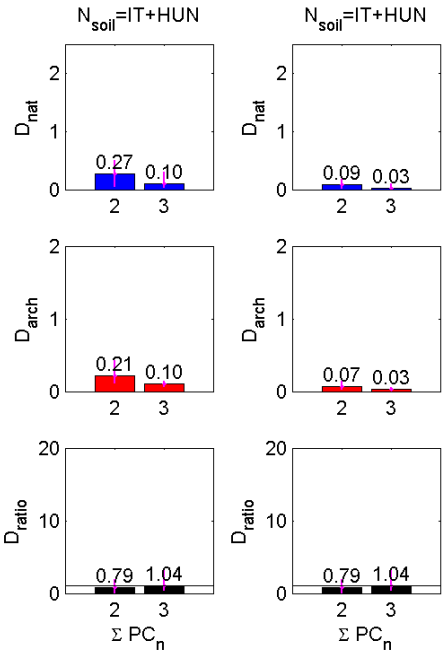
Appendix 5 XRF results for site 1,2 and 3 of Hungary. LOD stands for limit of detection.

Appendix 6

D values which present comparison between original spectra and resampled spectra by every tenth-nanometre for the Hungarian sites.



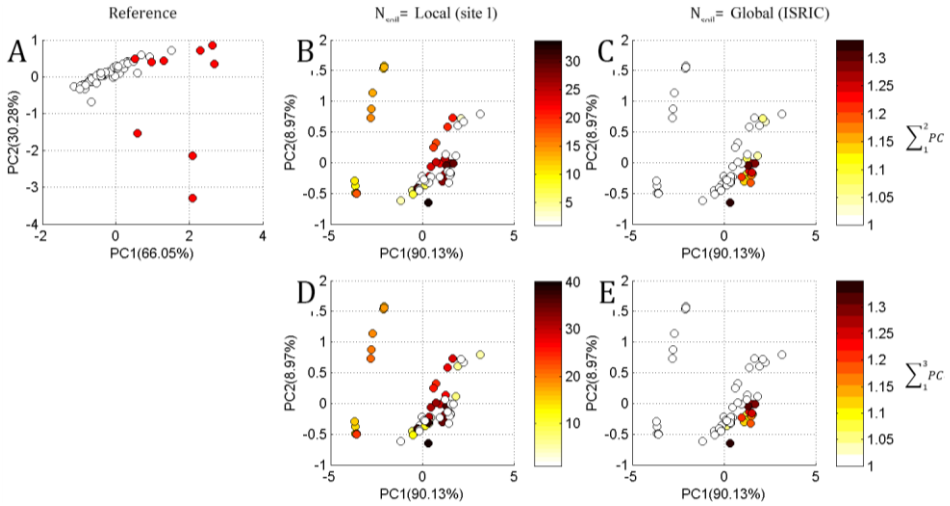
Appendix 6.1 D value (D_{nat}, D_{arch} and D_{ratio}) results for site 2 Hungary when N_{soil} is from both Italy and Hungary (IT+HUN).
A: Original spectra. B: Spectra resampled by a tenth-nanometre.



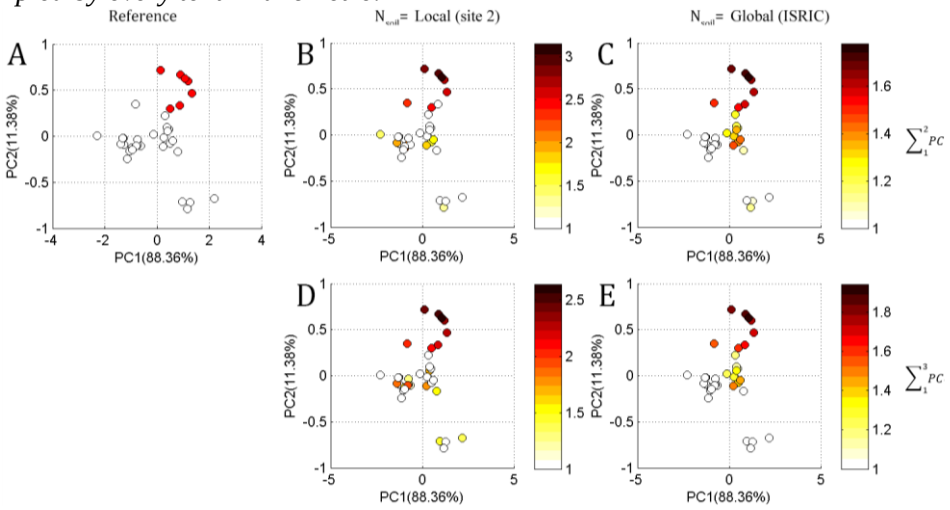
Appendix 6.2 D value (D_{nat}, D_{arch} and D_{ratio}) results for site 3 Hungary when N_{soil} is from both Italy and Hungary (IT+HUN).
A: Original spectra. B: Spectra resampled by a tenth-nanometre.

Appendix 7

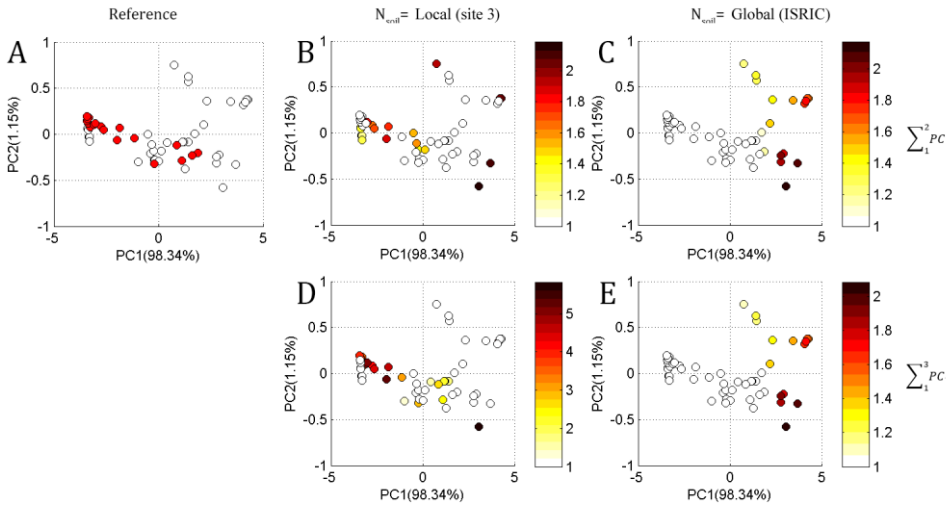
PCA score plot with marker colours indicating D_{ratio} values for the sites in Hungary.



Appendix 7.1 D_{ratio} values represented as colour on the PC1-PC2 score plot of site 1. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. Figure A) is a reference score plot where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. B) N_{soil} is used from the original site (site 1) with $\sum_1^2 PC$. C) is the result when N_{soil} is used from ISRIC with $\sum_1^2 PC$. D) N_{soil} is used from the original site (site 1) with $\sum_1^3 PC$ and E) N_{soil} is used from ISRIC with $\sum_1^3 PC$. Spectra used for this plot result are resampled by every tenth-nanometre.



Appendix 7.2 D_{ratio} values represented as colour on the PC1-PC2 score plot of site 2. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. Figure A) is a reference score plot where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. B) N_{soil} is used from the original site (site 2) with $\sum_1^2 PC$. C) is the result when N_{soil} is used from ISRIC with $\sum_1^2 PC$. D) N_{soil} is used from the original site (site 2) with $\sum_1^3 PC$ and E) N_{soil} is used from ISRIC with $\sum_1^3 PC$. Spectra used for this plot result are resampled by every tenth-nanometre.



Appendix 7.3 D_{ratio} values represented as colour on the PC1-PC2 score plot of site 3. D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. Figure A) is a reference score plot where D_{ratio} values below 1 are represented by white (empty) markers and D_{ratio} higher than 1 by yellow to dark red markers. B) N_{soil} is used from the original site (site 3) with $\sum_1^2 PC$. C) is the result when N_{soil} is used from ISRIC with $\sum_1^2 PC$. D) N_{soil} is used from the original site (site 3) with $\sum_1^3 PC$ and E) N_{soil} is used from ISRIC with $\sum_1^3 PC$. Spectra used for this plot result are resampled by every tenth-nanometre.

References

Agapiou, A., Alexakis, D. D., Sarris, A., and Hadjimitsis, D. G. (2013). Orthogonal equations of multi-spectral satellite imagery for the identification of un-excavated archaeological sites. *Remote Sensing*, *5*(12), 6560–6586.

Agapiou, A., and Hadjimitsis, D. G. (2011). Vegetation indices and field spectroradiometric measurements for validation of buried architectural remains: verification under area surveyed with geophysical campaigns. *Journal of Applied Remote Sensing*, *5*(1), 053554–053554.

Agapiou, A., Hadjimitsis, D. G., Papoutsa, C., Alexakis, D. D., and Papadavid, G. (2011). The Importance of accounting for atmospheric effects in the application of NDVI and interpretation of satellite imagery supporting archaeological research: The case studies of Palaepaphos and Nea Paphos sites in Cyprus. *Remote Sensing*, *3*(12), 2605–2629.

Agapiou, A., Hadjimitsis, D. G., Sarris, A., Georgopoulos, A., and Alexakis, D. D. (2012). Linear Spectral Unmixing for the Detection of Neolithic Settlements in the Thessalian Plain, central Greece. *Advances in Geoscience, Proc. 32nd EARSeL Symposium* (pp. 21-24).

Agapiou, A., Hadjimitsis, D., Sarris, A., Themistocleous, K., and Papadavid, G. (2010). Hyperspectral ground truth data for the detection of buried architectural remains. *Euro-Mediterranean Conference* (pp. 318-331). Springer, Berlin, Heidelberg.

Alexakis, D., Sarris, A., Astaras, T., and Albanakis, K. (2009). Detection of neolithic settle-

ments in thessaly (Greece) through multispectral and hyperspectral satellite imagery. *Sensors*, *9*(2), 1167–1187.

Aqdus, S. A., Drummond, J., & Hanson, W. S. (2008). Discovering archaeological cropmarks: a hyperspectral approach. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (pp. 361–365). Beijing, China.

Arcenegui, V., Guerrero, C., Mataix-Solera, J., Mataix-Beneyto, J., Zornoza, R., Morales, J., and Mayoral, A. M. (2008). The presence of ash as an interference factor in the estimation of the maximum temperature reached in burned soils using near-infrared spectroscopy (NIR). *Catena*, *74*(3), 177–184.

Aston, M. A., Martin, M. H., and Jackson, A. W. (1998). The use of heavy metal soil analysis for archaeological surveying. *Chemosphere*, *37*(3), 465-477.

Barba, L. A., Ortiz, A., Link, K. F., Luján, L. L., and Lazos, L. (1996). Chemical analysis of residues in floors and the reconstruction of ritual activities at the Templo Mayor, Mexico.

Bartholomeus, H. M., Schaepman, M. E., Kooistra, L., Stevens, a., Hoogmoed, W. B., and Spaargaren, O. S. P. (2008). Spectral reflectance based indices for soil organic carbon quantification. *Geoderma*, *145*(1-2), 28–36.

Bassani, C., Cavalli, R. M., Goffredo, R., Palombo, A., Pascucci, S., and Pignatti, S. (2009). Specific spectral bands for different land cover contexts to improve the efficiency of remote sensing archaeological prospection: The Arpi case study. *Journal of Cultural Heritage*, *10*, e41–e48.

Batayneh, A. T. (2011). Archaeogeophysics-archaeological prospection - A mini review. *Journal of King Saud University - Science*, *23*(1), 83–89.

Beck, A. (2010). Contrast, archaeological site detection and the non-visible component of the electromagnetic spectrum [Powerpoint slides]. Retrieved from http://slideplayer.com/slide/801079.

Beck, A., Philip, G., Abdulkarim, M., and Donoghue, D. (2007). Evaluation of Corona and Ikonos high resolution satellite imagery for archaeological prospection in western Syria. *Antiquity*, *81*(311), 161–175.

Ben-Dor, E., and Banin, A. (1990). Near-infrared reflectance analysis of carbonate concentra-

tion in soils. *Applied Spectroscopy, 44(6),* 1064–1069.

Ben-Dor, E., and Banin, A. (1994). Visible and near-infrared (0.4-1.1 μm) analysis of arid and semiarid soils. *Remote Sensing of Environment, 48*(3), 261–274.

Ben-Dor, E., and Banin, A. (1995a). Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. *Soil Science Society of America Journal, 59*(2), 364–372.

Ben-Dor, E., and Banin, A. (1995b). Near Infrared Analysis (Nira) As a Method To Simultaneously Evaluate Spectral Featureless Constituents in Soils. *Soil Science, 159(4),* 259-270.

Ben-Dor, E., Inbar, Y., and Chen, Y. (1997). The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. *Remote Sensing of Environment*, *61*(1), 1-15.

Ben-Dor, E., Patkin, K., Banin, A., and Karnieli, A. (2002). Mapping of several soil properties using DAIS-7915 hyperspectral scanner data-a case study over clayey soils in Israel. *International Journal of Remote Sensing*, *23*(6), 1043-1062.

Ben-Dor, E., Taylor, R. G., Hill, J., Demattê, J. A. M., Whiting, M. L., Chabrillat, S., and Sommer, S. (2008). Imaging Spectrometry for Soil Applications. *Advances in Agronomy*, *97*, 321–392.

Berlin, G. L., Ambler, J. R., Hevly, R. H., and Schaber, G. G. (1977). Identification of a Sinagua agricultural field by aerial thermography, soil chemistry, pollen/plant analysis, and archaeology. *American Antiquity*, *42*(4), 588-600.

Bernick, M. B., Getty, D., Prince, G., and Sprenger, M. (1995). Statistical evaluation of field-portable X-ray fluorescence soil preparation methods. *Journal of Hazardous Materials*, *43*(1-2), 111-116.

Bethell, P. H., and Smith, J. U. (1989). Trace-element analysis of an inhumation from Sutton Hoo, using inductively coupled plasma emission spectrometry: an evaluation of the technique applied to analysis of organic residues. *Journal of Archaeological Science*, *16*(1), 47-55.

Bethell, P., and Máté, I. (1989). The Use of Soil Phosphate Analysis in Archaeology: A Critique. I: Henderson, J.(ed) Scientific Analysis in Archaeology and its interpretation.

Bevan, A., and Conolly, J. (2004). GIS, archaeological survey, and landscape archaeology on the island of Kythera, Greece. *Journal of Field Archaeology*, *29*(1-2), 123-138.

Bishop, J. L., Pieters, C. M., and Edwards, J. O. (1994). Infrared spectroscopic analyses on the nature of water in montmorillonite. *Clays and Clay Minerals, 42(6)*, 702-716.

Bølviken, B., Bogen, J., Jartun, M., Langedal, M., Ottesen, R. T., and Volden, T. (2004). Overbank sediments: a natural bed blending sampling medium for large-scale geochemical mapping. *Chemometrics and Intelligent Laboratory Systems*, *74*(1), 183-199.

Bousquet, B., Sirven, J. B., and Canioni, L. (2007). Towards quantitative laser-induced breakdown spectroscopy analysis of soil samples. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *62*(12), 1582-1589.

Bowen, H. C. (1975). Air photography and the development of the landscape in central parts of southern England. *Aerial reconnaissance for Archaeology. Research Report*, *12*, 103-118.

Bowers, S. A., and Hanks, R. J. (1965). Reflection of radiant energy from soils. *Soil Science*, *100*(2), 130–138.

Brancier, J., Cammas, C., Todisco, D., and Fouache, E. (2014). A micromorphological assessment of anthropogenic features in pre-Columbian French Guiana dark soils (FGDS): first results. *Zeitschrift Für Geomorphologie, Supplementary Issues*, *58*(2), 109–139.

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

Buck, P. E., Sabol, D. E., and Gillespie, A. R. (2003). Sub-pixel artifact detection using remote sensing. *Journal of Archaeological Science*, *30*(8), 973–989.

Caneva, C., and Ferretti, M. (2000). XRF spectrometers for non-destructive investigations in art and archaeology: the cost of portability. In *The 15th International Conference on Nondestructive Testing, Roma (Italy)*.

Canti, M. G., and Linford, N. (2000). The effects of fire on archaeological soils and sediments: temperature and colour relationships. In *Proceedings of the Prehistoric Society*.

Cambridge University Press.

Cantoro, G., Pelgrom, J., and Stek, T. D. (2017). Reading a difficult landscape from the air. A methodological case-study from a WWII airfield in South Italy. *Journal of Cultural Heritage*, *23*, 12-19.

Carter, T. (2009). Elemental characterization of Neolithic artefacts using portable X-ray fluorescence [PXRF]. *Çatalhöyük 2009 Archive Report*, 126-128.

Cavalli, R. M., Pascucci, S., and Pignatti, S. (2009). Optimal spectral domain selection for maximizing archaeological signatures: Italy case studies. *Sensors*, *9*(3), 1754–1767.

Cecil, L. G., Moriarty, M. D., Speakman, R. J., and Glascock, M. D. (2007). Feasibility of Field-Portable XRF to Identify Obsidian Sources in Central Petén, Guatemala. *Archaeological Chemistry*, 506–521.

Ceraudo, G. (2013). Aerial Photography in Archaeology. In *Good Practice in Archaeological Diagnostics*. Springer, Cham.

Chabrillat, S., Goetz, A. F., Krosley, L., and Olsen, H. W. (2002). Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote sensing of Environment*, *82*(2-3), 431-445.

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

Chaya, H. J. (1996). Studies of soils from an Aleutian Island site. *Archaeological Chemistry: Organic, Inorganic and Biochemical Analysis*, Oran MV (ed.). American Chemical Society: Washington, DC.

Clark, R. N. (1999). Spectroscopy of rocks and minerals, and principles of spectroscopy. *Manual of remote sensing*, *3*(3-58), 2-2.

Clark, R. N., and Roush, T. L. (1984). Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. *Journal of Geophysical Research*, *89*(B7), 6329–6340.

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

Clark, R. N., King, T. V., Gorelick N. (1987). Automatic continuum analysis of reflectance spectra.

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

Cook, S. F., and Heizer, R. F. (1964). Studies on the chemical analysis of archaeological sites. *University of California Publications in Anthropology*, 2.

Costantini, E. A. C., and Damiani, D. (2004). Clay minerals and the development of Quaternary soils in central Italy. *Revista Mexicana de Ciencias Geológicas*, *21*(1), 144–159.

Costantini, E. A., and Damiani, D. (2004). Clay minerals and the development of Quaternary soils in central Italy. *Revista Mexicana de Ciencias Geológicas*, *21*(1).

Courchesne, F., Turmel, M. C., and Chapdelaine, C. (2015). Chemical and Mineralogical Signatures of Archaeological Features at the Mailhot-Curran Iroquoian Site, Eastern Canada. *Geoarchaeology*, *30*(5), 414-429.

Cox, C. (1992). Satellite imagery, aerial photography and wetland archaeology: an interim report on an application of remote sensing to wetland archaeology: the pilot study in Cumbria, England. *World Archaeology*, *24*(2), 249-267.

Crawford, O. G. S. (1939). Air Reconnaissance of Roman Scotland. *Antiquity*, *13*(51), 280–292.

Curcio, D., Ciraolo, G., D'Asaro, F., and Minacapilli, M. (2013). Prediction of Soil Texture Distributions Using VNIR-SWIR Reflectance Spectroscopy. *Procedia Environmental Sciences*, *19*, 494–503.

De Oliveira, J. F., Brossard, M., Corazza, E. J., Marchão, R. L., Vendrame, P. R. S., Brito, O. R., and Guimarães, M. D. F. (2015). VIS-NIR spectrometry, soil phosphate extraction methods and interactions of soil attributes. *Quimica Nova*, *38*(3), 342–350.

Doneus, M., Verhoeven, G., Atzberger, C., Wess, M., and Ruš, M. (2014). New ways to extract archaeological information from hyperspectral pixels. *Journal of Archaeological Science*, *52*, 84–96.

Donoghue, D., and Shennan, I. (1988). The application of remote sensing to environmental archaeology. *Geoarchaeology*, *3*(4), 275–285.

Drzewiecki, M. and Rączkowski, W., (2008). Following OGS Crawford: satellite images and field archaeology in Sudan, *In Proceedings of the 1st International EARSeL Workshop CNR*, 3-6.

Eckmeier, E., and Gerlach, R. (2012). Characterization of Archaeological Soils and Sediments Using VIS Spectroscopy. *Journal for Ancient Studies*, *3*, 285–290.

Eidt, R. C. (1984). *Advances in abandoned settlement analysis: application to prehistoric anthrosols in Colombia, South America*. Milwaukee, WI: Center for Latin America, University of Wisconsin-Milwaukee.

Entwistle, J. A., Abrahams, P. W., and Dodgshon, R. A. (1998). Multi-element analysis of soils from Scottish historical sites. Interpreting land-use history through the physical and geochemical analysis of soil. *Journal of archaeological science*, *25*(1), 53-68.

Entwistle, J. A., Abrahams, P. W., and Dodgshon, R. A. (2000a). The geoarchaeological significance and spatial variability of a range of physical and chemical soil properties from a former habitation site, Isle of Skye. *Journal of archaeological science*, *27*(4), 287-303.

Entwistle, J. A., Dodgshon, R. A., and Abrahams, P. W. (2000b). An investigation of former land-use activity through the physical and chemical analysis of soils from the Isle of Lewis, Outer Hebrides. *Archaeological Prospection*, *7*(3), 171-188.

Escadafal, R. (1993). Soil optical properties and environmental applications of remote sensing. *International Archives of Photogrammetry and Remote Sensing*, *29*, 709-709.

Escadafal, R., Girard, M. C., and Courault, D. (1989). Munsell soil color and soil reflectance in the visible spectral bands of Landsat MSS and TM data. *Remote Sensing of Environment*, *27*(1), 37-46.

Evans, R., and Jones, R. J. A. (1977). Crop marks and soils at two archaeological sites in Britain. *Journal of Archaeological Science*, *4*(1), 63–76.

Evelpidou, N., Barczi, A., Vona, M., Penksza, K., and Centeri, C. (2010). Preparing the soil loss prediction map of Sarviz, Hungary. *Πανελλήνια Και Διεθνή Γεωγραφικά Συνέδρια, Συλλογή Πρακτικών*, *1*, 61–71.

Fernández, F. G., Terry, R. E., Inomata, T., and Eberl, M. (2002). An Ethnoarchaeological Study of Chemical Residues in the Floors and Soils of Q'eqchi' Maya Houses at Las Pozas, Guatemala. *Geoarchaeology*, *17*(6), 487–519.

Foard, G. (1977). Systematic fieldwalking and the investigation of Saxon settlement in Northamptonshire. *World Archaeology*, *9*(3), 357-374.

Fontes, M. P. F., and Gomes, P. C. (2003). Simultaneous competitive adsorption of heavy metals by the mineral matrix of tropical soils. *Applied Geochemistry*, *18*(6), 795-804.

Fowler, M. J. (1996). High-resolution satellite imagery in archaeological application: a Russian satellite photograph of the Stonehenge region. *Antiquity*, *70*(269), 667-671.

Fowler, M. J. (2002). Satellite remote sensing and archaeology: a comparative study of satellite imagery of the environs of Figsbury Ring, Wiltshire. *Archaeological Prospection*, *9*(2), 55-69.

Frahm, E. (2013). Validity of "off-the-shelf" handheld portable XRF for sourcing Near Eastern obsidian chip debris. *Journal of Archaeological Science*, *40*(2), 1080-1092.

Gaffney, C., and Gaffney, V. (2000). Non-invasive investigations at Wroxeter at the end of the twentieth century. *Archaeological Prospection*, *7*(2), 65-67.

Gallo, D., Ciminale, M., Becker, H., and Masini, N. (2009). Remote sensing techniques for reconstructing a vast Neolithic settlement in Southern Italy. *Journal of Archaeological Science*, *36*(1), 43–50.

Gao, B. C., and Goetz, A. F. (1990). Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. *Journal of Geophysical Research: Atmospheres*, *95*(D4), 3549-3564.

Garrity, D., and Bindraban, P. (2004). A Globally Distributed Soil Spectral Library Visible Near Infrared Diffuse Reflectance Spectra. *CRAF (World Agroforestry Centre)/ISRIC (World Soil Information) Spectral Library: Nairobi, Kenya*.

Gerlach, R., Fischer, P., Eckmeier, E., and Hilgers, A. (2012). Buried dark soil horizons and archaeological features in the Neolithic settlement region of the Lower Rhine area, NW Germany: Formation, geochemistry and chronostratigraphy. *Quaternary International*, *265*, 191–204.

Gianfreda, L., Rao, M. A., Piotrowska, A., Palumbo, G., and Colombo, C. (2005). Soil enzyme activities as affected by anthropogenic alterations: Intensive agricultural practices and organic pollution. *Science of the Total Environment*, *341*(1-3), 265–279.

Golia, E. E., Dimirkou, A., and Mitsios, I. K. (2008). Levels of heavy metals pollution in different types of soil of central Greece. *Bulletin of environmental contamination and toxicology*, *80*(3), 206-210.

Greweling, T. (1962). Plant tissue analysis, an extraction procedure for the determination of total calcium, magnesium, and potassium in plant tissue. *Journal of Agricultural and Food Chemistry*, *10*(2), 138-140.

Hall, E. T. (1960). X-ray fluorescent analysis applied to archaeology. *Archaeometry*, *3*(1), 29-35.

Hao, X., and Chang, C. (2003). Does long-term heavy cattle manure application increase salinity of a clay loam soil in semi-arid southern Alberta?. *Agriculture, ecosystems & environment*, *94*(1), 89-103.

Harris Geospatial Solutions. (2004). *ENVI User's Guide.* Retrieved from http://aviris.gl.fcen.uba.ar/Curso_SR/biblio_sr/ENVI_userguid.pdf.

Haslam, R., and Tibbett, M. (2004). Sampling and analyzing metals in soils for archaeological prospection: a critique. *Geoarchaeology*, *19*(8), 731-751.

Hill, J., and Mégier, J. (Eds.). (2007). Imaging Spectrometry--a Tool for Environmental Observations (Vol. 4). Springer Science & Business Media.

Holliday, V. T., and Gartner, W. G. (2007). Methods of soil P analysis in archaeology. *Journal of archaeological science*, *34*(2), 301-333.

Hong, S., Candelone, J. P., Patterson, C. C., and Boutron, C. F. (1994). Greenland ice evidence of hemispheric lead pollution two millennia ago by Greek and Roman civilizations. *Science*, *265*(5180), 1841-1843.

Hu, N. K., and Li, X. (2017). Historical ruins of remote sensing archaeology in arid desertified environment, northwestern China. In *IOP Conference Series: Earth and Environmental Science* (Vol. 57, No. 1, p. 012028). IOP Publishing.

Huang, Z., Turner, B. J., Dury, S. J., Wallis, I. R., and Foley, W. J. (2004). Estimating foliage nitrogen concentration from HYMAP data using continuum removal analy-

sis. *Remote Sensing of Environment*, *93*(1-2), 18-29.

Hunt, A. M., and Speakman, R. J. (2015). Portable XRF analysis of archaeological sediments and ceramics. *Journal of Archaeological Science*, *53*, 626-638.

Hunt, G. R. (1977). Spectral signatures of particulate minerals in the visible and near infrared. *Geophysics*, *42*(3), 501-513.

Hunt, G.R. and Salisbury, J.W. (1970). Visible and near infrared spectra of minerals and rocks. I. Silicate minerals. *Modern Geol. 1*, 283–300.

ICRAF-ISRIC. (2010). A Globally Distributed Soil Spectral Library: Visible Near Infrared Diffuse Reflectance Spectra. World Agroforestry Centre (ICRAF) and ISRIC - World Soil Information. Retrieved from http://www.africasoils.net/afsis_files/ICRAF-ISRICSoilVNIRSpectralLibrary.pdf.

James, P. (1999). Soil variability in the area of an archaeological site near Sparta, Greece. *Journal of archaeological Science*, *26*(10), 1273-1288.

Jenkins, D. A. (1989). Trace element geochemistry in archaeological sites. *Environmental geochemistry and health*, *11*(2), 57-62.

Johnson, J. R., Lucey, P. G., Horton, K. A., and Winter, E. M. (1998). Infrared measurements of pristine and disturbed soils 1. Spectral contrast differences between field and laboratory data. *Remote Sensing of Environment*, *64*(1), 34-46.

Jones, M. U., Evison, V. I., and Myres, J. N. L. (1968). Crop-mark sites at Mucking, Essex. *The Antiquaries Journal*, *48*(2), 210-230.

Jones, R. E., and Sarris, A. (2000). Geophysical and related techniques applied to archaeological survey in the Mediterranean: a review. *Journal of Mediterranean Archaeology*, *13*(1), 3-75.

Jones, R. J. A., and Evans, R. (1975). Soil and crop marks in the recognition of archaeological sites by air photography. *Aerial reconnaissance for archaeology*, *12*, 1-11.

Jones, A., and MacGregor, G. (2002). Colouring the past: the significance of colour in archaeological research. Oxford, UK. Berg.

Kalnicky, D. J., and Singhvi, R. (2001). Field portable XRF analysis of environmental samples. *Journal of hazardous materials*, *83*(1-2), 93-122.

Kattenberg, A. E. (2008). *The application of magnetic methods for Dutch archaeological resource management.* Institute for Geo and Bioarchaeology, Vrije Universiteit.

Keeley, H. C. M. (1981). Recent work using soil phosphorus analysis in archaeological prospection. *Revue d'Archéométrie*, *5*(1), 89-95.

Knudson, K. J., Frink, L., Hoffman, B. W., and Price, T. D. (2004). Chemical characterization of Arctic soils: activity area analysis in contemporary Yup'ik fish camps using ICP-AES. *Journal of Archaeological Science*, *31*(4), 443-456.

Kruse, F. A., Boardman, J. W., and Huntington, J. F. (2003). Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(6), 1388-1400.

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., and Goetz, A. F. H. (1993). The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, *44*(2-3), 145-163.

Kvamme, K. L. (2003). Geophysical surveys as landscape archaeology. *American Antiquity*, *68*(3), 435-457.

Kvamme, K. L. (2005). Terrestrial remote sensing in archaeology. *Handbook of Archaeological Methods*, *1*.

Kvamme, K. L., Johnson, J. K., and Haley, B. S. (2006). Multiple methods surveys: case studies. *Remote sensing in archaeology: an explicitly North American perspective*, 251-268.

Lagacherie, P., Baret, F., Feret, J. B., Netto, J. M., and Robbez-Masson, J. M. (2008). Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment*, *112*(3), 825-835.

Lasaponara, R., Leucci, G., Masini, N., Persico, R., and Scardozzi, G. (2016). Towards an operative use of remote sensing for exploring the past using satellite data: The case study of Hierapolis (Turkey). *Remote sensing of Environment*, *174*, 148-164.

Lasaponara, R., and Masini, N. (2005). QuickBird-based analysis for the spatial characterization of archaeological sites: Case study of the Monte Serico medieval village. *Geophysical research letters*, *32*(12).

Lasaponara, R., and Masini, N. (2006). On the potential of QuickBird data for archaeological prospection. *International Journal of Remote Sensing*, *27*(16), 3607-3614.

Lasaponara, R., and Masini, N. (2007). Detection of archaeological crop marks by using satellite QuickBird multispectral imagery. *Journal of archaeological science*, *34*(2), 214-221.

Lasaponara, R., and Masini, N. (2008). Advances in remote sensing for archaeology and cultural heritage management. In *Proc. of I International EARSeL Workshop "Advances in Remote Sensing for Archaeology and Cultural Heritage Management", Rome*.

Lasaponara, R., and Masini, N. (2011). Satellite remote sensing in archaeology: past, present and future perspectives, *Journal of Archaeological Science, 38*, 1995–2002.

Lasaponara, R., and Masini, N. (2014). Beyond modern landscape features: New insights in the archaeological area of Tiwanaku in Bolivia from satellite data. *International Journal of Applied Earth Observation and Geoinformation*, *26*, 464-471.

Lauer, F., Prost, K., Gerlach, R., Pätzold, S., Wolf, M., Urmersbach, S., … Amelung, W. (2014). Organic fertilization and sufficient nutrient status in prehistoric agriculture? - Indications from multi-proxy analyses of archaeological topsoil relicts. *PLoS ONE*, *9*(9), e106244.

Lee, H.J. (2011). Aerial archaeology and palaeotopography analysis. Retrieved from http://www.nrich.go.kr/kr/Journal/2011_kor/sup_index.jsp?page=178.

Leone, A. P., and Escadafal, R. (2001). Statistical analysis of soil colour and spectroradiometric data for hyperspectral remote sensing of soil properties (example in a southern Italy Mediterranean ecosystem). *International Journal of Remote Sensing*, *22*(12), 2311-2328.

Linker, R., Shmulevich, I., Kenny, A., and Shaviv, A. (2005). Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy. *Chemosphere*, *61*(5), 652-658.

Liu, W., Baret, F., Gu, X., Zhang, B., Tong, Q., and Zheng, L. (2003). Evaluation of methods for soil surface moisture estimation from reflectance data. *International Journal of Remote Sensing*, *24*(10), 2069-2083.

Lopo, M., dos Santos, C. T., Páscoa, R. N. M. J., Graça, A. R., and Lopes, J. A. (2017). Near

infrared spectroscopy as a tool for intensive mapping of vineyards soil. *Precision Agriculture*, 1–18.

Loum, M., Diack, M., Ndour, N. Y. B., and Masse, D. (2016). Effect of the Continuum Removal in Predicting Soil Organic Carbon with Near Infrared Spectroscopy (NIRS) in the Senegal Sahelian Soils. *Open Journal of Soil Science*, *06*(09), 135.

Mac Arthur, A., MacLellan, C. J., and Malthus, T. (2012). The fields of view and directional response functions of two field spectroradiometers. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(10), 3892-3907.

Macqueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*(233), 281–297.

Malvern Panalytical (2002). *ASD FieldSpec Pro User's Guide*.

Maly, P., Zamansky, V., Ho, L., and Payne, R. (1999). Alternative fuel reburning. *Fuel*, *78*(3), 327-334.

Mantler, M., and Schreiner, M. (2000). X-ray Fluorescence Spectrometry in Art and Archaeology. *X-Ray Spectrometry*, *29*(1), 3–17.

Manzanilla, L. (1996). Soil analyses to identify ancient human activities. *Canadian journal of soil science*, *76*(2), 107-108.

Marcos, A., Fisher, A., Rea, G., and Hill, S. J. (1998). Preliminary study using trace element concentrations and a chemometrics approach to determine the geographical origin of tea. *Journal of Analytical Atomic Spectrometry*, *13*(6), 521-525.

Martens, H., & Naes, T. (1984). Multivariate calibration. In *Chemometrics*. Springer, Dordrecht.

Martin, M. Z., Labbé, N., André, N., Harris, R., Ebinger, M., Wullschleger, S. D., and Vass, A. A. (2007). High resolution applications of laser-induced breakdown spectroscopy for environmental and forensic applications. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *62*(12), 1426-1432.

Maskall, J. E., and Thornton, I. (1998). Chemical partitioning of heavy metals in soils, clays and rocks at historical lead smelting sites. *Water, Air, and Soil Pollution*, *108*(3-4), 391-409.

Mathieu, R., Pouget, M., Cervelle, B., and Escadafal, R. (1998). Relationships between satellite-based radiometric indices simulated using laboratory reflectance data and typic soil color of an arid environment. *Remote sensing of environment*, *66*(1), 17-28.

Matneya, T., Barrett, L. R., Dawadi, M. B., Maki, D., Maxton, C., Perry, D. S., … Whitman, L. G. (2014). In situ shallow subsurface reflectance spectroscopy of archaeological soils and features: A case-study of two Native American settlement sites in Kansas. *Journal of Archaeological Science*, *43*, 315–324.

Matschullat, J., Ottenstein, R., and Reimann, C. (2000). Geochemical background–can we calculate it?. *Environmental geology*, *39*(9), 990-1000.

McCawley, J. C., and Mckenell, H. (1971). Soil phosphorus levels at archaeological sites. *Proceedings of the Society of Antiquaries of Scotland*, *104*, 301–306.

Menze, B. H., and Ur, J. A. (2007). Classification of multispectral ASTER imagery in archaeological settlement survey in the Near East. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 244-249.

Metternicht, G. I., and Zinck, J. A. (2003). Remote sensing of soil salinity: potentials and constraints. *Remote sensing of Environment*, *85*(1), 1-20.

Middleton, W. D., and Price, D. T. (1996). Identification of activity areas by multi-element characterization of sediments from modern and archaeological house floors using inductively coupled plasma-atomic emission spectroscopy. *Journal of archaeological science*, *23*(5), 673-687.

Mielke, C., Boesche, N. K., Rogass, C., Kaufmann, H., and Gauert, C. (2015). New geometric hull continuum removal algorithm for automatic absorption band detection from spectroscopic data. *Remote Sensing Letters*, *6*(2), 97-105.

Miller, L. (1977). Soil spectra contributions to grass canopy spectral reflectance. *Photogrammetric Engineering and Remote Sensing*, *43*(6), 721-726.

Milton, E. J., Schaepman, M. E., Anderson, K., Kneubühler, M., and Fox, N. (2009). Progress in field spectroscopy. *Remote Sensing of Environment*, *113*, S92–S109.

Moholy-Nagy, H., Meierhoff, J., Golitko, M., and Kestle, C. (2013). An Analysis of pXRF Obsidian Sourch Attributions from Tikal, Guatemala. *Latin American Antiquity*, *24*(1), 72–97.

Monna, F., Galop, D., Carozza, L., Tual, M., Beyrie, A., Marembert, F., ... and Grousset, F. E. (2004). Environmental impact of early Basque mining and smelting recorded in a high ash minerogenic peat deposit. *Science of the total environment*, *327*(1-3), 197-214.

Morris, R. V., Lauer, H. V., Lawson, C. A., Gibson, E. K., Nace, G. A., and Stewart, C. (1985). Spectral and other physicochemical properties of submicron powders of hematite (α-Fe2O3), maghemite (γ-Fe2O3), magnetite (Fe3O4), goethite (α-FeOOH), and lepidocrocite (γ-FeOOH). *Journal of Geophysical Research: Solid Earth*, *90*(B4), 3126–3144.

Murphy, R. J., and Wadge, G. (1994). The effects of vegetation on the ability to map soils using imaging spectrometer data. *Remote Sensing*, *15*(1), 63-86.

Neff, J. C., Harden, J. W., and Gleixner, G. (2005). Fire effects on soil organic matter content, composition, and nutrients in boreal interior Alaska. *Canadian Journal of Forest Research*, *35*(9), 2178-2187.

Noomen, M. F., Skidmore, A. K., van der Meer, F. D., and Prins, H. H. T. (2006). Continuum removed band depth analysis for detecting the effects of natural gas, methane and ethane on maize reflectance. *Remote Sensing of Environment*, *105*(3), 262–270.

Oonk, S., Slomp, C. P., and Huisman, D. J. (2009a). Geochemistry as an aid in archaeological prospection and site interpretation: Current issues and research directions. *Archaeological Prospection*, *16*(1), 35–51.

Oonk, S., Slomp, C. P., Huisman, D. J., and Vriend, S. P. (2009b). Geochemical and mineralogical investigation of domestic archaeological soil features at the Tiel-Passewaaij site, The Netherlands. *Journal of geochemical exploration*, *101*(2), 155-165.

Ottaway, J. H., and Matthews, M. R. (1988). Trace element analysis of soil samples from a stratified archaeological site. *Environmental geochemistry and health*, *10*(3-4), 105-112.

Panishkan, K., Swangjang, K., Sanmanee, N., and Sungthong, D. (2012). Principal Component Analysis for the Characterization in the Application of Some Soil Properties. *World Academy of Science, Engineering and Technology*, *6*, 727–729.

Parnell, J. J., Terry, R. E., and Golden, C. (2001). Using in-field phosphate testing to rapidly

identify middens at Piedras Negras, Guatemala. *Geoarchaeology*, *16*(8), 855-873.

Parnell, J. J., Terry, R. E., and Nelson, Z. (2002). Soil Chemical Analysis Applied as an Interpretive Tool for Ancient Human Activities in Piedras Negras, Guatemala. *Journal of Archaeological Science*, *29*(4), 379–404.

Parson, R. B. (1962). Indian mounds of northeas Iowa as soil genesis benchmarks. *Journal of The Iowa Archaeological Society*, *12*, 1–70.

Pastor, A., Gallello, G., Cervera, M. L., and de la Guardia, M. (2016). Mineral soil composition interfacing archaeology and chemistry. *TrAC - Trends in Analytical Chemistry*, *78*, 48–59.

Pendleton, R. L., and Nickerson, D. (1951). Soil Colours and Special Munsell Soil Color Charts. *Soil Science*, *71*(1), 35–44.

Persson, K. (2005). *Integrated geophysical-geochemical methods for archaeological prospecting* (Doctoral dissertation, KTH Royal Institute of Technology).

Proudfoot, B. (1976). The analysis and interpretation of soil phosphorus in archaeological contexts. *Geoarchaeology*, 93-113.

Pyatt, F. B., Amos, D., Grattan, J. P., Pyatt, A. J., and Terrell-Nield, C. E. (2002). Invertebrates of ancient heavy metal spoil and smelting tip sites in southern Jordan: their distribution and use as bioindicators of metalliferous pollution derived from ancient sources. *Journal of Arid Environments*, *52*(1), 53-62.

Qafoku, N. P., Kukier, U., Sumner, M. E., Miller, W. P., and Radcliffe, D. E. (1999). Arsenate displacement from fly ash in amended soils. *Water, Air, and Soil Pollution*, *114*(1-2), 185-198.

Ray, S. S., Singh, J. P., Das, G., and Panigrahy, S. (2004). Use Of High Resolution Remote Sensing Data For Generating Site-Specific Soil Mangement Plan. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *35*, 127–132.

Reid, M. K., and Spencer, K. L. (2009). Use of principal components analysis (PCA) on estuarine sediment datasets: The effect of data pre-treatment. *Environmental Pollution*, *157*(8-9), 2275-2281.

Richter, N., Jarmer, T., Chabrillat, S., Oyonarte, C., Hostert, P., and Kaufmann, H. (2009).

Free Iron Oxide Determination in Mediterranean Soils using Diffuse Reflectance Spectroscopy. *Soil Science Society of America Journal*, *73*(1), 72–81.

Rodger, A., Laukamp, C., Haest, M., and Cudahy, T. (2012). A simple quadratic method of absorption feature wavelength estimation in continuum removed spectra. *Remote Sensing of Environment*, *118*, 273–283.

Rowlands, A., and Sarris, A. (2007). Detection of exposed and subsurface archaeological remains using multi-sensor remote sensing. *Journal of Archaeological Science*, *34*(5), 795–803.

Salehi, A., and Zahedi Amiri, G. (2005). Study of physical and chemical soil properties variations using principal component analysis method in the forest, North of Iran. *Caspian Journal of Environmental Sciences*, *3*(2), 131-137.

Salisbury, J. W., and Wald, A. (1992). The role of volume scattering in reducing spectral contrast of reststrahlen bands in spectra of powdered minerals. *Icarus*, *96*(1), 121-128.

Sánchez, J., Boluda, R., Morell, C., Colomer, J. C., Artigao, A., and Tébar, J. I. (1996). Assessment of soil degradation in desertification threatened areas: a case study in Castilla-La Mancha (Spain). *EFEDA-II Final Report: Desertification Processes in the Mediterranean Area and their Interlinks with the Global Climate: EFEDA-II, Subgroup II: Vegetation, Soil Physics, Inventory and Impacts'.(Ed. FM Santa-Olalla)*, 19-53.

Sánchez-Martín, M. J., García-Delgado, M., Lorenzo, L. F., Rodríguez-Cruz, M. S., and Arienzo, M. (2007). Heavy metals in sewage sludge amended soils determined by sequential extractions as a function of incubation time of soils. *Geoderma*, *142*(3-4), 262-273.

Schiffer, M. B., Sullivan, A. P., and Klinger, T. C. (1978). The design of archaeological surveys. *World Archaeology*, *10*(1), 1–28.

Schlezinger, D. R., and Howes, B. L. (2000). Organic Phosphorus and Elemental Ratios as Indicators of Prehistoric Human Occupation. *Journal of Archaeological Science*, *27*(6), 479–492.

Sellitto, V. M., Fernandes, R. B. a, Barrón, V., and Colombo, C. (2009). Comparing two different spectroscopic techniques for the characterization of soil iron oxides: Diffuse

versus bi-directional reflectance. *Geoderma*, *149*(1), 2–9.

Shackley, M. S. (2011). *X-Ray fluorescence spectrometry (XRF) in geoarchaeology*. University of California.

Shepherd, K. D., and Walsh, M. G. (2002). Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Science Society of America Journal*, *66*(3), 988–998.

Sherman, D. M., and Waite, T. D. (1985). Electronic spectra of Fe3+ oxides and oxide hydroxides in the near IR to near UV. *American Mineralogist*, *70*(11-12), 1262–1269.

Shugar, A. N., and Mass, J. L. (Eds.). (2012). *Handheld XRF for art and archaeology* (Vol. 3). Leuven University Press.

Simpson, I. A., Dockrill, S. J., Bull, I. D., and Evershed, R. P. (1998). Early Anthropogenic Soil Formation at Tofts Ness, Sanday, Orkney. *Journal of Archaeological Science*, *25*(8), 729–746.

Singh, V., Agrawal, H. M., Joshi, G. C., Sudershan, M., and Sinha, A. K. (2011). Elemental profile of agricultural soil by the EDXRF technique and use of the Principal Component Analysis (PCA) method to interpret the complex data. *Applied Radiation and Isotopes*, *69*(7), 969-974.

Slager, S., and Van de Wetering, H. T. J. (1977). Soil formation in archaeological pits and adjacent loess soils in Southern Germany. *Journal of Archaeological Science*, *4*(3), 259-267.

Small, C. (2001). Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing*, *22*(7), 1305–1334.

Speakman, R. J., and Shackley, M. S. (2013). Silo science and portable XRF in archaeology: a response to Frahm. *Journal of Archaeological Science*, *40*(2), 1435-1443.

Stenberg, B. (2010). Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma*, *158*(1-2), 15-22.

Stenberg, B. O., Nordkvist, E., and Salomonsson, L. (1995). Use of Near Infrared Reflectance Spectra of Soils for Objective Selection of Samples. *Soil Science*, *159*(2), 109–114.

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., and Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, *107*, 163–215.

Stoner, E. R., and Baumgardner, M. F. (1981). Characteristic Variations in Reflectance of Surface Soils. *Soil Science Society of America Journal*, *45*(6), 1161–1165.

Stoner, E. R., Baumgardner, M. F., Biehl, L. L., and Robinson, B. F. (1979). Atlas of soil reflectance properties.

Stoner, E. R., Baumgardner, M. F., Weismiller, R. A., Biehl, L. L., and Robinson, B. F. (1980). Extension of laboratory-measured soil spectra to field conditions. *Soil Science Society of America Journal*, *44*(3), 572–574.

Stoner, E. R., Baumgardner, M. F., Weismiller, R. A., Biehl, L. L., and Robinson, B. F. (1982). Extension of Laboratory-Measured Soil Spectra to Field Conditions. *LARS Technical Reports*.

Summers, D., Lewis, M., Ostendorf, B., and Chittleborough, D. (2011). Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecological Indicators*, *11*(1), 123-131.

Taylor, T. P. (1979). Soil mark studies near Winchester, Hampshire. *Journal of Archaeological Science*, *6*(1), 93-100.

Terhoeven-Urselmans, T., Michel, K., Helfrich, M., Flessa, H., and Ludwig, B. (2006). Near-infrared spectroscopy can predict the composition of organic matter in soil and litter. *Journal of Plant Nutrition and Soil Science*, *169*(2), 168-174.

Terry, R. E., Fernández, F. G., Parnell, J. J., and Inomata, T. (2004). The story in the floors: chemical signatures of ancient and modern Maya activities at Aguateca, Guatemala. *Journal of Archaeological Science*, *31*(9), 1237–1250.

Terry, R. E., Nelson, S. D., Carr, J., Parnell, J., Hardin, P. J., Jackson, M. W., and Houston, S. D. (2000). Quantitative phosphorus measurement: A field test procedure for archaeological site analysis at Piedras Negras, Guatemala. *Geoarchaeology*, *15*(2), 151–166.

Themistocleous, K., Agapiou, a., Cuca, B., and Hadjimitsis, D. G. (2015). Unmanned Aerial Systems and Spectroscopy for Remote Sensing Applications in Archaeology. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Infor-*

mation Sciences, *XL-7/W3*(May), 1419–1423.

Theocaris, P. S., Liritzis, I., Lagios, E., and Sampson, A. (1996). Geophysical prospection, archaeological excavation, and dating in two hellenic pyramids. *Surveys in Geophysics*, *17*(5), 593–618.

Torrent, J., and Barrón, V. (1993). Laboratory measurement of soil color: theory and practice. *Soil color*, 21-33.

Torrent, J., and Barrón, V. (2003). Iron oxides in relation to the colour of Mediterranean soils. *Applied study of cultural heritage and clays*, 377-386.

Traviglia, A. (2006). Archaeological usability of Hyperspectral images: successes and failures of image processing techniques. *BAR INTERNATIONAL SERIES*, *1568*, 123–130.

Trier, Ø. D., Larsen, S. Ø., and Solberg, R. (2009). Automatic Detection of Circular Structures in High-resolution Satellite Images of Agricultural Land. *Archaeological Prospection*, *16*(1), 1–15.

Tuniz, C., Bernardini, F., Cicuttin, A., Crespo, M. L., Dreossi, D., Gianoncelli, A., ... and Zanini, F. (2013). The ICTP-Elettra X-ray laboratory for cultural heritage and archaeology. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *711*, 106-110.

Ulrich, B. (1986). Natural and anthropogenic components of soil acidification. *Zeitschrift Für Pflanzenernährung Und Bodenkunde*, *149*(6), 702–717.

Várallya, G. (1989). Soil degradation processes and their control in Hungary. *Land Degradation & Development*, *1*(3), 171-188.

Verhoeven, G., and Sevara, C. (2016). Trying to break new ground in aerial archaeology. *Remote Sensing*, *8*(11), 918.

Viscarra Rossel, R. A. (2008a). ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and intelligent laboratory systems*, *90*(1), 72-83.

Viscarra Rossel, R. A., and Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, *158*(1), 46–54. Retrieved from http://doi.org/10.1016/j.geoderma.2009.12.025.

Viscarra Rossel, R. A., Cattle, S. R., Ortega, A., and Fouad, Y. (2009). In situ measurements

of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma*, *150*(3), 253–266.

Viscarra Rossel, R. A., Chappell, A., De Caritat, P., and Mckenzie, N. J. (2011). On the soil information content of visible-near infrared reflectance spectra. *European Journal of Soil Science*, *62*(3), 442–453.

Viscarra Rossel, R. A., and McBratney, A. B. (1998). Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. *Geoderma*, *85*(1), 19–39.

Viscarra Rossel, R. A., and McBratney, A. B. (2008b). Diffuse reflectance spectroscopy as a tool for digital soil mapping. In *Digital Soil Mapping with Limited Data* (pp. 165-172). Springer, Dordrecht.

Viscarra Rossel, R. A., McGlynn, R. N., and McBratney, A. B. (2006a). Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma*, *137*(1), 70–82.

Viscarra Rossel, R. A., Minasny, B., Roudier, P., and McBratney, A. B. (2006c). Colour space models for soil science. *Geoderma*, *133*(3-4), 320-337.

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

Walkington, H. (2010). Soil science applications in archaeological contexts: a review of key challenges. *Earth-Science Reviews*, *103*(3-4), 122-134.

Webster, R., and Butler, B. E. (1976). Soil classification and survey studies at Ginninderra. *Soil Research*, *14*(1), 1-24.

Weidong, L., Baret, F., Xingfa, G., Qingxi, T., Lanfen, Z., and Bing, Z. (2002). Relating soil surface moisture to reflectance. *Remote sensing of environment*, *81*(2-3), 238-246.

Wells, E. C., Novotny, C., and Hawken, J. R. (2007). Quantitative modeling of soil chemical data from inductively coupled plasma—optical emission spectroscopy reveals evidence for cooking and eating in ancient Mesoamerican plazas.

Wells, E. C., Terry, R. E., Parnell, J. J., Hardin, P. J., Jackson, M. W., and Houston, S. D.

(2000). Chemical analyses of ancient anthrosols in residential areas at Piedras Negras, Guatemala. *Journal of Archaeological Science*, *27*(5), 449-462.

Wetterlind, J., Stenberg, B., and Viscarra Rossel, R. A. (2013). Soil analysis using visible and near infrared spectroscopy. *Plant Mineral Nutrients: Methods and Protocols*, 95–107.

Wilson, C. A., Cresser, M. S., and Davidson, D. A. (2006). Sequential element extraction of soils from abandoned farms: an investigation of the partitioning of anthropogenic element inputs from historic land use. *Journal of Environmental Monitoring*, *8*(4), 439-444.

Wilson, C. A., Davidson, D. A., and Cresser, M. S. (2005). An evaluation of multielement analysis of historic soil contamination to differentiate space use and former function in and around abandoned farms. *The Holocene*, *15*(7), 1094-1099.

Wilson, C. A., Davidson, D. A., and Cresser, M. S. (2008). Multi-element soil analysis: an assessment of its potential as an aid to archaeological interpretation. *Journal of Archaeological Science*, *35*(2), 412–424.

Wilson, C. A., Davidson, D. A., and Cresser, M. S. (2009). An evaluation of the site specificity of soil elemental signatures for identifying and interpreting former functional areas. *Journal of Archaeological Science*, *36*(10), 2327-2334.

Wilson, D. R. (1975). *Aerial reconnaissance for archaeology*. Council for British archaeology.

Winterbottom, S. J., and Dawson, T. (2005). Airborne multi-spectral prospection for buried archaeology in mobile sand dominated systems. *Archaeological Prospection*, *12*(4), 205–219.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), 37-52.

Wynn, J. C. (1990). Applications of high-resolution geophysical methods to archaeology. *Archaeological Geology of North America, Geological Society of America, Centennial Special*, *4*, 603-617.

Zhang, J. (2010). Potential of continuum removed reflectance spectral features estimating nitrogen nutrition in rice canopy level. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on. IEEE* (pp. 1–4).