Miguel Pedro Estrada Lopo

From soil to plant advanced monitoring with infrared spectroscopy: towards tailor-made strategies for a more accurate and sustainable vineyard management

Tese do 3º Ciclo de Estudos Conducentes ao Grau de Doutor em Ciências Farmacêuticas na Especialidade de Química Analítica

Trabalho realizado sob orientação do Professor Doutor João Pedro Martins de Almeida Lopes e co-orientação do Professor Doutor José Luís Fontes da Costa Lima

Maio de 2017

Agradecimentos

Gostaria de agradecer a todas as pessoas que contribuíram directa ou indirectamente para a realização deste trabalho, em particular os membros que fazem parte assim como os que passaram pelo grupo Chemometrics and Process Analytical Technology Unit (PAT):

João Lopes
Ricardo Páscoa
Cláudia Teixeira dos Santos
Mafalda Sarraguça
Jorge Sarraguça
Clara Sousa
Silvana Moreira
Paulo Ribeiro

Um agradecimento especial:

Ao Professor Doutor João Pedro Martins de Almeida Lopes, meu orientador, pela liderança, sugestões, ideias e conselhos.

Ao Ricardo pela inestimável ajuda durante todos estes anos, inumeráveis explicações, sugestões e momentos de bom humor.

À Cláudia por me manter sempre a par, através da sua meticulosa capacidade organizacional, de todos os fastidiosos trâmites e necessárias diligências burocráticas a cumprir durante todo o Doutoramento.

Ao Jorge e à Mafalda por toda a ajuda e esclarecimentos.

A todos os membros do Laboratório de Química Aplicada.

Gostaria também de estender um agradecimento especial:
Ao Dr. António R. Graça e à empresa Sogrape Vinhos, S.A. por toda a colaboração, ajuda, disponibilização de amostras e locais de estudo.

Ao Dr. Miguel Braga e à Quinta do Mourão por toda a colaboração, cedência de instalações e disponibilização de amostras.

To all the people I had the opportunity to meet during my stay in Australia, particularly the people of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Land in Water in Adelaide: Martín Soriano-Disla, Mike McLaughlin, Danni Oliver and Les Janick. Especially:
Martín Soriano-Disla and Mike McLaughlin without whom that important journey would not have been possible.

Agradeço à Fundação para a Ciência e Tecnologia (FCT) pelo suporte financeiro referente à bolsa de Doutoramento com referência SFRH/BD/ 91521/2012 e ao projecto PTDC/AGR-PRO/6817/2014 (Modelação das vias metabólicas da videira através da combinação de dados climatéricos, solos e espectroscopia vibracional não invasiva e aplicada *in-situ*).

UNIÃO EUROPEIA
Fundo Social Europeu

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of variance
ATR	Attenuated total reflectance
B.C.	Before Christ
DNA	Deoxyribonucleic acid
EC	Electric conductivity
eCEC	Effective cation exchange capacity
ESP	Exchangeable sodium percentage
FT	Fourier transform
FT-IR	Fourier transform infrared
HTT	High-throughput transmission
InGaAs	Indium galium arsenide
IR	Infrared
LV	Latent variables
MSE	mean square error
MIR	Mid infrared
MSE	Mean square error
MLR	Multiple linear regression
NIPALS	Non-iterative partial least squares
NIR	Near infrared
NIRS	Near infrared spectroscopy
OC	Organic carbon
OIV	Office International de la Vigne et du Vin
PC	Principal component
PCA	Principal component analysis
PLS	Partial least squares
PLS1	Partial least squares with one dependent variable

PLS2	Partial least squares with multiple dependent variables
PLS-DA	Partial least squares discriminant analysis
RER	Range error ratio
RMSE	Root mean square error
RMSEC	Root mean square error of calibration
RMSECV	Root mean square error of cross-validation
RMSEP	Root mean square error of prediction
SG	Savitzky-Golay
SNV	Standard normal variate
SOM	Soil organic matter
TN	Total nitrogen
TOC	Total organic carbon
US	United States
USA	United States of America
USDA	United States Department of Agriculture
WRB	World Reference Base for Soil Resources
meq	Milliequivalent

List of Symbols

A	Absorbance
P	Loading matrix
T	Scores matrix
X	Independent data matrix
$\widehat{W}$	PLS loading weights
$\hat{b}$	PLS regression vector
p	Loading vector
q	Loading vector from dependent data vector
t	Scores vector
x	Independent data vector
y	Dependent data vector
$\hat{y}$	Predicted dependent data vector

Abstract

Thorough, sustainable vineyard management is paramount for the success of any endeavour within the wine industry. Without a healthy vineyard there is no quality wine and without quality wine there is no wine industry. In light of this and due to the socio-economic importance of wine in many countries as well as public opinion and exigency, there has recently been an interest in state of the art, environmentally friendly, and if possible, cost-effective technology. The unwavering demand from consumers, coupled with state legislation and regulation has led wine producers to seek out means of production that could, not only maintain and improve product quality, but also comply with strict modern regulations, attract more customers and diminish production costs.

Infrared spectroscopy, mainly near infrared (NIR) and mid infrared (MIR), combined with chemometric techniques, has been used in many industries with great advantages. These spectroscopic techniques have already been used during certain crucial steps of wine production. The work presented in this thesis explores infrared spectroscopy as a methodology to support specific aspects related to vineyard management, namely direct, *in-situ*, soil identification and classification with the purpose of vineyard soil mapping; indirect soil classification using grapevine leaves spectra; ampelographic differentiation of grapevine varieties and determination of essential biological, chemical and physical soil properties, for a healthy development and growth of the grape. Furthermore, comparative studies to ascertain the capabilities of portable and handheld spectrometers for such purposes were also undertook.

Throughout several studies undergone during the duration of this thesis, infrared spectroscopy is presented has an efficient alternative to lengthy laboratory based methods and is shown to possess several advantages when compared to traditional wet-chemistry methods. The main advantages are speed of analysis, simplicity in sample preparation, multiplicity of analysis and absence of chemicals. Moreover, this technology does not originate any by-products, making it eco-friendly.

This work intended to follow a logical, should we say chronological, sequence in vineyard management and wine production. Starting at the soil and moving up to the plant. The first objective was to understand if infrared spectroscopy could be used for the purpose of soil mapping. Attempts at classifying and identifying a large number of soil types coming from, not only different areas of a single vineyard, but also from distant geographical locations. Very distinct wine regions were studied, each with its own characteristics and specificities, climate conditions and topographic features that together with the planted grapevine varieties make up for very diverse *terroirs*. Using chemometric

methods such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), soil identification and mapping in all different vineyards were successful concluded. Infrared spectroscopy was also successfully applied, during the duration of this thesis, to the determination of chemical, physical and biological properties of different soil samples originating from different vineyards. Several spectrometers (benchtop, portable and handheld) performance was analysed and compared during the course of this study with the objective of assessing the quality of results on the aforementioned topics with great emphasis, obviously on efficiency, but also on portability. The advent of small, portable, performant spectrometers enables swift *in-situ* measurements that are able to provide precious insight to the producer about the state and specific needs of the vineyard. This pivotal know-how undoubtedly leads to a more accurate and sustainable vineyard management by improving efficiency in resource usage and defining tailor-made strategies for the future. Results indicated that overall, portable spectrometers spanning into the MIR range performed better than their NIR counterparts for the purpose of soil mapping and determination of important biological, physical and chemical soil properties.

Moving on to the plant, attempts to indirectly identify and classify different soil types using only spectra of grapevine leaves was successfully accomplished. It has long been known that grapevines absorb soil nutrients that will ultimately be reflected on the characteristics of the wine. This knowledge led to the assumption that nutrients passed on from soil to plant are so distinct that it would be possible to identify and classify a specific soil type using spectra of leaves form a grapevine variety planted on that same soil. Grapevine leaves are known to be repositories of important information for the health status of the plant. In addition to their role in photosynthesis, plant water status (leaf water potential can be used as an indicator of overall plant water stress), grapevine leaves are also important in ampelographic differentiation of the cultivars. In the current worldwide industrial context with constant demand for efficiency and quality in crop maintenance and food production, plant phenotyping cannot be underestimated. Grapevine ampelographic differentiation through spectra of grapevine leaves was attempted and successfully achieved.

Overall, this thesis addresses state of the art technology with practical applications into an important socio-economic industry, traditionally characterised by ancestral practices and still somewhat adverse to change. The results presented reveal that infrared spectroscopy can be rapidly implemented into routine methodologies as a tool for more efficient vineyard management. The work in this thesis also opened possibilities for further work important in vineyard management using infrared spectroscopy, such as establishing

correlations between the multielemental compositon of the grape and the soil in which the grapevine is planted. Other important parameters for the grapevine health such as the plant's water potential and to improve the correlation between spectra and soil/plant components, can be further explored.

Keywords: Near infrared spectroscopy; mid infrared spectroscopy; soil, vineyard; chemometrics; grapevine leaves.

Resumo

A gestão adequada e eficiente de um ecossistema complexo como a vinha é fundamental para o sucesso de qualquer iniciativa dentro da indústria vitivinícola. Uma vinha em mau estado não produzirá, obviamente, vinho de qualidade, sem vinho de qualidade não há indústria vitivinícola. Tendo isto em conta e também devido à importância socioeconómica do vinho em vários países, assim como à exigência da opinião pública, o interesse por tecnologia de ponta que seja amiga do ambiente e também, se possível, economicamente mais atractiva do que a já existente tem aumentado significativamente por parte dos produtores. A pressão constante dos consumidores, assim como novas normas regulativas inerentes à legislação de cada país, tem levado os produtores de vinho a procurar meios de produção que possam não só manter, mas também melhorar a qualidade do produto. Essas alterações visam fazer face a regras e controlos mais rígidos, mas também atrair novos compradores e diminuir os custos de produção.

A espectroscopia de infravermelho, principalmente de infravermelho próximo (NIR) e médio (MIR), acoplada a métodos quimiométricos tem várias aplicações e tem sido utilizada em diversas indústrias com assinalável sucesso. Estas técnicas espectroscópicas foram já testadas com resultados positivos durante alguns passos importantes da produção de vinho. O trabalho apresentando no âmbito desta tese pretende descobrir se o uso de espectroscopia de infravermelho pode ser extrapolada para aspectos específicos relacionados com a gestão da vinha, nomeadamente identificação e classificação de solos directamente, *in-situ*, com o intuito de mapear de forma precisa os solos da vinha; identificação e classificação de solos de forma indirecta, através dos espectros da folha da videira. Procedeu-se também à diferenciação ampelográfica de castas, assim como à determinação de importantes parâmetros químicos, biológicos e físicos para o desenvolvimento e crescimento da vinha. Foi também realizado um estudo comparativo para verificar as capacidades de vários espectrofotómetros portáteis para realizar os propósitos supracitados.

Através de vários estudos, a espectroscopia de infravermelho é aqui apresentada como uma alternativa eficiente a morosos métodos laboratoriais e com significativas vantagens relativamente a métodos analíticos tradicionais. As principais vantagens são a rapidez de análise, a simplicidade na preparação de amostra, a capacidade multi-paramétrica e a ausência de compostos químicos. Este último facto leva a que esta tecnologia não origine resíduos e seja totalmente amiga do ambiente.

Tentou-se neste trabalho seguir uma sequência lógica, quase “cronológica”, no desenvolvimento do produto e gestão da vinha. Começando com o solo e acabando na

planta. O primeiro objectivo foi tentar perceber se a espectroscopia de infravermelho poderia ser utilizada para o mapeamento de solos em várias vinhas. Tentou-se identificar e classificar um importante número de solos vindos não só de diferentes zonas de uma mesma vinha, mas também regiões vitivinícolas distintas e a distâncias consideráveis uma da outra. As várias regiões demarcadas utilizadas neste estudo apresentam, cada uma, características específicas de clima e topografia que junto com as tradicionais castas plantadas formam a especificidade do que se denomina de *terroir*. A identificação de solos e mapeamento de todas as diferentes vinhas foi atingido com sucesso usando métodos quimiométricos como análise de componentes principais (PCA) e análise discriminante de mínimos quadrados parciais (PLS-DA). A espectroscopia de infravermelho foi também aplicada com sucesso na determinação de propriedades químicas, físicas e biológicas de diferentes amostras de solo provenientes de diversas vinhas. O desempenho de vários espectrofotómetros (de bancada, portátil e *handheld*) foi analisado e comparado com o intuito de avaliar a qualidade dos resultados nos tópicos supracitados. Colocou-se um enfâse assinalável na qualidade dos resultados, mas também na portabilidade dos aparelhos. O desenvolvimento de espectrofotómetros cada vez mais pequenos, ao ponto de se tornarem facilmente portáteis, e com desempenhos satisfatórios permite a realização de medições rápidas no próprio local, que por seu turno proporcionarão uma informação quase imediata acerca do estado e necessidades específicas da vinha. A possibilidade de obter este tipo de conhecimento de uma forma quase instantânea levará, sem dúvida, a uma gestão mais sustentada e controlada da vinha. Os resultados demonstraram que, no geral, espectrofotómetros que se estendam pela região MIR apresentavam melhores resultados para determinação de propriedades físicas, biológicas e químicas do solo assim como para mapeamento, do que aparelhos que trabalham na região NIR.

Procedeu-se à identificação e classificação indirecta de diferentes tipos de solo usando apenas espectros de folhas de vinha. É sabido, através de vários estudos científicos, que a vinha absorbe nutrientes do solo que em última instância se reflectirão nas características do vinho. Este conhecimento levantou a hipótese que os nutrientes passados do solo para a planta seriam de tal maneira específicos que seria possível identificar e classificar um tipo de solo específico usando espectros da folha da casta plantada nesse mesmo solo. As folhas da vinha funcionam também como repositórios importantes de informação acerca do vigor da planta. Para além do papel importante que desempenham na fotossíntese, na regulação hídrica da planta (o potencial hídrico da folha pode ser usado como indicador de stress da planta), as folhas da vinha são também importantíssimas na diferenciação ampelográfica de castas. Na actual conjectura e

contexto industrial do planeta, dominado por exigências de manutenção de colheitas e produção de alimento, a identificação fenotípica de plantas não pode ser menosprezada. A diferenciação ampelográfica de vinhas através da análise de espectral de vinhas foi tentada e concluída com sucesso.

Globalmente, esta tese foca-se numa tecnologia de ponta com aplicações práticas numa área industrial de importante relevância socioeconómica, caracterizada tradicionalmente por práticas ancestrais e ainda, de certa maneira, avessa a grandes mudanças. Os resultados obtidos demonstram que a espectroscopia de infravermelho pode ser rapidamente implementada dentro de metodologias de rotina como importante ferramenta para uma gestão mais eficiente da vinha e práticas vitivinícolas. O trabalho desenvolvido durante esta tese também abriu possibilidades para novos estudos como por exemplo, perceber se existem correlações entre a composição multielementar da uva e o solo onde a planta se encontra plantada. Tentar determinar usando espectroscopia de infravermelho, outros parâmetros importantes para o bem-estar da planta, tais como o stress hídrico; melhorar a correlação entre espectros e componentes do solo/planta, entre outros. Métodos quimiométricos como análise de componentes principais, método de mínimos quadrados parciais e análise discriminante de mínimos quadrados parciais foram usados em todos os trabalhos para desenvolvimento dos modelos.

Palavras-chave:
Espectroscopia de infravermelho próximo; Espectroscopia de infravermelho médio; solo; vinha; quimiometria; folhas de videira.

Chapter I Aim, Scope and Structure

1.1. Aim and Scope

The central objective of the work presented in this thesis was to demonstrate the possibilities and advantages of using infrared spectroscopy, mainly near infrared and mid infrared, combined with chemometric techniques, during certain crucial steps of the wine industry. The work is predominantly directed to field applications in the vineyard, particularly the soil and the plant and intended to explore the use of infrared spectroscopy during important routine steps related to viticulture and vineyard management. The main objectives of this thesis can be briefly summarised into several key points:

- Demonstrate that spectroscopic techniques can be an efficient alternative to lengthy laboratory based methods and possess several advantages when compared to traditional wet-chemistry methods.

- Understand if infrared spectroscopy could be used for the purpose of direct swift soil mapping by means of identification and classification of very distinct soil types.

- Determine chemical, physical and biological properties of different soil samples originating from different vineyards.

- Analyse the performance of several benchtop, portable and handheld spectrometers with emphasis both on efficiency and portability.

- Demonstrate that indirect soil mapping is possible with infrared spectroscopy just by using spectra of grapevine leaves.

- Apply infrared spectroscopy to the ampelographic differentiation of grapevine varieties.

The development and validation of infrared spectroscopy methodologies in vineyard management could provide solutions with environmental and financial advantages, when compared to reference methods.

1.2. Structure

Chapter I: **Aim, Scope and Structure**. In this chapter, a brief outline of the main objectives as well as the motivation behind the work is provided. A general description of the remaining chapters is also included.

Chapter II: General Introduction. This chapter provides a general background about the subjects discussed within this thesis as well as the different environments studies and the techniques used. A brief historical background and socio-economic importance about the vineyard as well as its intricate constituents for the matters at study is presented. An outline on infrared spectroscopy is presented with emphasis on near infrared and mid infrared. Chemometric methods and procedures used in the various works of this thesis are also described.

Chapter III: **Progress Beyond the State-of-art.**

3.1. Soil mapping of two vineyards using near infrared spectroscopy

3.2. Classification of vineyard soils using portable and benchtop near-infrared spectrometers: a comparative study.

3.3. Modelling of soil parameters with infrared spectroscopy in Australian vineyards: an instrument comparative study.

3.4. Infrared spectroscopy suitability for the prediction of important soil properties for vine's growth and soils discrimination in Australian vineyards.

3.5. Exploratory study on vineyards soil mapping by visible/near-infrared spectroscopy of grapevine leaves.

3.6. Grapevine ampelographic differentiation using near infrared spectroscopy.

Chapter IV: **Concluding Remarks and Future Perspectives.** In this last chapter, the main concluding remarks based on the findings of the studies presented in this thesis are described, as well as the future perspectives opened by the work developed.

Chapter II General Introduction

2.1. The Vineyard

A vineyard's ecosystem is undoubtedly an intricate complex environment. It is not known when the concept of vineyard first appeared, but historians date its earliest, still rather modest and artisanal, origins to the early years of antiquity, around the same time that wine is believed to have emerged, swiftly forging its way to become an omnipresent feature in human culture, lore and everyday life. The first archaeological vestiges of wine date back to the Neolithic, as early as the sixth millennium B.C., but historians have hypothesised, based on DNA sequencing results that the winemaking process could extend even further back [1]. The concept of vineyard, winemaking and viniculture developed throughout the ages as the importance of wine also grow within the different societies, moving from ancient Mesopotamia and Persia to Egypt and Greece. With the rise of the Roman Empire and its dominion all over Europe, the dissemination of winemaking spread all around the Mediterranean basin and even further. By 500 B.C., wine was being produced in Sicily, Italy, France, Spain, Portugal and northern Africa. Cultivation of the vine also spread into the Balkan States, and the Romans took it into Germany and other parts of northern Europe, eventually reaching as far as Britain [2]. The renaissance period thriving on the back and coinciding with the great epoch of discoveries also led to the introduction of winemaking in the New World and the discovery of different grapevine species that proved fundamental to the survival of Europe's entire wine industry. Spanish conquistadors firstly planted *Vitis vinifera* in Mexico and most of South America in the sixteenth century. By the second half of the seventeenth century French vine cuttings had been planted in South Africa by Dutch settlers. The emergence of winemaking in the current United States occurred soon after, mainly in California. More than a century later, viniculture arrived in Australia and New Zealand [2]. However, like in all good romantic tragedies, disaster struck the world of wine in the Old World during the late 1800's, when the introduction of the native American *Phyloxera vastatrix* (*Dactylasphaera vitifoliae*), a grapevine roots and leaves feeding insect, almost decimated Europe's vines. The solution was to graft new *V. vinifera* vines onto phyloxera-resistant rootstocks from native American vines [3]. There is still no cure today for phylloxera.

Not much is known about vineyard management and practices in antiquity, most of the archaeological vestiges and found residues are more directly related to winemaking processes such as fermentation, used ingredients and wine storage. Some depictions of the winemaking and grape collecting can be found in ceremonial representations, on

sculptures and adorned pottery [4]. What is known is that wine production, consumption and worship thrived throughout antiquity up to the Middle Ages. Upon the fall of the Roman Empire, its production severely declined and remaining mostly under the dominion of the Christian Church due to its need for ceremonial purposes during mass celebration. Production began to flourish again at the end of the Middle Ages and beginning of the renaissance, flourishing with the great new technological and scientific leaps forward that also favoured the wine industry. The settling and development of New World nations and colonies with the inherent trade that ensues also greatly helped the progress of the wine industry [2], leading to the present day golden age of wine production.

Today, vineyard management implies a thorough scientific know-how about specifically different converging fields and an inter-relationship between intrinsic controllable factors as well as external, uncontrollable ones. According to the Office International de la Vigne et du Vin (OIV) there are today about eight million hectares of vineyards across the world, mainly concentrated within the planet's temperate zones. Each of these vineyards reflects the history, culture and traditions of its region, making up that yet elusive, semi-romantic, almost mythical entity that constitutes the *terroir*.

2.1.1. The *Terroir*

Unesco defines *terroir* as “a geographical limited area where a human community generates and accumulates along its history a set of cultural distinctive features, knowledge and practices based on a system of interactions between biophysical and human factors. The combination of techniques involved in production reveals originality, confers typicality and leads to a reputation for goods originating from this geographical area, and therefore for its inhabitants. The *terroirs* are living and innovative spaces that cannot be reduced only to tradition” [5]. The word *terroir* is believed to have originated in the XIX century form the French term *territoire* (territory). Their etymologic origin is the same, but the meaning is rather different. The use of the term *terroir is* particularly associated to viticulture, although it is also sometimes used to mention different cultivars. It is a term still characterised by a certain vagueness and is also one of the most debated issues in oenology and viticulture not only regarding its definition but also due to its commercial and marketing relevance. *Terroir* frequently refers to soil, sometimes to a vineyard site as a whole, and often to a complex effect of environmental (and occasionally viticultural and oenological) factors on fruit and wine attributes. There are indeed many complex environmental factors within a vineyard that create the viticultural potential of a *terroir* [6]. It can be defined as "a complex of natural environmental factors, which cannot

easily be modified by the producer" [7]. In a non-scientific context, the concept of *terroir* implies that a wine produced in a given region is unique and cannot be reproduced elsewhere even if the grape and winemaking techniques are meticulously duplicated [8].
When studying a specific *terroir*, no single environmental component should be studied on its own, but rather the combination of all factors present. All the relevant natural factors (such as soil type, effective soil depth, geology, water supply to the vine, altitude, aspect, among others) must be identified and characterised. These factors will obviously be expressed in the final product, but the influence of specific managerial decisions will also play an important part, resulting in distinctive wines with an identifiable origin [7]. Therefore, it is obvious that the *terroir* cannot be viewed in isolation from management and cultivation practices. A good *terroir* is considered to be one that ensures a slow but complete maturation of grapes with a certain regularity in quality of the product from vintage to vintage [9].

In terms of biology, the *terroir* is reflected in the differences found in fruit composition caused by growing the vine in different locations and environment, given that the accumulation of metabolites by grape berries is influenced by both biotic and abiotic factors. Determining the impact of a specific *terroir* on a vineyard and ultimately the wine is challenging because there may be an involvement of numerous interactions between many different characteristics, including climate, soil, topography, vineyard characteristics, cultivar, vine water status, rootstock and viticultural practices [8]. The dynamic interaction among all these diverse factors including the environment, the grapevine plant and the imposed viticultural techniques means that the wine produced in a given *terroir* is unique
Of all the environmental factors affecting the dynamics and performance of the *terroir*, the heterogeneity of the soils in which vines are grown play a major role in vine behaviour, grape quality and wine sensory characteristics [10].

2.2. Soil

Soil is the natural medium for the growth of land plants which supply food, fibres, and drugs, among other essential processes for human beings. Additionally, soil filters water and recycles wastes. It covers the earth's surface as a continuum, except on bare rock, in areas of perpetual frost or deep water, or on the ice of glaciers. According to the United States Department of Agriculture (USDA), soil is a natural body comprised of solids (organic matter and minerals), liquid, and gases that occurs on the land surface, and is characterized by horizons or layers, distinguishable from the initial material as a result of additions, losses, transfers, and transformations of energy and matter [11]. The

aforementioned official institution defines the upper limit of soil as the boundary between soil and air, shallow water, live plants, or plant materials that have not begun to decompose. The lower boundary, however, for classification purposes has a fixed depth set at 200 cm. In soils where either biological activity or other processes extend to depths greater than 200 cm, the lower limit used for classification is still 200 cm. In theory, it is the limit that separates soil from the non-soil, even though that boundary is hard to define because it flows in an ever-evolving and changing pattern through interactions between climate, relief, and living organisms. Commonly, soil gradually shifts at its lower boundary to hard rock or to earthy materials virtually devoid of animals, roots, or other marks of biological activity. These changes are often gradual and therefore difficult to discern [11] [USDA, 1999]. Pedology is the sub-discipline of soil science that integrates and quantifies the distribution, morphology, genesis, and classification of soils as natural landscape bodies [12]. It emphasises the study of soil as a natural phenomenon on the surface of the earth [13].

In vineyards the importance of soil is nowadays carefully analysed and never underestimated. It is well known by wine producers that soil attenuates the harmful effects of extreme climatic conditions such as long drought or heavy rainfall. Analyses are often carried out on physical, chemical and physicochemical analysis of the soil and the subsoil. Such analyses are indispensable since it indicates the degree to which chemicals, manure and soil conditioners are necessary for the development and growth of the vine [9]. The effect of the soil on vine behaviour and grape composition is complex, because the soil influences vine mineral nutrition and water uptake conditions, but also rooting depth and temperature in the root zone. In addition, the soil texture has a major influence on vine development and consequently on the features of the wine. Water availability depends greatly on each specific soil and can thus influence vine development and wine quality [14]. In the climate-soil-vine ecosystem it is difficult, when studying soil in isolation, to determine its influence on the constitution and the quality of grapes and wines. Human factors must be added to the natural factors, since the wine grower may happen to transform the characteristics and properties of the soil with fertilizers, chemicals, manures and occasional irrigation.

For many years now, authors have tried to establish a relationship between the quality of wines and the soil content of various important assimilable elements (potassium, magnesium, phosphorous and various oligo-elements such as boron, iron, manganese, copper and so on). It is obvious that certain tendencies can be established, particularly on a local lever, but it is now known that it does not hold on a regional or global basis [9]. At the moment, no single soil constituent or element may be considered an unquestionable

decisive factor in wine quality, leading to the basic principle that, from a winemaking point of view, soils cannot be studied in isolation. It is impossible to define the best possible soil for growing high-quality wines in terms of pebble, clay or lime content, soil depth or mineral content. These factors of the natural environment have to be considered in conjunction with their interaction with the vine. Human factors, such as history, socioeconomics, as well as viticultural and oenological techniques, also have to be taken into account [15].

2.2.1. Soil Classification

Soil classifications are often compared with biological ones. However, unlike the animal and plant kingdoms, there are numerous soil classifications, hence raising the issue of duality regarding convenience and ultimate goal. This diversity in terms of soil classification merely reflects the variety of approaches of different scientific schools and also the pressure for practical applications. This circumstance limits, to some extent, the classification activity of soil scientists, confining their knowledge to scientific journals and/or monographs. In some countries, this situation resulted in a bilingual terminology. There was a time when in Australia, for instance, there was a soil classification for scientists and one for layman [16]. Theory-based (systematics) and properties-oriented (classification) approaches of soil grouping vary in different taxonomic systems from almost completely speculative to numerical ad hoc empirical. Normally, the more effective soil classification systems combine both approaches. However, the basic theoretical concepts used to classify soils vary among the different scientific schools. Many of the variations within the different soil classifications occur because of the desire to provide information for practical uses of the soil resources, such as agriculture or forestry [17]. In most existing classifications, soils are still grouped according to their conceptual genesis. For the development of these classifications, soil morphology and properties are commonly used as distinct features and evidence to define the specific soil groupings. Furthermore, measurable and observable criteria are generally taken into account when an objective identification for the placement of the objects in the classification system is needed. Various scientific schools of thought have the same foundation for their soil grouping and apply rather similar abstract core concepts. The known differences between the numerous soil classifications are mainly due to the different theoretical approaches regarding the methodology of measurement and diagnostic definition. For a long time there lingered the belief that soil properties depended on soil processes which in turn depended on soil-forming factors. For instance, the US Soil Taxonomy [11] and Chinese

Soil Taxonomic Classification [18, 19] both use temperature regimes and internal water at high levels of taxonomic hierarchies for quantitative diagnostic properties. Nowadays, the concept changed into a more practical one and can be expressed in the following order: factors of soil formation → internal soil system functioning → specific pedogenic processes → soil properties and features → external soil functions [20]. This type of approach is routinely used in soil mapping.

Soil classifications can be characterized both qualitatively and quantitatively. Four main types of classifications exist: nominal systems, tables, reference bases and hierarchical taxonomies. All classification systems in nature break the continuum into a number of classes and the problems of classifying any natural resource is conceptually more or less the same.

Nominal systems are the most basic classifications and are usually proposed at the first stages of the onset of a comprehensive classification and are typical of most indigenous soil classifications [21]. An old system of soil series once used in the USA was nominal and in some smaller countries, the soil series are still sometimes listed in alphabetical order [22]. However, nominal systems are not comprehensive and in most cases are transformed into taxonomies during the process of accumulating empirical data [23]. Every class of objects of interest is unique, and the objects are not grouped into taxa of higher levels. In the USA, for instance, the old system of soil series was incorporated into Soil Taxonomy [11] and each series was redefined to be within the limits imposed by all higher categories.

Some soil classifications also group soil classes into a table format, similar to the periodic table of elements. The South African Republic uses such a system where the lines include the sequences of mineral horizons and the columns consist of the topsoil classes [24].

In soil classifications known as reference bases, soil classes are optionally grouped and regarded as regions or points in an n-dimensional space of properties. However, the groupings are not part of the system. If some specification is needed, an explanatory modifier is added to the name of a reference group. The French soil classification system [25] is known as a reference bases and so is the World Reference Base for Soil Resources (WRB) [26], even though it has a distinct hierarchical structure.
Hierarchical taxonomy is a more complex system with classes grouping into bigger ones at the higher levels.

The complexity of the different classifications is quite evident, but so can be the interpretation of theoretical concepts that can lead to the development of a classification within a single country. Some classifications group several of these types within different

hierarchical orders of the same classification. For instance, earlier versions of the US Soil Taxonomy had the structure of a reference base at the family level whereas the series level was a nominal system. If such a complexity is found within a single system, it is not surprising that a consensus may be hard to attain in view of terminology and concepts between different countries in the matter of soil classification. The difference between these types is not very great but the formats are very specific.

2.2.2. Soil Taxonomy

The first objective of soil taxonomy is to establish hierarchies of classes that enable the understanding of the relationship among soils and between soils. A second objective is to provide a means of communication for the discipline of soil science. Originally, soil taxonomy was developed to serve the purposes of soil survey. Today soil taxonomy attempts to make the most important statements possible about the taxa and is capable of providing taxa for all soils on a landscape wherever they may be [11].

Determining similarities among soils is not, by any means, a simple matter. It may be possible to find similarity in particle-size distribution within the members of one taxon and to the members of another taxon in base status. It is not always easy, but it is imperative to decide which property is more important. This decision must, undoubtedly, rest on the nature of the statements that it is possible to make about the classes, leading to specific groupings of the different kinds of soil. However, soils form a continuum. This continuum is broken into a manageable number of segments with limited and defined ranges in properties so that quantitative interpretations of overall soil behavior can be made [27]. When forming and defining the taxa all the known properties must be considered, even though only a few are differentiating. These differentiating properties should be the most important ones for the defined purpose or those that have the most important accessory characteristics. Normally, properties that are important to plant growth and that result from or influence soil genesis are placed into higher categories. Properties important to plant growth but unrelated to genesis are only considered for the lowest categories.

A taxonomic classification requires flexibility in the classes if it is to be effective in its prime objective which, most commonly, is soil survey. It is necessary to subdivide taxa and regroup those subdivisions into new classes of another classification for the greatest number and the most precise interpretations possible.

There is no unique “true” soil taxonomy, classification or terminology. One of the strongest arguments against uniformity in terminology is that a change of classification

would make obsolete all the existing soil maps made with older national systems. However, it is quite unfortunate that it is so, despite the obvious practicality of a common terminology, many kinds of soil are poorly represented or are unknown in a specific country. A system that would include all known soils would help to see the soils of a said country in a better perspective, particularly at both ends of the spectrum, i.e. if a kind of soil is poorly represented or is very extensive. Furthermore, it would help joint ventures, collaborative work and research as well as acquired knowledge from the experience of other countries with kinds of soil that are poorly represented or are not extensive in said country.

Soil taxonomy has been and still is a subject of much controversy. This controversy reflects, to a certain extent, the differences in concepts of soil and the different opinions about the taxonomy of soils. It is impossible to say that one taxonomic classification is better that another without mentioning the objectives to which both were developed and referring the merits each system has.

2.3. Grapevine Leaves

Grapevine leaves can be very plain and simple or exhibit various complex shapes and forms, revealing a remarkable diversity. The lengths and angles between the superior (distal) and inferior (proximal) lateral veins create an array of leaf morphs, including orbicular (circular), reniform (kidney shaped), and cordate (heart shaped) [28]. Grapevine leaves consist mainly of the petiole, by which they are attached to the shoot and at its basis two stipula encircle the shoot and the leaf-blade, which in turn is intersected by a network of veins (vascular bundles). The blade supported by the petiole positions the leaf for optimal light capture while the lamina is usually interconnected by five main veins that arise together from the point of attachment of the petiole [29]. Leaves vary in colour, size, surface shape, hirsuteness and dentation. Variations occur mainly in morphology, between genotypes, but some cultivars also exhibit profound complex shapes within a single shoot [30]. Leaves normally grow as large and thin as possible as the result of an adaptive response to gas exchange maximization during periods of shade. However, this adaptation also brings vulnerability, mainly to dehydration and photodamage [31]. Leaf cells require about two weeks to reach full size. This time frame allows the leaf to adapt to the environment in terms of optimal thickness and dimensions. Cell division is obviously crucial for leaf size, environmental limitations during cell division and expansion may limit leaf size, thus hindering the leaves and consequently the vines, true potential [32].

Water represents more than 66% of grapevines leaf wet mass, the remainder is composed of cellulose, hemicellulose, lignin, starch, protein, lipids, minerals, amino acids, soluble sugars and other secondary metabolites [33]. Despite diurnal fluctuations, leaf water potential is commonly used to describe crop water status and water stress dynamics. Water potential gradients explain most of the water flux in the soil-plant-atmosphere environment. Monitoring and knowledge of plant water status is essential to understand the interactions between the plant and its surroundings [34]. Leaf water potential has been used in grapevine monitoring, not only but mainly, for irrigation purposes [35]. Water stress may also reduce turgor pressure, leading to a lesser cell expansion. This will result in having approximately the same dry mass within a smaller leaf area, and thus increasing leaf density [36].

In viticulture, analysis of the chemical composition of leaves is extremely important for vineyard management. The nutrient status of leaves will directly affect total biomass production including the allocation of mineral nutrients to the fruits [37]. Anomalies in chemical elements will manifest in color changes. Lack of calcium (Ca) causes a narrow necrotic border at the leaf margin that moves in steps toward the petiole. Potassium (K) deficiency lightens the color of young leaves and also causes necrotic spots along the leaf margin: older leaves turn violet brown to dark brown. A deficiency in Magnesium (Mg) leads to interveinal chlorosis of older leaves. Lack or shortage of phosphorous (P) causes dark green foliage in young leaves and dark brown foliage in older leaves [38].

Vineyard leaf area is also a key determinant of grape characteristics and wine quality. A number of remote sensing studies demonstrated the relationship between canopy reflectance and grapevine biomass as well as the relationship between grape quality and final yield [39]. Vineyard leaf area has been related to important features that deeply influence the overall success yield of a vineyard such as water status, fruit ripening rate fruit characteristics and wine quality, but also infestation and disease [40].

Grapevine leaves are regarded as extremely valuable because they reflect the overall health of a vineyard. Leaves exhibit the first symptoms if disease or lack of mineral elements hails the plant. Furthermore, their importance in ampelographic differentiation of cultivars should not be undermined as well as the known possibility of differentiate soil types using leaves infrared spectra.

2.4. Infrared Spectroscopy

Spectroscopy is the study of the interaction of electromagnetic radiation with a chemical substance. The nature of that interaction depends upon the specific properties of the substance. When the radiation passes through a sample (gas, liquid or solid), certain frequencies are absorbed by the molecules of the substance, leading to molecular vibrations. The frequencies of absorbed radiation are unique for each molecule, thus providing the characteristics of the substance [41]. Infrared spectroscopy allows complex chemical information to be determined regarding a very wide range and nature of different samples. Specific spectroscopic techniques operate within different frequency ranges, depending on the process being studied and the magnitude of the associated energy change. Infrared spectroscopic techniques, such as near infrared (NIR) and mid-infrared (MIR) have been successfully used as fast and reliable analytical techniques for both quality analysis and authentication of a variety of products within the agro-food industry. However, infrared spectroscopy is not, by far, limited to agro-food products; it has been used with extreme high rates of success and reliability particularly in the pharmaceutical and petrochemical industry. Recently, infrared spectroscopy has been overlapped to other scientific sectors such as life sciences, drugs, environmental analysis and chemistry with new publications, revealing the extreme versatility of this technology, emerging almost on a daily basis, [42]. These techniques can be used for both qualitative and quantitative analysis of a great number of fields and become an alternative to laboratory, wet-chemistry techniques which are time-consuming, expensive and often generate environmentally hazardous wastes. These techniques have the great advantage of being non-destructive and extremely cost-effective [43, 44]. This unrivalled combination of simplicity, accuracy, and expeditiousness as well as a very low level (if any) of sample preparation makes infrared spectroscopy one of the most popular techniques for determining essential properties in a wide range of food products.

Infrared spectroscopy also has some disadvantages. It is a rather complex methodology; its intricate multi-disciplinary facets such as spectral technique, analytical methods and data analysis, namely in the form of chemometrics, are dependent on the availability of skilled users which sometimes are not familiar with all the necessary know-how of every intervenient field. Furthermore, because it is a secondary analytical method, infrared spectroscopy requires reference methods with accurate physical and chemical analysis. The localized nature of the scanning may also be considered a hindrance, because it may affect predictions of the internal quality of the sample. When using infrared spectroscopy, spectra is collected from a relatively small part of the specimen and the

result is thus considered an average one [45]. Instrumentation is expensive, even though the cost for sample analysis is low. The initial high investment on expensive equipment might be enough to deter the purchase of such instrumentation and limit its application for routine measurements. The influence of water is also a problematic limitation. Most biological samples and inorganic systems contain water or have water associated; specific sample procedures and preparation are therefore needed to decrease the effects of the present water or moisture. Furthermore, many compounds are not infrared active and cannot be detected (e.g. mineral compounds in the NIR region) [46]. NIR and MIR methods also have constraints on sample thickness, uniformity and dilution to avoid saturation.

2.4.1. Near Infrared Spectroscopy

Near infrared spectroscopy (NIRS) is a fast and non-destructive technique that provides multi-parametric analysis of virtually any matrix. It covers the wavelength range adjacent to the mid infrared and extends up to the visible region. It was discovered in 1800 by English astronomer Sir William Herschel who, using a prism, separated the electromagnetic spectrum and found out that the temperature increased markedly towards and beyond the red in what is nowadays called the near-infrared. However, it was not until the late 1960's that NIRS was used in a thorough scientific and practical way at an industrial level [47]. The development of equipment with improved electronic and optical components followed by the advent of increasingly powerful computers, capable of processing the information contained in NIR spectra in an acceptable time-frame enabled the expansion of this technique into an increasing number of fields. This innovation was accompanied by the introduction of NIRS as an industrial monitoring tool leading to the development of novel spectrometer configurations based on fibre optic probes and the onset of efficient chemometric data processing techniques [48].

In recent years, NIR spectroscopy has gained wide acceptance within the agro-food industry for raw material testing, product quality control and monitoring process. This growing interest is most probably due to its major advantages over other analytical techniques, i.e. extremely rapid measurements (just a few seconds), easy sample preparation, non-destructive, the possibility of varying the sample measurement position by use of fibre optic probes (extremely important for in-situ measurements), and the prediction of both chemical and physical parameters from one single spectrum [49].

Near-infrared radiation ranges from 780 – 2526 nm in the electromagnetic spectrum, which corresponds to the wave number range 12820 – 3959 cm^{-1}. The result of the

interaction between this radiation and a sample is a spectrum characterized by weak absorption bands that are broad and superimposed. The typically observed bands in NIR spectra correspond to bonds containing the hydrogen atom, such as C–H, N–H, O–H, and S–H, which are frequently present in most organic and some inorganic compounds. A NIR spectrum is essentially composed of overtones and combination bands containing useful chemical and even physical information [50].

NIRS also has some disadvantages, since it is a relative methodology, the construction of models requires prior knowledge of the value for the target parameter, which must be previously determined using a reference method. Moreover, there are no accurate models to take account of the interaction between NIR light and matter meaning that in many cases, calibration has to be purely empirical. Accurate, robust calibration models are difficult to obtain because their development means using a large enough set of samples to encompass all variations present in chemical and/or physical properties [41].

There are three measuring modes in NIR spectroscopy that are dictated by the samples optical properties: reflectance, transmittance, and transflectance. Reflectance is generally used for measuring the spectra of solids, transmittance for liquids, and transflectance for thin or clear samples. The adopted mode depends on the specific properties of the sample and its characteristics [44].

The reflectance mode measures the light that is reflected or scattered from the surface of the sample. Specular reflectance contains little information about the composition of a sample and occurs upon reflection of a smooth surface. Diffuse reflectance occurs when there is reflection off a rough surface and is more useful for analysis of the chemical and physical properties of a sample. This mode has gained much attention in the agro-food industry because of its compatibility with the characteristics and physical properties of food products.

In the transmittance mode light passes through the sample carrying information about its internal qualities. This mode can be adopted to detect both the external and internal qualities of a sample; it is possibly the simplest sampling technique for analysing solid, liquid, and gaseous samples.

Transflectance is the combination of transmission and reflectance measurements and it is not as common as the two previous modes. It is specially designed for measuring the spectra of thin or clear samples. Transflectance has been successfully used for the analysis of liquid streams, frequently in conjunction with optical bundle probes.

2.4.2. Mid Infrared Spectroscopy

Mid infrared spectroscopy monitors the fundamental vibrational and rotational stretching modes of molecules, producing a chemical profile of the sample and it spans over the MIR region (4000 – 400 cm^{-1}) of the electromagnetic spectrum. The MIR region can be segmented into two distinct regions: from 4000 to 1500 cm^{-1} and from 1500 to 500 cm^{-1}. The first region is known as the functional group region and the second one as the fingerprint region. Most of the relevant information that is used to interpret MIR spectra is usually found in the functional group region which can be further divided into specific regions representative of functional groups in biological materials. These are: the X-H stretching (4000 – 2500 cm^{-1}, where X is C, N, O, or S), including the N-H and O-H stretching frequencies (3700 – 2500 cm^{-1}), C-H stretching (3300 – 2800 cm^{-1}), and C-H stretching in aldehydes (2900 – 2700 cm^{-1}). The 2700 – 1850 cm^{-1} region corresponds to triple bonds (C≡C, C≡N, or C=C=C). The region from 1950 to 1450 cm^{-1} is attributed to a wide variety of double-bonded functional groups (C=C, C=N, C=O, and so on) [51]. The fingerprint region is more complex, exhibiting many overlapped bands because each different compound produces a different and unique absorption pattern. [52]. Nevertheless, the fingerprint region has been adopted for the determination and estimation of carbohydrates (polysaccharides) in different varieties of food products, chemical analysis of soils and in the pharmaceutical industry for the identification of drugs, among others.

MIR spectroscopy, like NIRS, enables the measurement of all types of samples (gaseous, liquid and solid), it measures the fundamental vibrations instead of the overtones and combination bands measured in the NIR region and thus, can be considered a richer analytical technique because it provides a greater amount of chemical information regarding the scanned sample. There are several different measurement modes in MIR, such as attenuated total reflectance (ATR), diffuse reflectance, high-throughput transmission (HTT), and transmission cell. Of all these modes, ATR is the most commonly used; it is based on the total internal reflectance of the MIR beam by an internal reflection element or crystal with a high reflective index [53]. The main advantage of ATR is that both qualitative and quantitative analysis can be carried out with no or minimum sample preparation. MIR transmission spectroscopy, whether using HTT or transmission cell, also has numerous applications, particularly in the food industry for quality assessment of natural products [44], as well as pharmaceutical products.

MIR sampling techniques have consistently evolved over the years, making this technology a powerful analytical method with a wide range of applications. NIR and MIR

spectroscopy can usually accomplish the same task and, depending on each specific case, one has some advantages over the other. Both techniques have positive as well as negative qualities that need to be considered if any are to be routinely implemented for any given sample analysis.

2.5. Chemometric Methods

Chemometrics is the use of mathematical and statistical methods to improve the understanding of chemical information and to correlate quality parameters or physical properties to analytical instrumental data. Spectroscopic methods are generally used for the identification or classification of samples by recognizing chemical and/or physical properties of the sample. However, spectral data normally consists of several hundreds to thousands of variables, which can be difficult to interpret without the help of a multivariate analytical method. In most cases the obtained spectra is quite complex and contains many overlapped and/or weak bands hindering an efficient analysis of the data. Furthermore, due to various factors such as light scattering, instrumental drift, base line shift or slope variation, caused by differences in particle size or physical properties of the samples, spectroscopic data cannot be interpreted directly. The need to effectively analyse the enormous amount of data generated led to the development of chemometrics. Using multivariate analysis, it became possible to amplify the information of interest and minimize the undesirable information in the spectra, retrieving both qualitative and quantitative information from the acquired data [44]. The combination of chemometrics and infrared spectroscopic has been rigorously used at laboratorial and industrial level to assess various quality parameters of different products.

There are several steps that need to be followed when analysing data using chemometrics. Firstly, and this happens before the multivariate analysis is performed, the data needs to be pre-treated to reduce, remove or correct interfering spectral parameters such as light scattering, path length variations, overlapped bands, baseline drift or random noise. This can be attained using a number of pre-processing techniques like standard normal variate (SNV) or Savitzky-Golay (SG) filter [49]. Pre-processing is followed by the selection of optimal variables from a large quantity of spectral data to develop an efficient and robust model. In this selection, it is necessary to reduce the number of variables by eliminating uncorrelated variables and keeping correlated variables containing the relevant information. Principal component analysis (PCA) is the most commonly used methodology for the reduction of variables. Variable reduction by PCA is applied before calibration and classification of the model occurs [54]. The final stage of chemometric

analysis is model validation and interpretation. At this stage there is the possibility of applying either a quantitative or classification multivariate analysis regression method, depending on the nature of the data and research objective. One of the most common multivariate methods used in quantitative analysis regression is partial least squares (PLS). This method combines the spectral information with the target property information to find the directions of largest variability. Among classification methods, partial least squares discriminant analysis (PLS-DA) is very frequently used. PLS-DA is a supervised method in which the group structure of the training set is known [49]. The nature and application of these pre-processing techniques, variable selection and multivariate methods will be further discussed in subsequent sections.

2.5.1. Spectral Pre-Processing

In infrared spectroscopic data analysis, the first step when using chemometrics is data pre-processing. Undesired effects such as light scatter effects or natural differences in samples thickness cause the acquisition of unwanted information within the spectra. Thus, pre-treatment is needed to reduce unwanted interference without affecting useful information, and improve the subsequent multivariate regression, classification model, or exploratory analysis. There are several mathematical techniques that can be used to remove, or at least, attenuate these effects, enabling that only relevant information is used to generate the model.

The most commonly used pre-processing techniques can be divided into two categories: scatter-correlation methods and spectral derivatives. Standard normal variate (SNV) is one of the most common scatter-correlation methods and it is used to reduce spectral variability caused by scattering effects [55]. Spectral derivation can be obtained using Savitzky-Golay (SG) polynomial derivative filters. This methodology is used in the resolution of overlapping peaks and removal of baseline variations, thus enabling a smoothing of the obtained spectra [56]. The first derivative is the slope in each point of the original spectrum; by applying a first derivative an additive baseline is removed. The second derivative is the slope of the first derivative and when implemented removes a linear baseline. Two parameters, besides the order of the derivative, have to be decided when performing SG. One is the choice of the window size, i.e. the number of points used in the smoothing which will always depend on the data set that is being used. This choice has to be balanced between the reduction of the noise (large window) and the distortion of the curve if the window is too large [54]. In short, SG emphasizes the analysis of the

shape rather than the height of the spectra, making it one of the most commonly adopted pre-processing techniques to reveal useful band information [57, 58].

2.5.2. Principal Component Analysis (PCA)

Optimal variable selection plays an important role in modelling construction. Even though multivariate analytical methods can handle large, noisy spectral data sets, if there are many more variables than samples, there is a high probability of over-fitting the data. Using chemometrics there is a wide range of variable selection methods that can be used for spectroscopic data. The selection of the most important variables will improve the performance of analytical methods. PCA is a fundamental unsupervised method, normally used to find hidden structures in unlabelled data and discover natural groupings within the data. The purpose of data grouping (or clustering) is to make a data cluster of samples with similar properties. Moreover, PCA is one of the most widely used methods in spectral analysis for data compression and data reduction. This method reduces the data information originating new variables (principal components) that are linear combinations of the original variables. The first principal component (PC) captures as much variability as possible and each successive PC accounts for as must as possible of the remaining variability [54]. This methodology leads to a more manageable set of variables containing almost all the information and variability within the original data.

In addition, PCA loading plots have great importance when determining which components are relevant and which variables have a significant effect on the components, thus enabling a comprehensive visualisation of the whole data in an easier way [59].

For a given matrix X, PCA decomposes the data as the sum of the product of vectors t_i and p_i plus a residual matrix E:

$$X = t_1 p^T_1 + t_2 p^T_2 + \ldots + t_A p^T_A + E \quad (2.1)$$

The process continues until the desired number of components *A*, has been extracted. The matrix consisting of the *A* most dominating principal component scores is represented by t_i and the corresponding matrix of loadings is represented by p_i [54]. Thusly, the model can be written as:

$$X = TP^T + E \quad (2.2)$$

2.5.3. Partial Least Squares (PLS)

One of the most popular supervised techniques used in infrared spectroscopy is undoubtedly PLS. It predicts data using a y vector of a known group of values (e.g., binary numbers or % concentration) instead of the natural groupings used in unsupervised methods by developing regression models that establish a model for the analysis of unknown samples. This technique permits the modelling of inter- and intra-block relationships from an X-block and Y-block of variables. The output will be real numerical values that can be compared with known targets to assess the accuracy of the multivariate model and therefore establish a model for the analysis and determination of physical and/or chemical properties of unknown samples. PLS tries to find factors that capture the greatest amount of variance in the prediction variables while seeking, at the same time, to find a single factor that best correlates predictor variables with predicted variables [60]. Thus, using PLS it is possible to predict the chemical, physical, and sensory properties of various products during, manufacturing, processing and distribution. There are several ways to calculate PLS model parameters, perhaps the most commonly used and possibly the most intuitive method is non-iterative partial least squares (NIPALS), which can be described by the following equations:

$$X = TP^T + E_x \tag{2.3}$$

$$y = Tq^T + e_y \tag{2.4}$$

The loadings P and q are determined by maximizing the correlation between the scores T. E_x and e_y are the X and y residuals. The direction of the first PLS component obtained by maximising the covariance criterion is given by $\hat{w}_1$ which is a unit length vector designated weight vector. The scores along this axis are computed as follows:

$$\hat{t}_1 = X\hat{w}_1 \tag{2.5}$$

All variables in X are regressed onto $\hat{t}$ in order to obtain the loading vector $\hat{p}_1$. The regression coefficient $\hat{q}_1$ is obtained similarly by regressing y on $\hat{t}_1$. The product of $\hat{t}_1$ and $\hat{p}_1$ is then subtracted from X and $\hat{t}_1 \hat{q}_1{}^T$ is subtracted from y. The second direction is found in the same way as the first, but using the residuals after the subtraction of the first component instead of the original data (this process is called deflation). The process is

repeated until the desired number of components *A* is achieved [54]. The regression coefficient vector used in the PLS predictor can be computed as follows:

$$\hat{b} = \hat{W}(\hat{P}^T \hat{W})^{-1} \hat{q} \quad (2.6)$$

were $\hat{W}$ is the matrix of weights. The predictions can then be calculated as:

$$\hat{y} = X\hat{b} \quad (2.7)$$

2.5.4. Partial Least Squares Discriminant Analysis (PLS-DA)

A supervised discriminant analysis such as PLS-DA works very much in the same way as the one previously described for PLS. However, a discriminant analysis is normally used for classification purposes instead of prediction. In general, classification techniques are used to separate the dataset into classes belonging to the response variable. In this particular methodology the group structure of the training set is known, allowing the construction of classification rules for a number of pre-defined sub-groups. These rules are then used to locate new samples and place them in the more appropriated group [49]. The goal of classification techniques such as PLS-DA is to establish mathematical criteria for parameterising spectral similarity, which in turn will allow for the similarity between a sample or samples and a class to be expressed quantitatively. For this purpose, comprehensive libraries of spectra that represent the natural variation of each product have to be constructed in the calibration process, with similarity being expressed by a correlation coefficient, such as the spectral match value [61].

2.5.5. Model Validation

2.5.5.1. Root Mean Square Error

Each time a calibration is performed, it is essential to decide how many components *A* are to be used. To that effect, the mean square error (MSE) and the root mean square error (RMSE) of the predictions ($\hat{y}$), which measure the difference between the actual values and the predicted values, are normally used. The RMSE is defined by the following equation:

$$\text{RMSE} = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}i - yi)^2}{N}} \quad (2.7)$$

N is number of samples. The objective is therefore to minimise the error when performing a regression model. The advantage of using RMSE instead of MSE is that it has the same units of the original measurements. Good calibration equations will have small RMSE values.

Calibration

An empirical estimate of prediction error used by some researchers is the root mean square error of calibration (RMSEC) defined as:

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}i - yi)^2}{N - A - 1}} \quad (2.8)$$

RMSEC is a measuring of the fit of the model to the calibration data. $\hat{y}$ represents the values obtained by testing the calibration equation directly on the calibration data. The problem with this error estimate is that it is an estimate of the model error and not of the prediction error. The estimation error of the regression coefficient $\hat{b}$ is not taken into account, giving an over-optimistic estimate of the prediction ability. For PLS models the difference between RMSEC and the true model prediction error can be quite large [54].

Cross-validation

Cross-validation is based solely on the calibration set. It is similar to using a prediction set, in the way that it only tests predictions on samples that were not used in the calibration. This differentiation of data is performed by removing samples from the data set in the following manner: one sample is removed from the set and calibration is performed using the remaining samples. The removed sample is then reintroduced in the data set and another sample is removed from the data set followed by a new calibration. The process continues until all samples have been removed once. This procedure is illustrated by the following equation:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_{CV},i-yi)^2}{N}} \quad (2.9)$$

where $\hat{y}_{CV,i}$ is the estimate for $\hat{y}_i$ based on the calibration equation with the sample *i* deleted.

There are several ways in which cross-validation may be performed. Leave-one-out, when only one sample is deleted at a time; Venetian blinds, in which every i^{th} sample is deleted; contiguous blocks, where a set of samples can be deleted during each iteration; and random subsets, where the samples are randomly chosen to be removed.
The statistical properties of the predictor are dependent on the number of calibration samples. This may cause the RMSECV to lead to biased estimations of the predictor based on the full dataset. However, for full cross-validation where N - 1 samples are used in each calibration, this is usually not a problem since N in most cases is large enough to ensure that (N - 1)/N ≈ 1. RMSECV does not directly estimate the actual predictor, but an approximation of the average prediction error of calibration models based on N - 1 samples [54].

Prediction

This concept is based on the splitting of the data into two sets, one for calibration and one for validation which then can be used to validate the calibration model. The prediction testing estimate of RMSE is appropriately named root mean square error of prediction (RMSEP) and is calculated according to:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}i-yi)^2}{N_P}} \quad (2.10)$$

N_P is the number of samples in the validation set and $\hat{y}$ and y are the predicted and measured reference values for the test samples. RMSEP is the simplest test that can be made to validate a model since it estimates the prediction ability of the actual predictor to be used. However, this methodology requires that several samples are set aside for testing purposes only. These samples could obviously be put back into the calibration set, giving more precise regression coefficients with better prediction ability, but in that case the predictor would be based on a larger set and its properties would become altered. Nevertheless, it is important that the test samples cover the highest possible range of samples. A generally accepted rule of thumb is to set aside about one third of the data for testing purposes and use the rest for calibration.

There are instances when the test set is an important part of the calibration, particularly in very flexible methods (e.g. in neural networks or in stepwise multiple linear regression (MLR), when the number of x-variables are determined by comparing RMSEP values).

A different option is to combine cross-validation and prediction testing. Cross-validation is used to determine model architecture while prediction is used to test the overall performance of the method.

2.5.5.2 Range Error Ratio

Model performance can also be assessed using the range error ratio (RER) parameter.

$$RER = Range/RMSEP \tag{2.11}$$

This ratio is calculated by dividing the amplitude of each parameter range by the correspondent RMSEP value [62]. With a RER > 10 the model can be considered good for quality control purposes. A low RMSEP is expected and the coefficient of determination of prediction (R^2_P) needs to be close to one.

References

1. Cavalieri, D., et al., *Evidence for S. cerevisiae fermentation in ancient wine.* Journal of molecular evolution, 2003. **57**: p. S226-S232.
2. Robinson, J. and J. Harding, *The Oxford companion to wine*. 2015: American Chemical Society.
3. Pretorius, I., *S.(2000) Tailoring wine yeast for the new millennium: novel approach to ancient art of wine making.* Yeast16: p. 675-729.
4. McGovern, P., et al., *The beginnings of winemaking and viniculture in the ancient Near East and Egypt.* Expedition, 1997. **39**(1).
5. Rouvellac, É., *Le terroir, essai d'une réflexion géographique à travers la viticulture, volume 2*. 2013, Université de Limoges.
6. Barbeau, G., et al., *Comportement du cépage Cabernet Franc dans différentes terroirs du Val de Loire. Incidence de la précocité sur la composition de la vendange en année climatique normale (exemple de 1988).* Journal International des Sciences de la Vigne et du Vin, 1998b. **32**(2): p. 69-81.
7. Carey, V., *Spatial characterization of natural terroir units for viticulture in the Bottelaryberg-Simonsberg-Helderberg winegrowing area*, in *Private Bag XI*. 2001, Stellenbosch University: Matieland (Stellenbosch), South Africa.
8. Anesi, A., et al., *Towards a scientific interpretation of the terroir concept: plasticity of the grape berry metabolome.* BMC plant biology, 2015. **15**(1): p. 191.
9. Seguin, G., *'Terroirs' and pedology of wine growing.* Experientia, 1986. **42**(8): p. 861-873.
10. Morlat, R. and F. Bodin, *Characterization of viticultural terroirs using a simple field model based on soil depth - II. Validation of the grape yield and berry quality in the Anjou vineyard (France).* Plant and Soil, 2006. **281**(1-2): p. 55-69.
11. ., E.U.D.o.A.S.C.S., *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*. 1999.
12. Simonson, R.W., *SOIL SCIENCE-GOALS FOR THE NEXT 75 YEARS.* Soil Science, 1991. **151**(1): p. 7-18.
13. Bridges, E.M., *World soils.* World Soils, by EM Bridges, pp. 176. ISBN 0521497779. Cambridge, UK: Cambridge University Press, October 1997., 1997: p. 176.
14. de Andrés-de Prado, R., et al., *Effect of soil type on wines produced from Vitis vinifera L. cv. Grenache in commercial vineyards.* Journal of agricultural and food chemistry, 2007. **55**(3): p. 779-786.

15. Van Leeuwen, C. and G. Seguin, *The concept of terroir in viticulture.* Journal of Wine Research, 2006. **17**(1): p. 1-10.
16. Moore, A., R. Isbell, and K. Northcote, *Classification of Australian soils.* Soils, an Australian viewpoint/sponsored by the Division of Soils, Commonwealth Scientific and Industrial Research Organization, Australia, 1983.
17. Ibáñez, J. and J. Boixadera, *The search for a new paradigm in pedology: a driving force for new approaches to soil classification.* Soil classification, 2001: p. 93-110.
18. Gong, Z., *Chinese soil taxonomic classification (First proposal)*. 1994: Institute of Soil Science, Academic Sinica.
19. Krasilnikov, P.V., *A handbook of soil terminology, correlation and classification*. 2009: Earthscan.
20. Targulian, V. and P. Krasilnikov, *Soil system and pedogenic processes: Self-organization, time scales, and environmental significance.* Catena, 2007. **71**(3): p. 373-381.
21. Tabor, J.A., *Ethnopedological surveys—soil surveys that incorporate local systems of land classification.* Soil Horizons, 1992. **33**(1): p. 1-5.
22. Beinroth, F., *Relationships between US soil taxonomy, the Brazilian soil classification system, and FAO/UNESCO soil units.* Soil Management in Tropical America; Proceedings of asiminar, 1975.
23. Kemp, C. and J.B. Tenenbaum, *The discovery of structural form.* Proceedings of the National Academy of Sciences, 2008. **105**(31): p. 10687-10692.
24. MacVicar, C., *Soil classification. A binomial system for South Africa*. 1977: Department of Agricultural Technical Services.
25. Baize, D. and M.-C. Girard, *Référentiel pédologique 2008*. 2009: Editions Quae.
26. *International Union of Soil Science (IUSS) Working Group World Reference Base. World Reference Base for Soil Resources. International soil classification system for naming soils and creating legends for soil maps*. 1st ed. World Soil Resources Reports. Vol. 106. 2014, Rome: FAO.
27. Heuvelink, G. and R. Webster, *Modelling soil variation: past, present, and future.* Geoderma, 2001. **100**(3): p. 269-301.
28. Chitwood, D.H., et al., *A modern ampelography: a genetic basis for leaf shape and venation patterning in grape.* Plant Physiology, 2014. **164**(1): p. 259-272.
29. Strever, A.A.E., *Non-destructive assessment of leaf composition as related to growth of the grapevine (Vitis vinifera L. cv. Shiraz)*. 2012, Stellenbosch: Stellenbosch University.
30. Galet, P., *Précis d'ampélographie pratique*. 1952: Impr. P. Déhan.

31. Tsukaya, H., *Mechanism of leaf-shape determination.* Annu. Rev. Plant Biol., 2006. **57**: p. 477-496.
32. Lindo, B. and M. Keller, *The Science of Grapevines. Anatomy & Physiology.* Experimental Agriculture, 2010. **46**(4): p. 569.
33. Jacquemoud, S., et al., *Estimating leaf biochemistry using the PROSPECT leaf optical properties model.* Remote sensing of environment, 1996. **56**(3): p. 194-202.
34. Santos, A.O. and O. Kaye, *Grapevine leaf water potential based upon near infrared spectroscopy.* Scientia Agricola, 2009. **66**(3): p. 287-292.
35. Wample, R.L., L. WILLIANS, and D. ADAMS, *Wine grape irrigation management in the San Joaquin Valley of California.* CURSO INTERNATIONAL MANEJO DE RIEGO Y SUELO EN VIDES PARA VINO Y MESA, October, Santiago de Chile, 2005: p. 38-46.
36. Witkowski, E. and B.B. Lamont, *Leaf specific mass confounds leaf density and thickness.* Oecologia, 1991. **88**(4): p. 486-493.
37. Peuke, A.D., *Nutrient composition of leaves and fruit juice of grapevine as affected by soil and nitrogen fertilization.* Journal of Plant Nutrition and Soil Science, 2009. **172**(4): p. 557-564.
38. Ordonez, C., et al., *Using hyperspectral spectrometry and functional models to characterize vine-leaf composition.* IEEE Transactions on Geoscience and Remote Sensing, 2013. **51**(5): p. 2610-2618.
39. Hall, A., et al., *Optical remote sensing applications in viticulture-a review.* Australian Journal of Grape and Wine Research, 2002. **8**(1): p. 36-47.
40. Johnson, L., et al., *Mapping vineyard leaf area with multispectral satellite imagery.* Computers and electronics in agriculture, 2003. **38**(1): p. 33-44.
41. Kumaravelu, C. and A. Gopal. *A review on the applications of Near-Infrared spectrometer and Chemometrics for the agro-food processing industries*. in *Technological Innovation in ICT for Agriculture and Rural Development (TIAR), 2015 IEEE*. 2015. IEEE.
42. Baeten, V. and P. Dardenne, *Spectroscopy: developments in instrumentation and analysis.* Grasas y aceites, 2002. **53**(1): p. 45-63.
43. Lohumi, S., et al., *Nondestructive evaluation for the viability of watermelon (citrullus lanatus) seeds using fourier transform near infrared spectroscopy.* Journal of Biosystems Engineering, 2013. **38**(4): p. 312-317.
44. Lohumi, S., et al., *A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration.* Trends in Food Science & Technology, 2015. **46**(1): p. 85-98.

45. Slaughter, D.C. and J.A. Abbott, *Analysis of fruits and vegetables.* Near-infrared spectroscopy in agriculture, 2004(nearinfraredspe): p. 377-398.
46. Burns, D.A. and E.W. Ciurczak, *Handbook of near-infrared analysis*. 2007: CRC press.
47. Williams, P., K. Norris, and W. Zarowski, *Influence of temperature on estimation of protein and moisture in wheat by near-infrared reflectance.* Cereal Chemistry, 1982.
48. Blanco, M. and I. Villarroya, *NIR spectroscopy: a rapid-response analytical tool.* TrAC Trends in Analytical Chemistry, 2002. **21**(4): p. 240-250.
49. Reich, G., *Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications.* Advanced drug delivery reviews, 2005. **57**(8): p. 1109-1143.
50. Dos Santos, C.A.T., et al., *A review on the applications of portable near-infrared spectrometers in the agro-food industry.* Applied spectroscopy, 2013. **67**(11): p. 1215-33.
51. Settle, F.A., *Handbook of instrumental techniques for analytical chemistry*. 1997: Prentice Hall PTR.
52. Jawaid, S., et al., *Rapid detection of melamine adulteration in dairy milk by SB-ATR–Fourier transform infrared spectroscopy.* Food chemistry, 2013. **141**(3): p. 3066-3071.
53. Cocciardi, R.A., *Evaluation of single-bounce attenuated total reflectanceFourier transform infrared and two-dimensional correlation spectroscopy in quantitative analysis.* 2003.
54. Naes, T., et al., *A user friendly guide to multivariate calibration and classification*. 2002: NIR publications.
55. Rinnan, Å., F. van den Berg, and S.B. Engelsen, *Review of the most common pre-processing techniques for near-infrared spectra.* TrAC Trends in Analytical Chemistry, 2009. **28**(10): p. 1201-1222.
56. Brown, C.D., L. Vega-Montoto, and P.D. Wentzell, *Derivative preprocessing and optimal corrections for baseline drift in multivariate calibration.* Applied Spectroscopy, 2000. **54**(7): p. 1055-1068.
57. Toher, D., G. Downey, and T.B. Murphy, *A comparison of model-based and regression classification techniques applied to near infrared spectroscopic data in food authentication studies.* Chemometrics and Intelligent Laboratory Systems, 2007. **89**(2): p. 102-115.

58. Lu, C., et al., *Rapid detection of melamine in milk powder by near infrared spectroscopy.* Journal of near infrared spectroscopy, 2009. **17**(2): p. 59.
59. Lachenmeier, D.W., *Rapid quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra.* Food Chemistry, 2007. **101**(2): p. 825-832.
60. Geladi, P. and B.R. Kowalski, *PARTIAL LEAST-SQUARES REGRESSION - A TUTORIAL.* Analytica Chimica Acta, 1986. **185**: p. 1-17.
61. Plugge, W. and C. Van Der Vliest, *The use of near infrared spectroscopy in the quality control laboratory of the pharmaceutical industry.* Journal of pharmaceutical and biomedical analysis, 1992. **10**(10-12): p. 797-803.
62. Santos, J.R., et al., *Exploiting near infrared spectroscopy as an analytical tool for on-line monitoring of acidity during coffee roasting.* Food Control, 2016. **60**: p. 408-415.

Chapter III: **Progress Beyond the State-of-art**

The following work was adapted from the article:

Lopo, M., Dos Santos, C. A. T., Páscoa, R. N., Graça, A. R., & Lopes, J. A. Soil mapping of two vineyards using near infrared spectroscopy. Accepted for publication in Precision Agriculture.

3.1. Soil mapping of two vineyards using near infrared spectroscopy

Abstract: The wine industry has always been particularly interested in the influence of the *terroir* characteristics on the features of a wine. Over the past few years a growing interest has spurred on the mechanisms by which a particular soil influences the vine's growth, grape variety characteristics and ultimately wine quality. Near-infrared spectroscopy (NIRS) is a rapid, non-destructive, low-cost and robust analytical method for chemical and physical property determination. Its use for soil characterization, discrimination and compound determination is rapidly increasing. In this work, NIRS data were collected in two vineyards, one in the Dão Wine Region (centre of Portugal) and one in the Vinhos Verdes Wine Region (North of Portugal) previously characterized in terms of soils. Wet, dried and dried-ground soil samples collected from specific vineyard locations were scanned on a Fourier-transform near infrared spectrometer (FTLA 2000, ABB, Quebec, Canada) in diffuse reflectance mode. Spectra were analysed with chemometric tools, namely principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). Results revealed that dried-ground soil samples presented better results, but not substantially so when compared with wet or dried samples. Discriminant models showed that the NIRS method is able to discriminate the different vineyard soil types, reproducing very accurately the mapping generated by pedology methods. Variations within the same soil type (present at different locations in the vineyard) were also detected by NIRS. The NIRS technology was shown to be suitable for correlating, complementing and perhaps eventually replacing costly, time-consuming vineyard soil mapping methods.

3.1.1. Introduction

There are many complex environmental factors within a vineyard that create the viticultural potential of a *terroir* [1]. The use of the term *terroir* in viticultural literature is characterised by vagueness. It frequently refers to soil, sometimes to a vineyard site as a whole, and often to a complex effect of environmental (and occasionally viticultural and oenological) factors on fruit and wine attributes. *Terroir* may be defined as "a complex of natural environmental factors, which cannot easily be modified by the producer" [2]. These factors will be expressed in the final product, with the aid of various management decisions, resulting in distinctive wines with an identifiable origin. Therefore the *terroir* cannot be viewed in isolation from management and cultivation practices, although such practices do not form part of the intrinsic definition [2]. Nevertheless, the overall body of scientific knowledge about the effect of *terroir* on fruit and wine attributes is relatively limited [3]. Of all the environmental factors affecting the dynamics and performance of the *terroir*, the different soils in which vines are grown play a major role in the vine behaviour, grape quality and wine sensory characteristics [4]. Some authors have shown that wine multi-elemental composition was strongly influenced by the solubility of inorganic compounds of the soil. In principle, the pattern of a wine will reflect the geochemistry of the provenance soil [5, 6]. Although it is not clear that trace elements in wine can substantially affect taste, it seems likely that they could reflect soil compositions and be useful for fingerprinting [7]. Knowledge of grapevine viticultural and oenological performance as influenced by site-related factors represents a significant value to the grape and wine industry. Adequate site selection will ensure the highest value of the final product. The knowledge of natural resources of a vineyard site, beyond giving indications for zoning, permits the application of an appropriate vineyard management that can optimize fruit and wine quality from a given area [8].

The currently adopted methods for determining soil chemical and physical characteristics are generally time consuming, as they need samples to be transported to the laboratory as well as the time spent to chemically analyse the sample itself [9]. Clearly, soil analysis techniques that are cheap and simultaneously require less sample processing are desirable. Near infrared spectroscopy (NIRS) has been used since the 1970s for the routine evaluation of foods and forages in the laboratory [10]. The NIR spectral region is dominated by weak overtones and combinations of vibrational bands from molecular bonds containing hydrogen attached to atoms such as nitrogen, oxygen and carbon. It often allows the analysis of several chemical properties at the same time [11, 12]. In NIRS, calibration requires mathematical tools since it involves multivariate

regression techniques. These techniques relate NIRS optical measurements at specific wavelengths with reference values previously obtained by conventional methods. The advantages of NIRS are the speed of analysis, simplicity in sample preparation, multiplicity of analysis, and absence of using chemicals [10]. During the past 20 years NIRS has been successfully applied to the determination of chemical, physical and biological properties in soil samples. There is a growing literature on successful calibrations between NIRS signals and amounts of one or more soil attributes sufficient to be used as surrogates for conventional soil testing [13-15]. One of these attributes may be moisture for instance, which can be highly variable in the soil [16], and it is known that water is extremely active in the NIRS range affecting the spectra significantly [17]. These facts may be very relevant when moisture is being estimated from spectra, but can prove counterproductive when other soil properties are investigated.

However, most published studies involving NIRS associations with C and/or N involve samples from multiple locations, across fields and sometimes even regions, often in the absence of in-depth knowledge of the soil system, cropping, and topsoil-management practices [18-22].

Particle size and soil structure are also known to affect the NIR signal [23]. Previous works in the literature have shown that as the particle size increased, the NIR reflectance decreased [24] and that the effect of soil structure is related to Fe-bearing minerals, clay minerals, and C–H functional groups of organic matter [25]. However, these effects can be minimized through the application of specific preprocessing techniques such as the first-order derivative [24].

The purpose of this study was to investigate the potential of NIRS as a rapid and low-cost technique to map vineyard soils with the objective to increase the producer's knowledge for planning crops and new vineyard plantations. NIRS was tested as a tool to discriminate between different types of vineyard soils. For this purpose, soil samples were collected on a vineyard previously characterized with pedology methods. This approach may provide an advantageous alternative to currently existing methods for soil mapping and a better understanding, from the producer's point of view, of the selection of soils best "matching" specific vine species, ultimately leading to the production of wines with targeted characteristics.

3.1.2. Material and methods

3.1.2.1. Sample collection

Soil samples were collected from two previously characterized vineyards: *Quinta dos Carvalhais* (Mangualde, 40°33'20"N 7°46'40"W) in the Dão Wine Region, center of Portugal (Fig. 3.1.1a) and *Quinta de Azevedo* (Barcelos, 41º34' 12.45''N 8º32'25.05''O) in the *Vinho Verde* Wine Region, north of Portugal (Fig. 3.1.1b). These vineyard soils were previously characterized using pedological reference methods [26] and classified according to the International Union of Soil Science (IUSS) [27]. The soils analysed in both vineyards were previously characterized through the measurement of multiple chemical and physical properties (Table 3.1.1). The vineyards are divided into numbered "blocks" of approximately one hectare each, containing a single variety (*Vitis vinifera* cultivar). However, in many blocks, multiple soil types co-exist (especially in *Quinta dos Carvalhais*). Forty-five sampling locations encompassing seven different soil types in *Quinta dos Carvalhais* and forty-four sampling locations comprising three different soil types in *Quinta de Azevedo* were selected in this work. These locations were defined according to the mapped soil characteristics and grape varieties. The sampling strategy required digging one hole (0.5m diameter) at each location. Approximately 500g soil sample was collected from each hole at specific pre-defined depths according to each soil analyzed (see Table 3.1.2). Depths were chosen according to a combination of horizon thickness and fixed depth considerations for each soil type. Samples were collected and taken into the laboratory for spectral data acquisition. Samples were measured in three different forms. 1) Wet samples: spectra were acquired immediately after sample reception at the laboratory, roughly 6 hours after sample collection; 2) Dried samples: samples were dried in an oven (Raypa DO-20, Barcelona, Spain) at 45 °C for two weeks and measured by NIRS; 3) Dried-ground samples: dried samples were milled using a mortar and pestle before spectra were collected. Wet (intact) samples spectra simulate what should be expected if a portable instrument was used to acquire spectra directly in the vineyard.

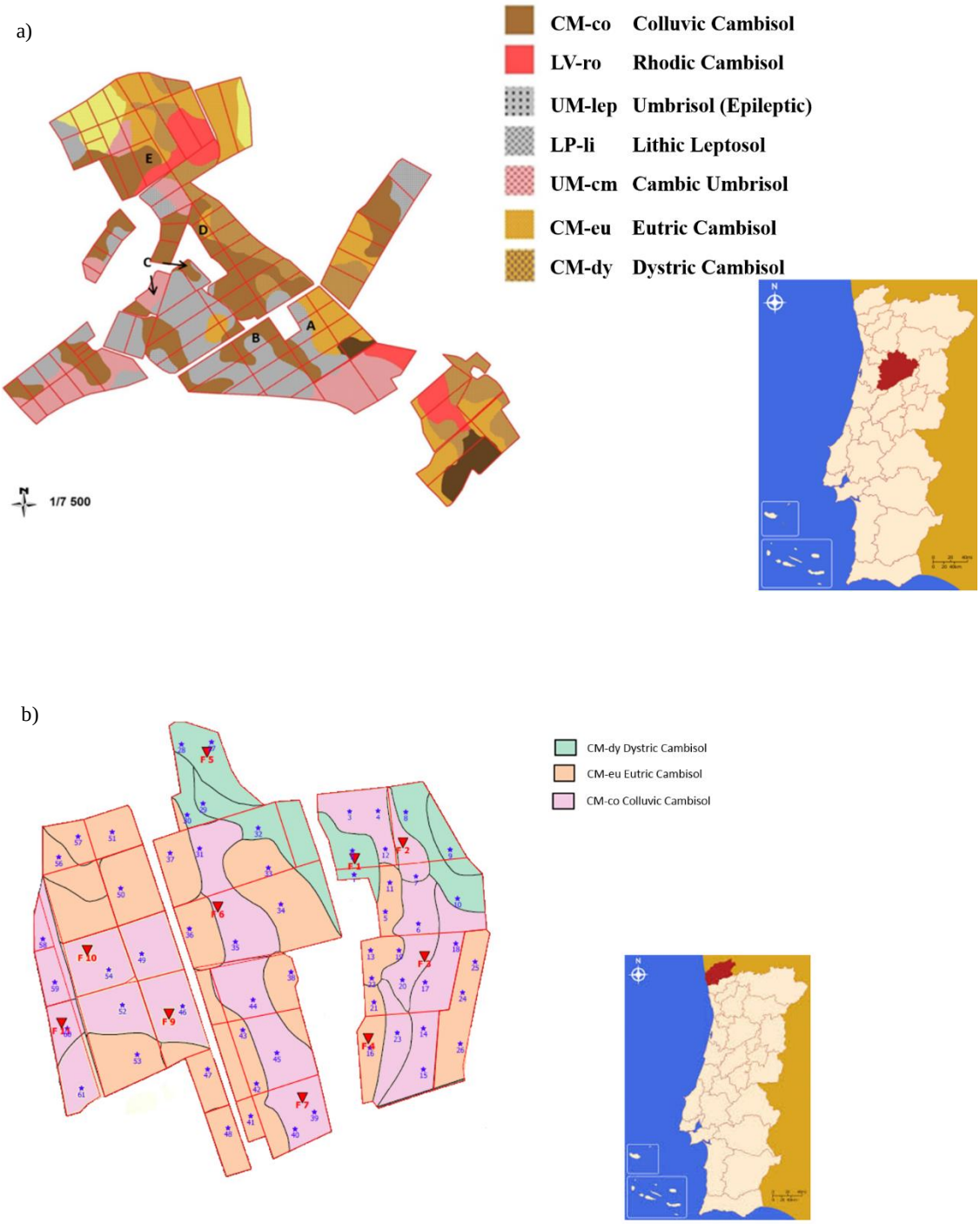


Figure 3.1.1. Vineyard maps coloured according to the soil types with indication of sampling locations: (a) *Quinta dos Carvalhais* (representative of the Dão wine region) and (b) *Quinta de Azevedo* (representative of the *Vinhos Verdes* wine region). Both regions are marked in red in Portugal's map.

Table 3.1.1. Soil types and selected properties obtained at specific depths for *Quinta dos Carvalhais* and *Quinta de Azevedo*.

Vineyard[a]	Soil type code[b]	Soil Texture	Depth (cm)	Total nitrogen (g/Kg)	pH	Organic matter (g/Kg)
QC	CM-eu	Sandy Loam	0-20	0.61	6.8	13.2
QC	UM-lep	Sandy Loam	0-20	[c]	6.8	3.4
QC	LP-li	Sandy Loam	0-10	0.7	5.2	10.0
QC	CM-co	Sandy Loam	0-20	0.86	6.5	19.1
QC	UM-cm	Sandy Loam	0-25	0.6	7.0	7.1
QC	LV-ro	Sandy Loam	0-15	0.68	7.0	16.7
QC	CM-dy	Sandy Loam	0-10	[c]	6.5	40.2
QA	CM-dy		0-20	7.0	5.8	13.0
QA	CM-co		0-20	1.1	6.1	23.4
QA	CM-co		0-20	1.0	6.0	18.2
QA	CM-eu		0-20	0.7	6.3	12.7
QA	CM-co		0-20	0.7	5.9	5.4

[a]QC (*Quinta dos Carvalhais*); QA (*Quinta de Azevedo*)

[b]CM-eu: Eutric Cambisol; UM-lep: Umbrisol (Epileptic); LP-li: Lithic Leptosol; CM-co: Colluvic Cambisol; UM-cm: Cambic Umbrisol; LV-ro: Rhodic Luvisol; CM-dy: Dystric Cambisol.

[c] Values not measured.

3.1.2.2. Near infrared spectral acquisition

Near infrared spectra of all soil samples were collected in diffuse reflectance mode on a Fourier-transform near infrared spectrometer (FTLA 2000, ABB, Québec, Canada) equipped with an indium–gallium-arsenide (InGaAs) detector. Each spectrum was recorded as the average of 64 scans with 8 cm^{-1} resolution over a wavenumber interval between 10000 cm^{-1} and 4000 cm^{-1}. The equipment was controlled via the Grams software (version 7, ABB, Québec, Canada). Approximately 30 g of each soil sample was transferred into borosilicate flasks in order to perform spectral acquisition. The background was taken at the beginning of each analysis using Spectralon (teflon) as reference. NIR spectra of wet, dried and dried-ground samples were collected. These spectra were pre-processed with a Savitzky–Golay filter (15-points filter size, second order polynomial, and first-order derivative). This pre-processing technique will minimize unwanted light scattering effects and other baseline drifts. Therefore, a total of 369 spectra were available for *Quinta dos Carvalhais,* including all depths and sample type

(wet, dried, dried-ground samples) and a total of 528 spectra were available for *Quinta de Azevedo* in the same conditions as *Quinta dos Carvalhais*.

Table 3.1.2. Sampling locations and respective soil types for *Quinta dos Carvalhais* and *Quinta de Azevedo*.

Sampling Spot	Vineyard[a]	Block code	Soil type code[b]	Sampling depths (cm)
1-3	QC	*a*	CM-eu	0-20/20-75
4-6	QC	*a*	UM-lep	0-20/20-50
7-9	QC	*a*	LP-li	0-10/10-70/80-115
10-12	QC	*a*	CM-co	0-20/20-110/>110
13-15	QC	*b*	CM-co	0-20/20-110/>110
16-18	QC	*b*	LP-li	0-10/10-70/80-115
19-21	QC	*c*	LP-li	0-10/10-70/80-115
22-24	QC	*c*	CM-co	0-20/20-110/>110
25-27	QC	*c*	UM-cm	0-25/25-100/100-150
28-30	QC	*d*	CM-eu	0-20/20-75
31-33	QC	*d*	CM-co	0-20/20-110/>110
34-36	QC	*d*	CM-dy	0-20/20-70
37-39	QC	*e*	LV-ro	0-15/15-50/50-75/75-110
40-42	QC	*e*	CM-co	0-20/20-110/>110
43-45	QC	*e*	CM-eu	0-20/20-75
1-4	QA	*a*	CM-dy	0-20/20-50/50-100/100-150
5-8	QA	*b*	CM-co	0-20/20-50/50-100/100-150
9-12	QA	*c*	CM-co	0-20/20-50/50-100/100-150
13-16	QA	*d*	CM-eu	0-20/20-50/50-100/100-150
17-20	QA	*e*	CM-co	0-20/20-50/50-100/100-150
21-24	QA	*f*	CM-dy	0-20/20-50/50-100/100-150
25-28	QA	*g*	CM-eu	0-20/20-50/50-100/100-150
29-32	QA	*h*	CM-co	0-20/20-50/50-100/100-150
33-36	QA	*i*	CM-co	0-20/20-50/50-100/100-150
37-40	QA	*j*	CM-co	0-20/20-50/50-100/100-150
41-44	QA	*k*	CM-co	0-20/20-50/50-100/100-150

[a] QC (*Quinta dos Carvalhais*); QA (*Quinta de Azevedo*)

[b] Acronyms are explained in Table 3.1.1.

3.1.2.3. Data analysis

Principal components analysis (PCA) [28] and partial least-squares discriminant analysis (PLS-DA) [29] were the selected data analysis methods. The former was used to detect common patterns within soil samples as well as putative outliers. The latter was used to develop calibration models for soil discrimination purposes. Before application of PCA and PLS-DA, the entire spectra were subjected to mean-centring as this is a requirement of both methods [30] For PLS-DA models, the data selected for the model were divided to form calibration (70% of the available samples) and test (remaining 30%) sets [30]. Samples were divided randomly while ensuring the same proportion between soil types was present in the calibration and test sets to avoid unbalanced classes (soil types) across sets [31]. The optimal number of latent variables (LVs) was estimated by leave-one-block-out cross-validation (considering blocks of 15 samples) using only the calibration set [28]. The test set results were projected onto each optimized PLS-DA model and soil predictions were expressed as confusion matrices [32]. Confusion matrices compare each known sample soil type with the corresponding NIRS prediction and entries are expressed as percentages. The objective of these matrices is not only to estimate the number of correctly predicted samples but also to define which samples are being incorrectly predicted, identifying the most similar soil types in terms of NIRS. For PCA analysis, all samples were used (wet, dried and dried-ground) including all depths. For PLS-DA analysis, only samples from the superficial layer were used (wet, dried, and dried-ground). All chemometric methods and spectra processing were performed using Matlab version 7.9 (Mathworks, Natick, MA) and the PLS Toolbox version 5.5.1 (Eigenvector Research, Inc., Wenatchee, WA).

3.1.3. Results

An initial spectral comparison was carried out between wet and dried samples. Fig. 3.1.2 compares the NIR spectrum profile of a wet and corresponding dried sample for a CM-eu soil sample from *Quinta dos Carvalhais*. The obtained soil spectra were similar to those previously reported within the literature [33] and therefore, no detailed analysis on spectral bands was made. The reduction of the bands due to OH vibration in water at 6900 cm^{-1} and 5200 cm^{-1} on the dried sample spectrum was quite clear.

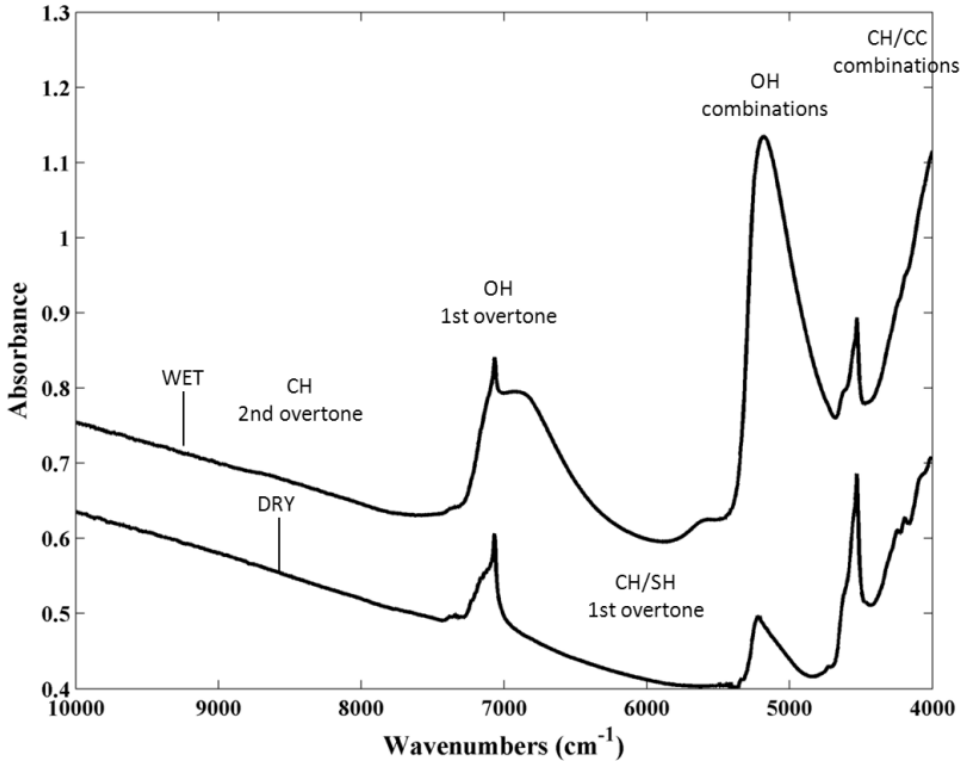


Figure 3.1.2. Unprocessed NIR spectra of a CM-eu wet (intact) and corresponding dried soil sample.

3.1.3.1. PCA

To verify the difference between modelling wet and dry samples, PCA models comprising the entire spectral region (10000-4000 cm^{-1}) were developed for samples of each soil type separately. A detailed analysis of scores and loadings revealed that similar scores distribution and an almost perfect matching between loadings was obtained for all soil types (data not shown). This shows that there are no significant differences between the NIR spectra of wet and dry samples.

A spectral analysis was performed to identify the effect of depth on the different soil types. For this task, dried samples were selected. PCA models for each soil type, using only dried samples, were developed considering the entire spectral region (10000-4000 cm^{-1}) and samples were analysed in terms of scores. The results obtained for the soil type CM-eu (appearing in three different blocks in *Quinta dos Carvalhais*) are shown in Fig. 3.1.3. Two depths for this particular type of soil were collected (0-20 cm and 20-75 cm). The PCA model reveals that no substantial difference between the different depths is evident. However, it is clear that, even though samples are of the same soil type, NIRS detects a clear separation between these samples. The separation is not due to sampling depth but to the geographical provenance of samples (different vineyard blocks), meaning

that the samples were differentiated by block and not by depth in this data set. Similar results showing separation by block and not by depth were found for the other soil types (data not shown). Since no substantial spectral differences were observed for the different depths, further models were developed using only spectra of the most superficial layer.

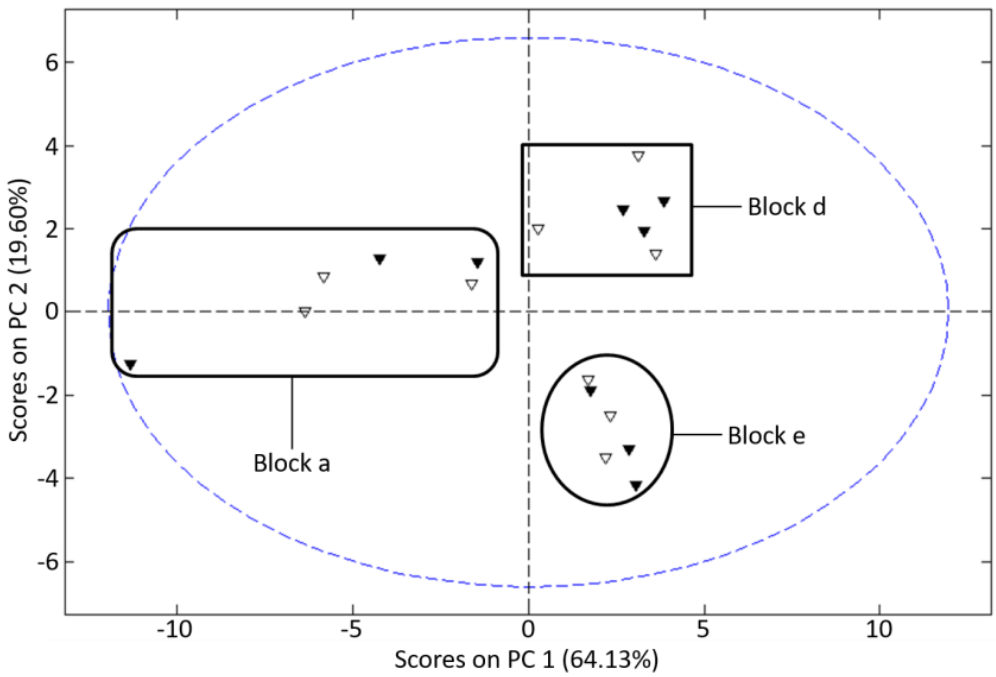


Figure 3.1.3. PCA score plot of NIR spectra obtained for CM-eu soil samples in *Quinta dos Carvalhais* considering the entire spectral range (▼= 0-20cm; ▽= 20-75cm).

To investigate the ability of NIRS to discriminate between soil types, PCA models using only dried samples were developed considering each vineyard block separately. The spectral profile of specific soil within a block is a combination between the soil type itself and the biological/chemical/physical characteristics of the analysed block. This type of analysis (separating each block) is therefore justified by the need to minimize soil geographical variations within each model. Additionally, models considering the entire spectra range (10000-4000 cm^{-1}) and models considering only sub-regions were tested. Sample scores were evaluated on each of these models. It was observed that a maximum level of discrimination was achieved for models considering blocks *a*, *b* and *e* when the entire spectral region was used and for blocks *c* and *d* when the region 7198-6623 cm^{-1} was used (Fig. 3.1.4a to 3.1.4e). These models revealed a good separation of samples according to soil type in most cases (especially in blocks *b*, *d* and *e*). This analysis, using only spectra of dried surface soil samples, was only performed for one of the vineyards (*Quinta dos Carvalhais*). In the other vineyard (*Quinta de Azevedo*), because each block

only had one sampling location, separation according to soil type within the same block was not feasible. For block *a* (Fig. 3.1.4a) results show the four different soil types clearly separated. A very similar result was obtained for the remaining blocks (Fig. 3.1.4b to 3.1.4e) where it was possible to identify separation (more or less evident depending on the different soil types). A global PCA model considering all blocks was also performed. The resulting score plot (data not shown) proved itself rather confusing and did not show a good separation amongst soil types. This result is in accordance with the previous observation (Fig. 3.1.3) that the same soil type occurring in different blocks formed different clusters in the PCA scores plot.

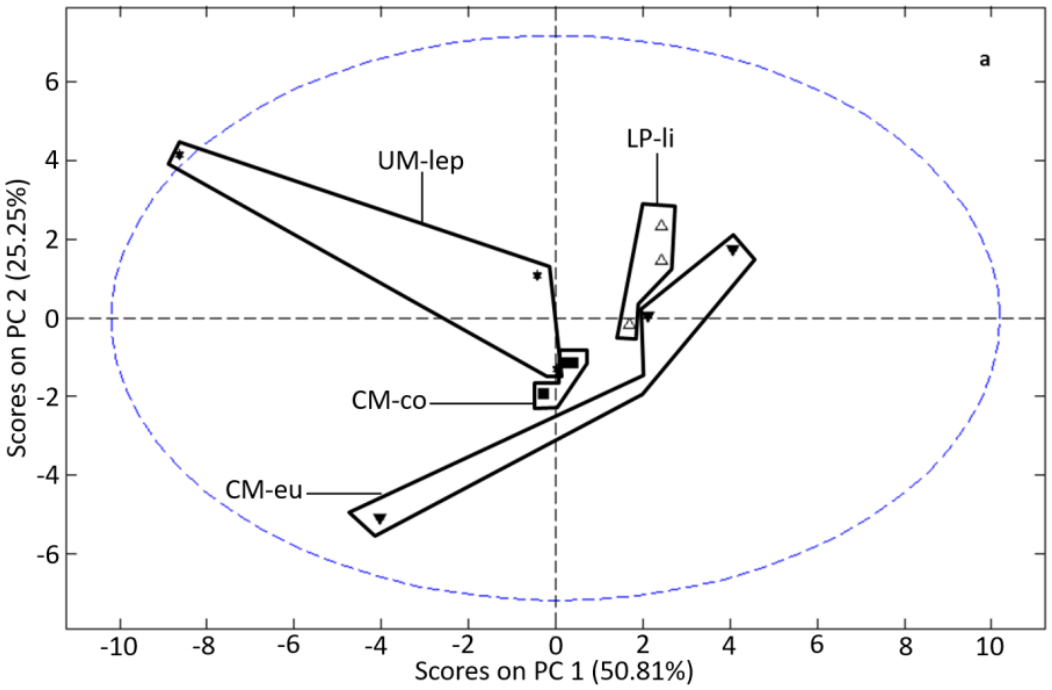
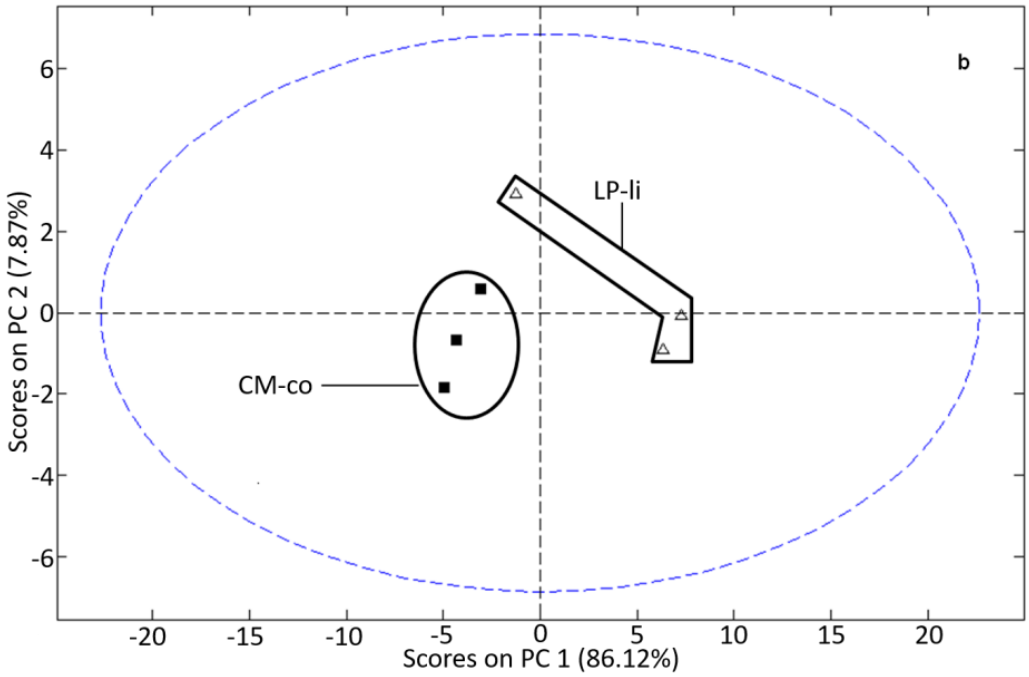


Figure 3.1.4. Score plots obtained from PCA models built on NIR spectra from shallower depth samples collected on different blocks of *Quinta dos Carvalhais*: *a* to *e* (score plots points are displayed according to the soil type, with acronyms explained in Table 1). Note: In figure 4c, the un-labelled points (■) belong to the soil type CM-co.



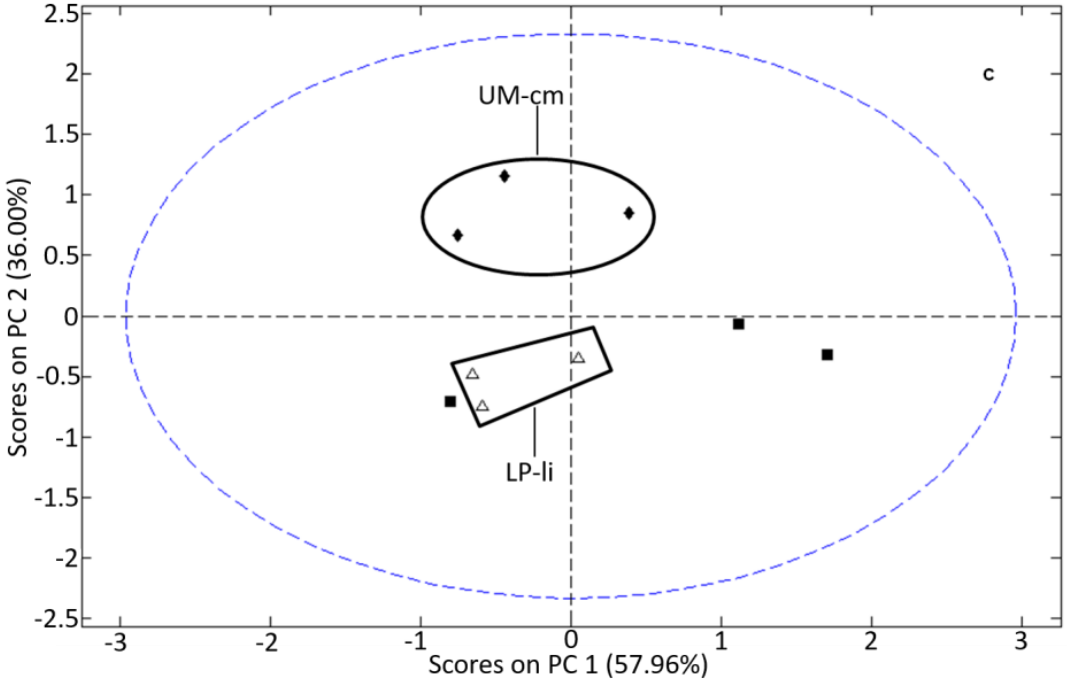
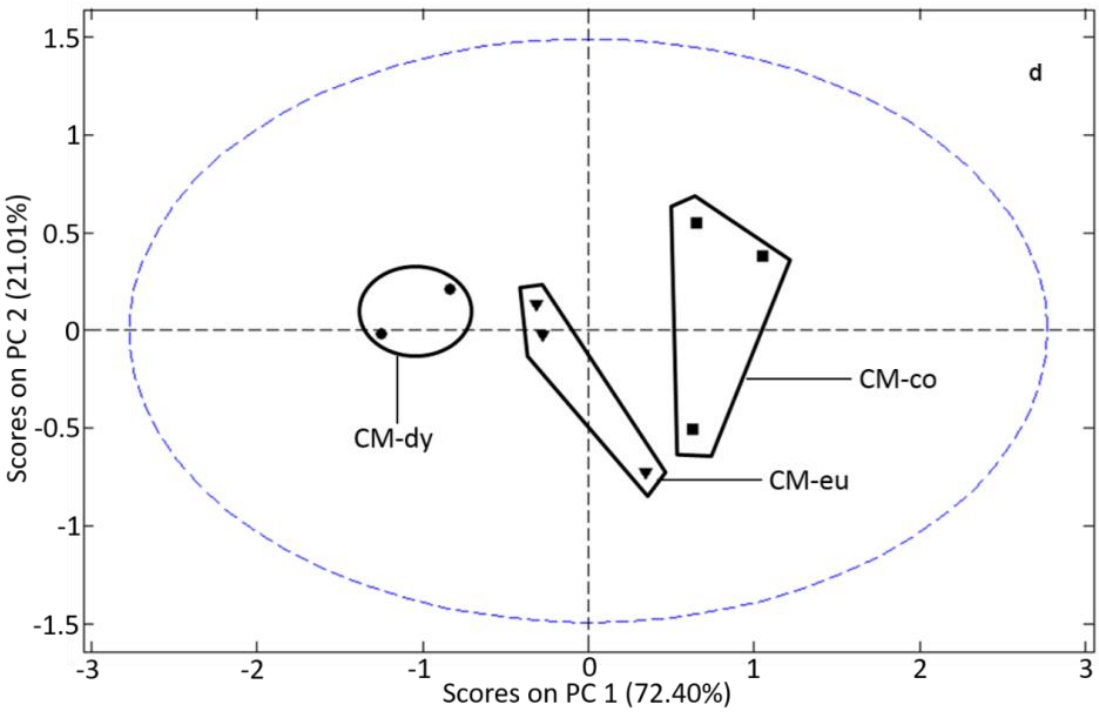


Figure 3.1.4. **(cont.).** Score plots obtained from PCA models built on NIR spectra from shallower depth samples collected on different blocks of *Quinta dos Carvalhais*: *a* to *e* (score plots points are displayed according to the soil type, with acronyms explained in Table 1). Note: In figure 4c, the un-labelled points (■) belong to the soil type CM-co.



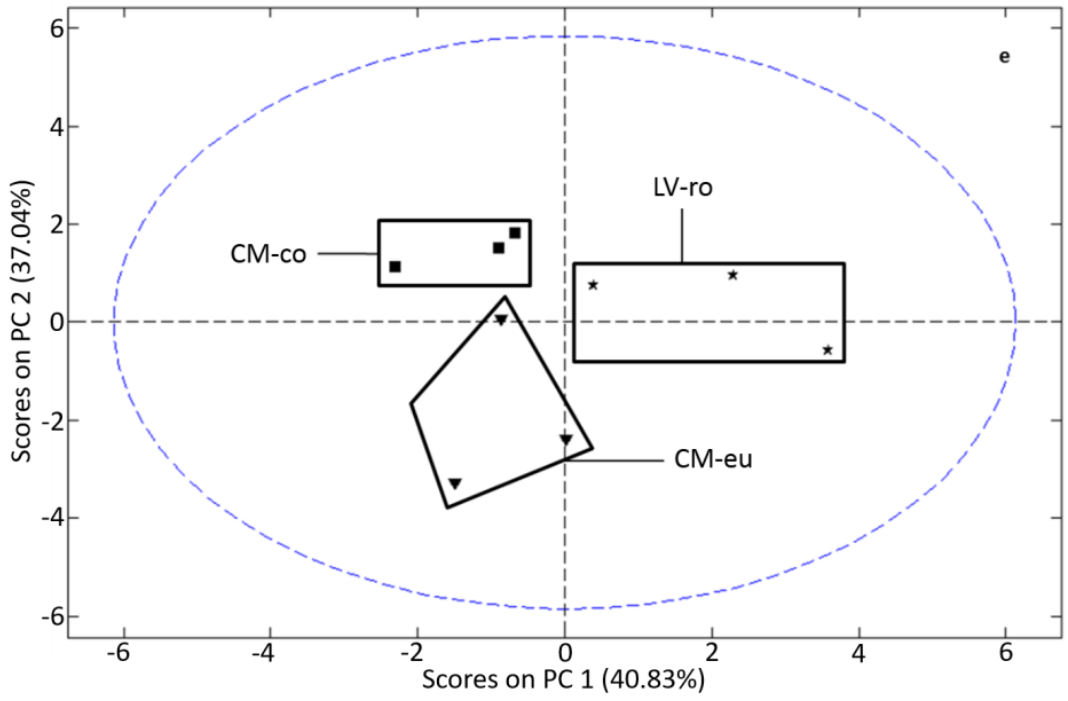


Figure 3.1.4. **(cont.).** Score plots obtained from PCA models built on NIR spectra from shallower depth samples collected on different blocks of *Quinta dos Carvalhais*: *a* to *e* (score plots points are displayed according to the soil type, with acronyms explained in Table 1). Note: In figure 4c, the un-labelled points (■) belong to the soil type CM-co.

Table 3.1.3. Soil discrimination PLS-DA models for *Quinta dos Carvalhais* considering individual blocks with indication of the respective number of latent variables.

Block	Dried samples		Dried-ground samples	
	Correct prediction rate (%)	LVs	Correct prediction rate (%)	LVs
a (n=36)	78	7	84	7
b (n=18)	93	3	99	3
c (n=27)	86	5	94	6
d (n=27)	93	5	84	6
e (n=27)	91	3	97	4

3.1.3.2. PLS-DA

To extend the previous analysis with PCA, supervised PLS-DA models were performed, firstly on each block separately. Models for *Quinta dos Carvalhais* were calibrated on wet, dried and dried-ground samples, using only the superficial layer (Table 3.1.3). It was observed that the milling process allowed better results. Roughly, the percentage of correct predictions for milled samples is 5% higher than when considering the dried samples independently of the considered vineyard block. It should be emphasized that even though the results obtained for individual blocks come from a limited data set, which may lead to overfitting, the number of samples were higher than the number of variables [34]. For that reason that reason, the same model type was then applied considering all vineyard blocks simultaneously, hereby designated as global models (Table 3.1.4). Two models were developed and validated: *Quinta dos Carvalhais* and *Quinta de Azevedo*. Again, models for dried-ground samples appear to have a better performance when compared with wet or dried sample models. In all cases, it was observed that the PLS-DA model was much better at correctly predicting soil type when analysing individual blocks than when analysing several blocks together (compare Tables 3.1.3 and 3.1.4). Overall, soil type was correctly predicted 79.3% of the time using dried-ground samples in *Quinta dos Carvalhais.* For wet soil samples the correct soil type was predicted at a 72.8% rate and at a 74.3% rate for dry samples, which reinforces our previous statement that there was no significant difference between using dry and wet soil samples. This contrasts with rates between 80 and 95% obtained when individual blocks are modelled. For *Quinta de Azevedo*, soil type was correctly predicted 88.6% of the time with dried-ground samples, contrasting with correct prediction rates of 79.9% for dried samples and 77.3% for wet

samples. This reinforces the idea that the milling process improves the model performance. The higher percentage of correct predictions found for *Quinta de Azevedo* is most likely due to a smaller number of different soils within the vineyard (three against seven in *Quinta dos Carvalhais*). Confusion matrices for both global models were obtained and analysed (Table 3.1.5 for *Quinta dos Carvalhais* and Table 3.1.6 for *Quinta de Azevedo*). For *Quinta dos Carvalhais* most bad predictions involve the soil types LP-li, CM-eu and CM-co. The remaining soils seem to be accurately predicted even though different blocks are considered. Bad predictions are for instance observed between CM-eu and CM-co soil types that are being incorrectly predicted with a 19.9% error rate. The same problem is observed between soil types LP-li and CM-co (17.9% error rate) and between UM-cm and CM-co (22.7% error rate). A total error rate of 7.4% was obtained for samples of both soils being incorrectly predicted. The PLS-DA model loadings (model weights) provide an insight about specific wavenumbers more involved in the soils discrimination. For simplicity only the first three principal components in the global model for *Quinta dos Carvalhais* were shown in Fig. 3.1.5 (the most important components encompassing more that 92.6% of variability in spectra). It is evident that there were two distinct wavenumber regions which were fundamental for the discrimination. These zones were between 4600-4000 cm^{-1} and 7300-7000 cm^{-1}. The first zone seems to be dominated by mineral compositions, namely 4630 cm^{-1} and 4529 cm^{-1} which are attributed to kaolin doublet from clay minerals; 4533 cm^{-1} is considered smectite and illite also from clay minerals and 4484 cm^{-1} is also smectite from clay minerals. The absorptions near 4619 cm^{-1} and 4537 cm^{-1} can be related to Al-OH bonds and at 4386 and 4396 cm^{-1} are ascribed to Fe-OH and aliphatic compounds respectively [35]. The second zone also has a strong dominance of minerals, but in this case shared with water bonds. Kaolin doublet from clay minerals at 7067 and 7169 cm^{-1}, hydroxyl bounds at 7143cm^{-1} and finally water bounds at 7082 cm^{-1} [36]. Very similar results were obtained for the *Quinta do Azevedo* global soil PLS-DA model (loadings not shown).

Table 3.1.4. Soils discrimination PLS-DA models for both vineyards considering all soil blocks (global models) with indication of the respective number of latent variables.

Samples form	*Quinta dos Carvalhais* (n=135)		*Quinta de Azevedo* (n=132)	
	Correct prediction rate (%)	LVs	Correct prediction rate (%)	LVs
Wet	73	10	77	10
Dried	74	10	80	10
Dried-ground	79	10	89	10

Table 3.1.5. Confusion matrix for the PLS-DA soil discrimination model based on the NIRS method applied to dried-ground samples of *Quinta dos Carvalhais* (79.3% overall correct prediction rate and 10 LVs, n=135). Values are in %.

Predicted soil types	Real soil types							
	CM-eu	UM-lep	LP-li	CM-co	UM-cm	LV-ro	CM-dy	Sum
CM-eu	**11.8**	0.0	0.1	2.1	1.3	0.0	0.0	15.4
UM-lep	0.0	**3.8**	1.1	0.0	0.3	0.0	0.0	5.1
LP-li	0.5	0.0	**18.3**	1.2	0.3	0.1	0.0	20.5
CM-co	3.3	0.5	4.3	**25.4**	2.1	0.3	0.0	35.9
UM-cm	0.4	0.0	0.0	2.0	**5.2**	0.1	0.0	7.7
LV-ro	0.0	0.0	0.0	0.3	0.0	**9.9**	0.0	10.3
CM-dy	0.4	0.0	0.0	0.0	0.0	0.0	**4.7**	5.1
Sum	16.5	4.3	23.8	31.1	9.2	10.4	4.7	100

Table 3.1.6. Confusion matrix for the PLS-DA soil discrimination model based on the NIRS method applied to dried-ground samples of *Quinta de Azevedo* (88.6% overall correct prediction rate and 10 LVs, n=132). Values are in %.

Predicted soil types	Real soil types			
	CM-dy	CM-eu	CM-co	Sum
CM-dy	**19.1**	0.6	0.3	20.0
CM-eu	1.0	**15.0**	4.0	20.0
CM-co	2.3	3.2	**54.5**	60.0
Sum	22.4	18.8	58.8	100

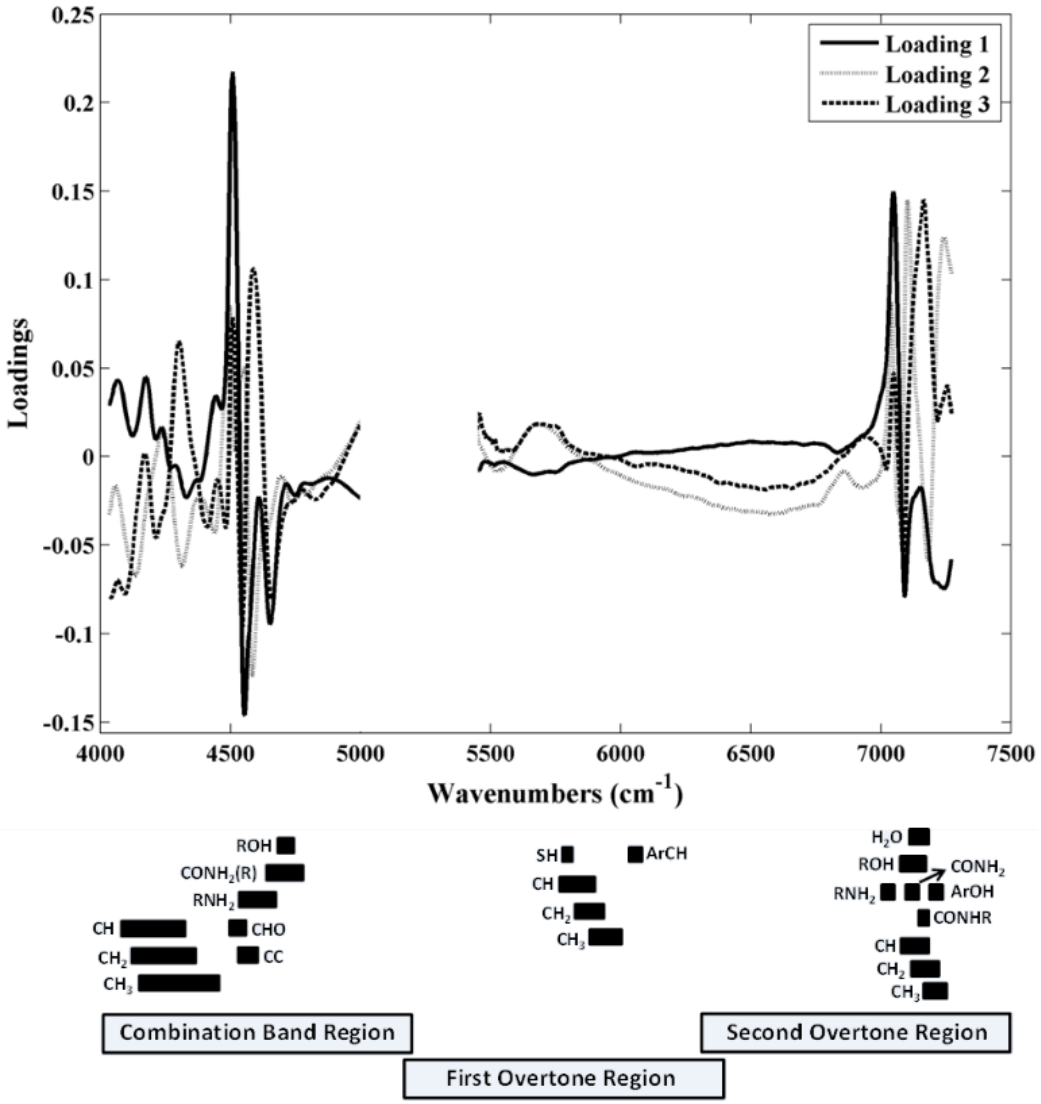


Fig. 3.1.5. Loadings for the PLS-DA model discriminating between all soil types of *Quinta dos Carvalhais* using dried-ground samples.

3.1.4. Discussion

The soil ability to capture and maintain water is an important factor and has to be considered when analysing soil for agricultural purposes. It is also known that water is extremely active in the NIRS range which will, therefore, influence the acquired spectra substantially [37]. This is very pertinent when moisture is being estimated from spectra, but less so when one wants to relate spectra with other soil properties. Several studies suggest that the variability in environmental conditions, especially soil moisture, decreases the predictive ability of NIRS [38]. Also, from direct observation, dried soil samples appear lighter in colour, whereas wet soils are darker, resulting in a deeper penetration of light and leading to greater light absorption and less reflection. Consequently, calibrations based on dry soil samples should exhibit more accurate results than those developed using wet samples [39]. However, some studies in the literature state that it is perfectly possible to analyse soil properties with NIRS using wet soil samples [9]. Some authors [40] even reported that NIRS calibrations for some soil chemical parameters, based on moist samples were slightly better than those obtained using dried ones. From our point of view, moisture does not bear a significant influence on the purpose of this study: there seems to be no significant difference between using wet or dry soil samples. Similar results were reported in a study where a comparison of soil spectra collected with a portable device and spectra collected in the laboratory was performed [41]. These authors concluded that for the most part, the differences between the spectra obtained by the two methods were related to soil water content and that those differences were not significant for soil characterization and mineral composition determination.

The sampling depth is another factor known to introduce variability in NIRS spectra. However, our results point to the fact that there is no substantial difference between dry samples from different depths, at least within the depths analysed in this study (Fig. 3.1.3). The depth analysis revealed, quite surprisingly, a clear pattern regarding vineyard geographical distribution: samples cluster according to the block which they belong to, instead of grouping together according to depth distribution. Therefore, we believe NIRS is able to separate soil samples of the same soil type appearing in different geographical areas of the same vineyard.

Since the major goal of this work was the attempt to discriminate soil types with NIRS, a first attempt was performed considering localized vineyard zones (blocks) in one of the studied vineyards (*Quinta dos Carvalhais*). According to the PCA models of individual blocks, some soil types are apparently closely related or seem to present some

overlapping (Fig. 3.1.4a to 3.1.4e). This is probably due to the proximity of so many different soil types in such small areas and of course, some uncertainty in the reference method used to outline the soil type boundaries. Vineyard soil types have transition zones (more or less extensive) and biological/chemical/physical characteristics naturally overlap each other in these areas. However, it should be stated that PCA model score plots presented only comprise information of the first two PC's (due to visual convenience), which means that there might be relevant information in subsequent PC's not taken into account. Additionally, scores are being extracted according to the criteria of maximizing variance, and give no information regarding soil types. For these reasons, a more adequate analysis regarding soil discrimination was required and the supervised method PLS-DA was adopted. PLS-DA model results corroborate the assumption that NIRS is able to distinguish rather well soil types within a localized geographical region (the block), but when several blocks are analysed in conjunction that separation is not that evident (Table 3.1.4). Furthermore, this may mean that samples of the same soil type, coming from different blocks, present specific characteristics which are discernible by NIRS, and that NIRS consequently recognizes these samples as being slightly different. PLS-DA scores analysis confirmed what was previously observed in the simple exploratory data analysis performed with PCA modelling. Dried-ground samples presented better correct predictions than intact dried ones, a result also observed in the other vineyard (*Quinta de Azevedo*) studied in this work. This is probably because milling leads to a more constant particle size, which will have an effect on the quality of the spectra as has been corroborated by other works [42]. Wetterlind and co-workers also concluded that milling soil samples increases the overall reflectance of the spectra [43]. However, reports in the literature regarding performance of results based on ground and intact soil samples are still contradictory [44, 45]. Some authors even infer that milling samples does not improve prediction accuracy [46]. Within this study, ground soil samples indeed presented more consistently correct predictions, but not significantly higher than intact soil samples.

The high cost of soil sampling as well as biological, chemical and physical analysis by reference methods has clearly spawned the urge for the development of alternative cost-effective solutions. NIRS proved to be part of that solution as an essential technology for fast and accurate soil mapping of areas of particular agricultural and economical interest. Further work should include experiments to establish a possible NIRS detectable correlation between the soil type and vines/grapes/leaves characteristics. The purpose would be to verify the extent to which soil spectral maps can be correlated with plants/leaves spectral maps and which chemical/physical/biological parameters are mostly involved. This can ultimately provide trademark characteristics and become a decision

support tool to direct wine quality towards established targets (quality-by-design). In light of this, it seems possible to apply NIRS in the field as a tool for monitoring variations in soil composition, and at the same time receive a quick estimate of some soil chemical properties. Furthermore, NIRS could also be used in the field for swift soil mapping, improving efficiency in resource usage and help define tailor-made strategies taking into consideration within-block variability where useful – leading to a more accurate and sustainable vineyard management.

3.1.5. Conclusions

This work demonstrated the ability of the high-throughput NIRS analytical method to discriminate between specific vineyard soil types. Soils characterized from two vineyards in Portugal (one in the Dão Wine Region and the other in the Vinho Verdes Wine Region) were employed in this study. Results revealed that, as expected, NIR spectra were dominated by clay minerals and water. The proposed method was able to accurately differentiate distinct soil types, which were well correlated (around 90%) to those obtained by pedology methods when the method was applied to specific blocks and with a slightly lower success rate (around 80%) when all blocks were analysed simultaneously. Incorrect predictions were most probably due to the uncertainty of the reference method to outline defined soil type boundaries. Vineyard soil types evidently have transition zones (more or less extensive) and chemical/physical characteristics naturally overlap in the transition zones. NIRS proved to be sensitive to these changes and can, therefore, provide a real fingerprint of the soil type. This work also showed that there is no significant difference between wet and dried samples for soil differentiation purposes and that milled soil samples yielded only slightly better prediction results. This may be of the utmost importance for in-situ measurements, opening the possibility for direct field analysis with a portable NIRS instrument.

3.1.6 References

1. Barbeau, G., et al., Comportement du cépage Cabernet Franc dans différentes terroirs du Val de Loire. Incidence de la précocité sur la composition de la vendange en année climatique normale (exemple de 1988). Journal International des Sciences de la Vigne et du Vin, 1998b. **32**(2): p. 69-81.
2. Carey, V., Spatial characterization of natural terroir units for viticulture in the Bottelaryberg-Simonsberg-Helderberg winegrowing area, in Private Bag XI. 2001, Stellenbosch University: Matieland (Stellenbosch), South Africa.
3. Tesic, D., et al., Environmental effects on cv Cabernet Sauvignon (Vitis vinifera L.) grown in Hawke's Bay, New Zealand. 1. Phenology and characterisation of viticultural environments. Australian Journal of Grape and Wine Research, 2002. **8**(1): p. 15-26.
4. Bodin, F. and R. Morlat, Characterization of viticultural terroirs using a simple field model based on soil depth I. Validation of the water supply regime, phenology and vine vigour, in the Anjou vineyard (France). Plant and Soil, 2006. **281**(1-2): p. 37-54.
5. Catarino, S., et al., Rare Earths Data for Geographical Origin Assignment of Wine: A Portuguese Case Study. Bulletin de l'OIV, 2011. **84**: p. 333-346.
6. Greenough, J.D., H.P. Longerich, and S.E. Jackson, Element fingerprinting of Okanagan Valley wines using ICP-MS: Relationships between wine composition, vineyard and wine colour. Australian Journal of Grape and Wine Research, 1997. **3**(2): p. 75-83.
7. Eschnauer, H., et al., USE AND LIMITATIONS OF ICP-OES IN WINE ANALYSIS. Mikrochimica Acta, 1989. **3**(3-6): p. 291-298.
8. Costantini, E.A.C., et al., Soil and climate functional characters for grape ripening and wine quality of "Vino Nobile di Montepulciano", in First Ishs Workshop on Strategies to Optimize Wine Grape Quality, S. Poni, et al., Editors. 1996. p. 45-55.
9. Cozzolino, D., et al., In Situ Measurement of Soil Chemical Composition by Near-Infrared Spectroscopy: A Tool Toward Sustainable Vineyard Management. Communications in Soil Science and Plant Analysis, 2013. **44**(10): p. 1610-1619.
10. Deaville, E.R. and P.C. Flinn, Near infrared (NIR) spectroscopy: an alternative approach for the estimation of forage quality and voluntary intake, in Forage Evaluation in Ruminant Nutrition, O.E. Givens DI, Axford RFE and Omedi HM, Editor. 2000, CABI Publishing: Wallingford, UK. p. 301 – 320.
11. Murray, I., Forage analysis by near infrared spectroscopy, in Sward Management Handbook, B.R. Davies A, Grant SA and Laidlaw AS, Editor. 1993, British Grassland Society: UK. p. 285–312.

12. Osborne, B.G., T. Fearn, and P.H. Hindle, Near Infrared Spectroscopy in Food Analysis. 1993, Essex, England, UK: Longman Scientific and Technical.
13. Theo, J., Short introduction to infrared and Raman spectroscopy, in The application of vibrational spectroscopy to clay minerals and layered double hydroxides, J. Theo, Editor. 2005, Clay Minerals Society: Chantilly.
14. Viscarra Rossel, R.A. and A.B. McBratney, Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. Geoderma, 1998. **85**(1): p. 19-39.
15. Viscarra Rossel, R.A., et al., Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, 2006. **131**(1-2): p. 59-75.
16. Bronick, C.J. and R. Lal, Soil structure and management: a review. Geoderma, 2005. **124**(1): p. 3-22.
17. Magalhães, L.M., et al., Rapid assessment of bioactive phenolics and methylxanthines in spent coffee grounds by FT-NIR spectroscopy. Talanta, 2016. **147**: p. 460-467.
18. Cozzolino, D. and A. Moron, Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. Soil & Tillage Research, 2006. **85**(1-2): p. 78-85.
19. Dalal, R.C. and R.J. Henry, Simultaneous determination of moisture, organic-carbon, and total nitrogen by near-infrared reflectance spectrophotometry. Soil Science Society of America Journal, 1986. **50**(1): p. 120-123.
20. Reeves, J.B., G.W. McCarty, and J.J. Meisinger, Near infrared reflectance spectroscopy for the analysis of agricultural soils. Journal of near Infrared Spectroscopy, 1999. **7**(3): p. 179-193.
21. Stevens, A., et al., Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. Geoderma, 2008. **144**(1-2): p. 395-404.
22. Zimmermann, M., J. Leifeld, and J. Fuhrer, Quantifying soil organic carbon fractions by infrared-spectroscopy. Soil Biology & Biochemistry, 2007. **39**(1): p. 224-231.
23. Sarathjith, M. C., Das, B. S., Vasava, H. B., Mohanty, B., Sahadevan, A. S., Wani, S. P., & Sahrawat, K. L. (2014). Diffuse reflectance spectroscopic approach for the characterization of soil aggregate size distribution. *Soil Science Society of America Journal*, **78**(2), 369-376.
24. Wu, C. Y., Jacobson, A. R., Laba, M., & Baveye, P. C. (2009). Accounting for surface roughness effects in the near-infrared reflectance sensing of soils. *Geoderma*, *152*(1), 171-180.

25. Baumgardner, M. F., Silva, L. F., Biehl, L.L., & Stoner, E. R.. 1985. Reflectance properties of soils. *Advances in Agronmy*, **38**, 1–44.
26. Pessanha, M. and A. Graça. Increasing efficiency in production management of a wine enterprise: the project I.C.O.N.E. in Proceedings of the 34th. World Congress of Vine and Wine. 2011. Porto, Portugal.
27. FAO-WRBSR. International union of soil science (IUSS) working group world reference base. World reference base for soil resources. In *International soil classification system for naming soils and creating legends for soil maps*. 2014, Rome: FAO.
28. Naes, T., et al., Interpreting PCR and PLS solutions, in A User-Friendly Guide to Multivariate Calibration and Classification. 2004, NIR Publications: Chichester, UK. p. 39-54.
29. Barker, M. and W. Rayens, Partial least squares for discrimination. Journal of Chemometrics, 2003. **17**(3): p. 166-173.
30. Geladi, P. and B.R. Kowalski, Partial least-squares regression - a tutorial. Analytica Chimica Acta, 1986. 185: p. 1-17.
31. Preisner, O., et al., Application of Fourier Transform Infrared Spectroscopy and Chemometrics for Differentiation of Salmonella enterica Serovar Enteritidis Phage Types. Applied and Environmental Microbiology, 2010. 76(11): p. 3538-3544.
32. Borrelli, P., et al., Assessment of the cover changes and the soil loss potential in European forestland: First approach to derive indicators to capture the ecological impacts on soil-related forest ecosystems. Ecological Indicators, 2016. **60**: p. 1208-1220.
33. Rossel, R.A.V. and C. Chen, Digitally mapping the information content of visible-near infrared spectra of surficial Australian soils. Remote Sensing of Environment, 2011. **115**(6): p. 1443-1455.
34. Grootveld, M. (2012). Introduction to the applications of chemometric techniques in 'omics' tesearch: common pitfalls, misconceptions and 'rights and wrongs'. In M. Grootveld (Ed.), *Metabolic profiling: disease and xenobiotics*, (pp. 1-34). London: Royal Society of Chemistry.
35. Rossel, R.A.V. and T. Behrens, Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma, 2010. **158**(1-2): p. 46-54.
36. Bishop, J.L., et al., Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite-serpentines, chlorites and micas. Clay Minerals, 2008. **43**(1): p. 35-54.

37. Tian, J. and W.D. Philpot, Relationship between surface soil water content, evaporation rate, and water absorption band depths in SWIR reflectance spectra. Remote Sensing of Environment, 2015. **169**: p. 280-289.
38. Wijewardane, N.K., Y.F. Ge, and C.L.S. Morgan, Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. Geoderma, 2016. **267**: p. 92-101.
39. Stevens, A., et al., Detection of carbon stock change in agricultural soils using spectroscopic techniques. Soil Science Society of America Journal, 2006. **70**(3): p. 844-850.
40. Waiser, T.H., et al., In situ characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy. Soil Science Society of America Journal, 2007. **71**(2): p. 389-396.
41. Viscarra Rossel, R.A., et al., In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma, 2009. **150**(3–4): p. 253-266.
42. Dunn, B.W., et al., The potential of near-infrared reflectance spectroscopy for soil analysis - a case study from the Riverine Plain of south-eastern Australia. Australian Journal of Experimental Agriculture, 2002. **42**(5): p. 607-614.
43. Wetterlind, J., B. Stenberg, and R.A. Viscarra Rossel, Soil analysis using visible and near infrared spectroscopy, in Plant Mineral Nutrients: methods and protocols., F.J.M. Maathuis, Editor. 2013, Humana Press, Springer: New York. p. 95-107.
44. Barthes, B.G., et al., Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: influence of replication and sample grinding and drying. Journal of near Infrared Spectroscopy, 2006. **14**(5): p. 341-348.
45. Russell, C.A., Sample preparation and prediction of soil organic matter properties by near infra-red reflectance spectroscopy. Communications in Soil Science and Plant Analysis, 2003. **34**(11-12): p. 1557-1572.
46. Fystro, G., The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis-NIR spectroscopy and comparative methods. Plant and Soil, 2002. **246**(2): p. 139-149.

The following work was adapted from the article:

Lopo, M., Páscoa, R. N., Graça, A. R., & Lopes, J. A. (2016). Classification of Vineyard Soils Using Portable and Benchtop Near-Infrared Spectrometers: A Comparative Study. *Soil Science Society of America Journal*, *80*(3), 652-661.

3.2. Classification of vineyard soils using portable and benchtop near infrared spectrometers: a comparative study

Abstract: Soils characterization is often accomplished by means of extensive field observations followed by laboratory analysis which are extremely time consuming and can be prohibitively expensive. Trying to address this issue, there is an on growing interest in using near infrared spectroscopy (NIRS) as a rapid and cost-effective tool for the prediction of soil's physical, biological, and chemical properties. This method is non-destructive, and provides spectra highly characteristic of soil properties and composition, enabling the analysis of many soil properties with a single measurement. The purpose of this study was the comparison between the quality of *in-situ* measurements with a dispersive NIR instrument (portable) device with a benchtop Fourier-transform NIR instrument (laboratory), thus investigating the potential of NIRS as a rapid and low-cost technique to map vineyards soil both in the field and in the laboratory. Soil samples collected from different areas of a soil fully characterized vineyard in the center of Portugal, were analyzed by NIRS and spectra were modelled by principal component analysis and partial least squares discriminant analysis. Both instruments proved to be able to differentiate the analyzed soil types. When samples are collected from nearby locations (e.g., within the same vineyard block), 75 to 100% successful soil identification rates are achieved depending on the soil type. Lower prediction percentages (around 70-75%) are obtained when soils from the entire vineyard are analyzed simultaneously. Results obtained with the portable instrument were, up to some point unexpectedly, equivalent to those obtained with the laboratory instrument.

3.2.1. Introduction

The conventional soil survey is accomplished by means of extensive field observations (sometimes very subjective) followed by laboratory analysis, which adds valuable information about the soil's properties in question [1]. With most used laboratory techniques (e.g., wet chemistry), soil analysis is often extremely time consuming and can be prohibitively expensive. Even though chemical analyses are effective, these techniques are generally restricted to specific areas being applied to a limited number of samples for thorough soil characterization. Such data will have little or no representative information on the spatial variability of soils in a large region [2]. Due to cost and time-consuming operation, such analyses are difficult to apply to large scale areas. For this reason, there is a need of quick, reliable and cost-effective technique able of being applied *in-situ* for soil analysis.

There is an on growing interest in using near infrared spectroscopy (NIRS) as a rapid and cost-effective tool for the prediction of soil's physical, biological, and chemical properties [2, 3]. Furthermore, NIRS requires minimal or no sample preparation, avoiding thus the use of environmentally harmful chemicals in the laboratory [4]. Infrared spectra of soils contain extensive information on the molecular and compositional chemistry (Rossel and Walter, 2004). These provide more information on the patterns of soil variation than conventional surveying, where only a few very accurate measurements are used enabling a better understanding of soils as a complete environmental system and as a long-term resource [5]. More importantly, this can be done via portable NIR devices without opening trenches or sending samples to the laboratory. Another advantage of NIRS portability is the possibility to estimate parameters that can only be measured *in-situ*. Most soil analysis using portable instruments are performed with the objective of measuring a specific set of parameters such as soil moisture content [6, 7], total N content [8], organic C [7], among others. One of the main disadvantages that may be pointed out to NIRS is that it is of limited use to monitor soil contamination by heavy metals [9]. A few models of portable NIR spectrometers of several brands are available, such as the microPHAZIR from Thermo Scientific; AgriSpec, FieldSpec4 and LabSpec4 from ASD Inc. or USB4000 from Ocean Optics [3].

The main aim of this study was to provide a comparison between the quality of *in-situ* measurements obtained with a portable NIR device with a benchtop Fourier-transform NIR spectrometer (FT-NIR) in the laboratory. Furthermore, the potential of NIRS as a rapid and low-cost technique to map vineyard soils both in the field and in the laboratory is also analyzed. The work presented in this study was based on a limited number of soil

types (seven soil types) and two specific NIR instruments, therefore should be considered a feasibility study. For this purpose, soil samples were measured on a vineyard previously characterized with pedology methods, collected and brought to the laboratory to be analyzed with a benchtop FT-NIR instrument. It is commonly known that portable instruments often have limitations when compared to benchtop FT instruments. Nevertheless, the use of portable NIR can work as a first-hand analysis of soil, giving valuable information for the development of a robust soil mapping model that can be further complemented by a laboratory instrument under more controlled measurement conditions.

3.2.2. Materials and methods

3.2.2.1. Samples collection and processing

Soil samples were collected from a 60 ha vineyard (*Quinta dos Carvalhais*, Mangualde, 40°33'20"N 7°46'40"W) in the Dão Wine Region, center of Portugal (Fig. 3.2.1). Soils of this vineyard were previously characterized by pedology reference methods in the scope of another project [10] and classified according to the International Union of Soil Science (IUSS) [11] and several constituents and properties (soil type, total nitrogen, pH, and organic matter) were measured for each soil type (Table 3.2.1). The vineyard is divided into "blocks" (listed *a* to *e*), each containing a single variety (*Vitis vinifera* cultivar). Forty-five sampling spots encompassing seven different soil types were identified on the vineyard. These spots were defined according to the mapped soil characteristics (and also grapes varieties). The sampling strategy required digging one hole at each spot. An approximately 500g soil sample was collected from each hole at specific pre-defined depths according to each soil analyzed (Table 3.2.2). Depths were chosen according to each type of soil and what was expected to bear more interesting results from a pedology perspective [10]. Immediately after collection, every sample was analyzed by introducing the portable's instrument probe in 15 different places. Samples were then transported to the laboratory and placed in an oven at 45°C for two weeks (Raypa DO-20, Barcelona, Spain). Dried samples were then measured with the benchtop NIR spectrometer.

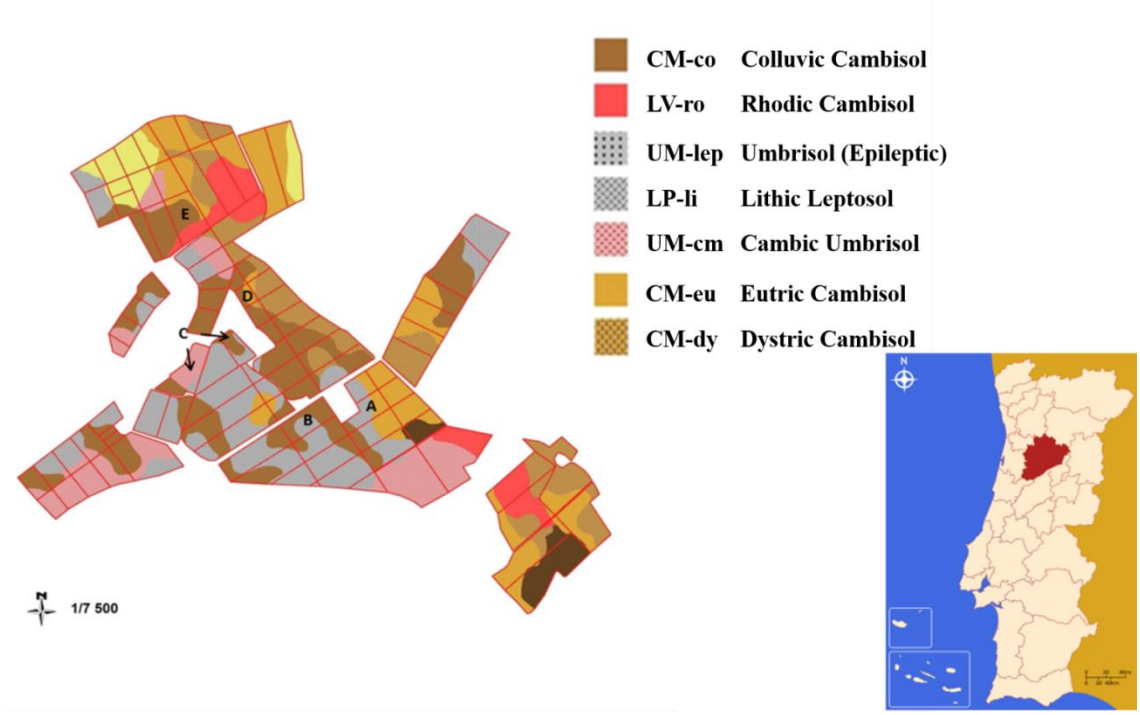


Figure 3.2.1. Vineyard map with indication of soil types with selected blocks (A to E) and representation of the Dão wine region (shaded region within Portugal's map).

3.2.2.2. Near infrared spectral acquisition

Near infrared spectra of *in-situ* samples were collected using a portable dispersive NIR spectrometer (model NIR-512, Ocean Optics, Dunedin, FL). This spectrometer features a temperature-regulated 512-element indium-gallium-arsenide (InGaAs) array detector effective in the 866-1670 nm wavelength range (11550-6000 cm^{-1}) providing a spectral resolution less than 5.0 nm (full width at half maximum). Spectra were acquired in diffuse reflectance mode, using a reflectance fiber optical probe (SabIR, ThermoNicolet, Madison, WI) with an irradiation area of 0.03 cm^2. Temperature, integration time and number of scans were set to -4°C, 1.5s and 40 respectively. Each NIR spectrum was obtained by averaging 40 spectra. Background was measured at the beginning of each analysis using Spectralon^R (teflon) as reference. The SpectraSuite software (Ocean Optics, Dunedin, FL) was used for spectrometer configuration, control and data acquisition. After each sample collection, 15 spectra were acquired from different sample locations, yielding 675 *in-situ* samples spectra (45 spots x 15 spectra).

Table 3.2.1 Soil types, textures (according to IUSS, 2014) and selected properties obtained at specific depths.

Soil Type	Soil taxonomy (suborder)[1]	Soil Texture	Depth (cm)	Total nitrogen (g/Kg)	pH	Organic matter (g/Kg)
Eutric Cambisol (CM-eu)	Udepts	Sandy Loam	0-20	0.61	6.8	13.2
Umbrisol (Epileptic) (UM-lep)	Anthrepts	Sandy Loam	0-20	[2]	6.8	3.4
Lithic Leptosol (LP-li)	Rendolls	Sandy Loam	0-10	0.7	5.2	10.0
Colluvic Cambisol (CM-co)	Udepts	Sandy Loam	0-20	0.86	6.5	19.1
Cambic Umbrisol (UM-cm)	Anthrepts	Sandy Loam	0-25	0.6	7	7.1
Rhodic Luvisol (LV-ro)	Udalfs	Sandy Loam	0-15	0.68	7	16.7
Dystric Cambisol (CM-dy)	Udepts	Sandy Loam	0-10	[2]	6.5	40.2

[1] From Krasilnikov 2009

[2] Values not measured.

Spectra of dried soil samples were collected in diffuse reflectance mode on a Fourier-transform near infrared spectrometer (FTLA 2000, ABB, Québec, Canada) equipped with an InGaAs detector. Each spectrum was recorded as the average of 64 scans with 8 cm^{-1} resolution over a wavenumber interval between 1000 and 2500 nm (10000-4000 cm^{-1}). The spectrometer was controlled via the Grams software (version 7, ABB, Québec, Canada). Soil samples were transferred into borosilicate flasks and measured on a diffuse reflectance accessory equipped with an integration sphere. The irradiation area for the benchtop equipment was 0.9 cm^2. Backgrounds were performed as for the *in-situ* samples. After drying, circa 30g of each sample was transferred to a borosilicate flask for spectral acquisition with the FT-NIR spectrometer and measured in triplicate (the average spectrum was considered). This process was repeated 9 times yielding 405 dry samples spectra (45 spots x 9 spectra).

Table 3.2.2 Sampling locations, soil types and depths.

Sampling spot	Block	Soil type[1]	Sampling depths(cm)
1-3	*a*	CM-eu	0-20
4-6	*a*	UM-lep	0-20
7-9	*a*	LP-li	0-10
10-12	*a*	CM-co	0-20
13-15	*b*	CM-co	0-20
16-18	*b*	LP-li	0-10
19-21	*c*	LP-li	0-10
22-24	*c*	CM-co	0-20
25-27	*c*	UM-cm	0-25
28-30	*d*	CM-eu	0-20
31-33	*d*	CM-co	0-20
34-36	*d*	CM-dy	0-10
37-39	*e*	LV-ro	0-15
40-42	*e*	CM-co	0-20
43-45	*e*	CM-eu	0-20

[1]Soil coding is explained in Table 3.2.1.

NIR spectra were divided in five regions according to the major chemical/physical properties captured by NIR spectra: region 1 (5000-4000 cm^{-1}), region 2 (5350-5000 cm^{-1}), region 3 (6700-5350 cm^{-1}), region 4 (7300-6700 cm^{-1}) and region 5 (10000-7300 cm^{-1}) (Bokobza, 1998). For the portable spectrometer, due to the limited detector range, regions 1 and 2 are not covered and region 3 is only partially covered. Regions 2 and 4 were delimited due to presence of water bands around 5250-5050 cm^{-1} and 7150-6900 cm^{-1} arising from the first O-H stretching overtone and O-H combination in the NIR spectra, respectively. Regions 1 to 4 capture essentially chemical information (combination and first overtone), while region 5 accounts essentially for physical information (second and third overtone).

3.2.2.3. Data analysis

Soils spectra were pre-processed with a Savitzky–Golay filter (15-points filter size, second order polynomial, and first-order derivative) [12] followed by the application of standard normal variate (SNV). Principal components analysis (PCA) [13] and partial least-squares discriminant analysis (PLS-DA) [14] were the chemometric methods used to analyze the spectral data. The PCA was used to extract common patterns from soil samples and to assist outlier detection. The PLS-DA was used to develop calibration models for soil discrimination purposes.

PLS-DA models were developed using the PLS-2 algorithm which handles multiple dependent variables which is the case here because PLS-DA codifies outputs (classes) in multiple variables [14]. Model predictions are converted in class assignments using the distribution of calibration predictions obtained from a PLS model built on two or more classes to determine the threshold level yielding the lowest level of false classifications. To every class prediction a probability level is assigned. In this work, only class assignments for which a probability value of more than 95% were considered [15]. To further investigate PLS-DA estimations precision, a strategy based on bootstrapping was performed. This strategy will generate an estimation of soils correct predictions distribution. The distribution was assessed by bootstrapping the PLS-DA models 1000 times. The available data were divided to form calibration and test sets. Spectra included in the calibration encompasses about 70% of each soil type with the remaining 30% used for the test set to prevent overfitting [16]. The best combination of the above mentioned spectral windows was estimated, by testing for each spectrometer, all possible combinations of spectral windows: 31 and 7 combinations for the benchtop (5 spectral windows) and portable (3 spectral windows), respectively. For each possible model, the optimal number of latent variables (LVs) was estimated by leave-one-block-out cross-validation (block sizes of 15 and 9 contiguous samples for the portable and benchtop spectrometers respectively) using only the calibration set and considering the root mean square error of cross validation (RMSECV) as the minimization criterion. Note that this process was performed considering a model encompassing all soil samples (all soil types). The optimized regions were used for all subsequent models. For all developed models, the test set was used to test the accuracy of the PLS-DA models and corresponding results expressed as confusion matrices. Confusion matrices compare each known soil type with the corresponding NIRS prediction and entries are expressed as percentages [16]. PLS-DA loadings were also analyzed in order to understand which specific wavenumbers are more important for soils discrimination.

Before applying PCA and PLS-DA the spectral sets were mean centered. All chemometric methods and spectra processing were performed using Matlab version 7.9 (MathWorks, Natick, MA) and the PLS Toolbox version 5.5.1 (Eigenvector Research Inc., Wenatchee, WA).

3.2.3. Results and discussion

A comparison between spectra acquired by both instruments for the same sample (CM-eu soil type sample) is shown in Fig. 3.2.2.

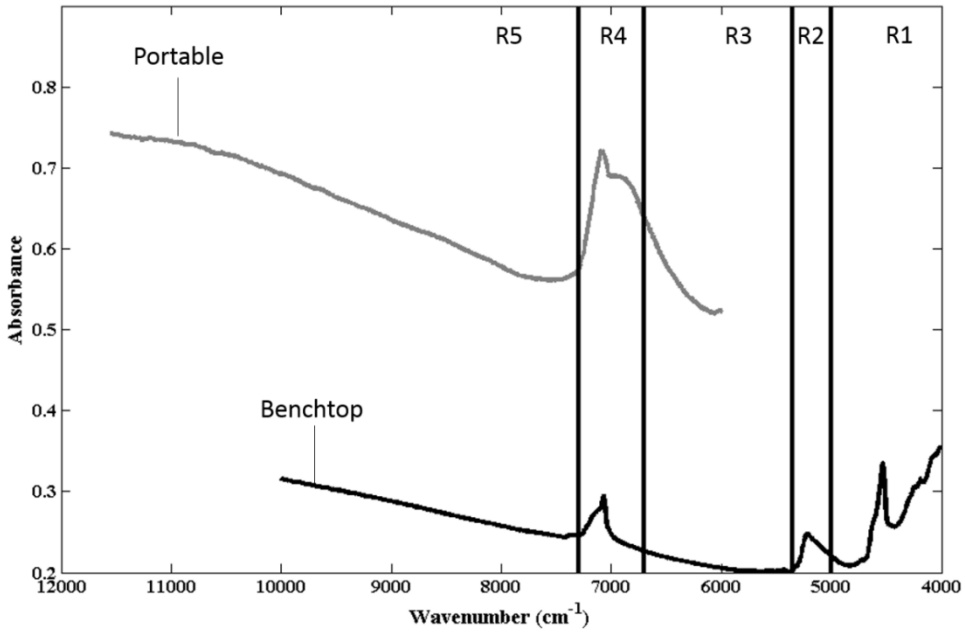


Figure 3.2.2. Comparison between a NIR spectrum obtained for a CM-eu soil sample using the portable instrument (measurement of the intact sample) and a benchtop instrument (measurement of the dried sample). R1: 5000-4000 cm^{-1}; R2: 5350-5000 cm^{-1}; R3: 6700-5350 cm^{-1}; R4: 7300-6700 cm^{-1}; R5: 10000-7300 cm^{-1}.

Separate PCA models for spectra acquired with the two instruments were developed, considering the maximum spectral window for each instrument, to identify any abnormal spectra. The eigenvalue (parameter that is proportional to the original spectra variance captured by each principal component) criterion for selecting number of components yielded 2 components for the portable instrument model (encompassing 60.5% of the total variance) and 3 components for the benchtop instrument model (encompassing 78.6% of the total variance). For the portable instrument, 15 spectra corresponding to the

measurements of a sample collected in block *e* belonging to the soil type CM-dy were considered abnormal both in the Hotelling's T^2 and squared residuals statistic (significantly above the 99% confidence limits) [17]. These spectra were, therefore removed and not used for further analysis. Quite surprisingly, these sample's spectra were not abnormal when the corresponding dried samples were analyzed with the benchtop instrument. From this observation, and because at this point no other treatment than drying had been applied to the samples, it is possible to assume that the reason for the abnormality in the *in-situ* sample's spectra may be the difference in water content between *in-situ* and dried samples. Be that as it may, and for consistency reasons, it was decided that these equivalent dried spectra would not be used in the modeling. Moisture in the soil may be highly variable [18] and it is known that water is extremely active in the NIRS range affecting significantly the spectra [19]. These facts may be very relevant when moisture is being estimated from spectra, but can prove itself counterproductive when other soil properties are investigated. PCA models of spectral data obtained in the individual blocks encompassing the highest quantity of different soil types (blocks *a* and *e*) were developed. To simplify presentation all spectra for each sample were averaged. Therefore, three spectral analyses are shown for each sampling spot (except for the CM-dy soil obtained from block e where one sample was excluded as detailed before). Calibrations were performed considering the entire spectral window for each instrument. Results show one block where a good separation according to soil type occurred both with the benchtop and portable instruments (Fig. 3.2.3c and 3.2.3d relative to block *e*) and one block where sample separation according to soil type was not as evident with both instruments (Fig. 3.2.3a and 3.2.3b relative to block *a*). It is interesting to note that PCA score plots generated from the portable instrument data are very similar to those obtained with the benchtop one. Measurements obtained with the portable instrument are therefore highly reliable even though they were obtained in the field with many variations, especially moisture. This finding is, in a somewhat ironic way, both in agreement and disagreement with the literature. Some studies suggest that the variability in environmental conditions, especially soil moisture, decreases the predictive ability of NIRS. Thus, the utilization of the portable NIRS equipment would be extremely limited, confined to very dry and hot areas, or to very specific times of year (e.g. summer peak), where soil moisture would supposedly be at its lowest. However, there are other studies that attest the possibility of analyzing soil properties with NIRS using *in-situ* soil samples [20]. Indeed, it has inclusively been reported that NIRS calibrations for some soil chemical parameters done *in-situ*, using a portable NIRS device were somewhat better than those obtained using dried ones [21]. In a comparison of soil spectra collected with a portable device and

spectra collected in the laboratory, it was concluded that for the most part, the differences between the spectra obtained by the two methods were related to soil water content and that those differences were not significant regarding soil mineral composition and chemical characterization [22].

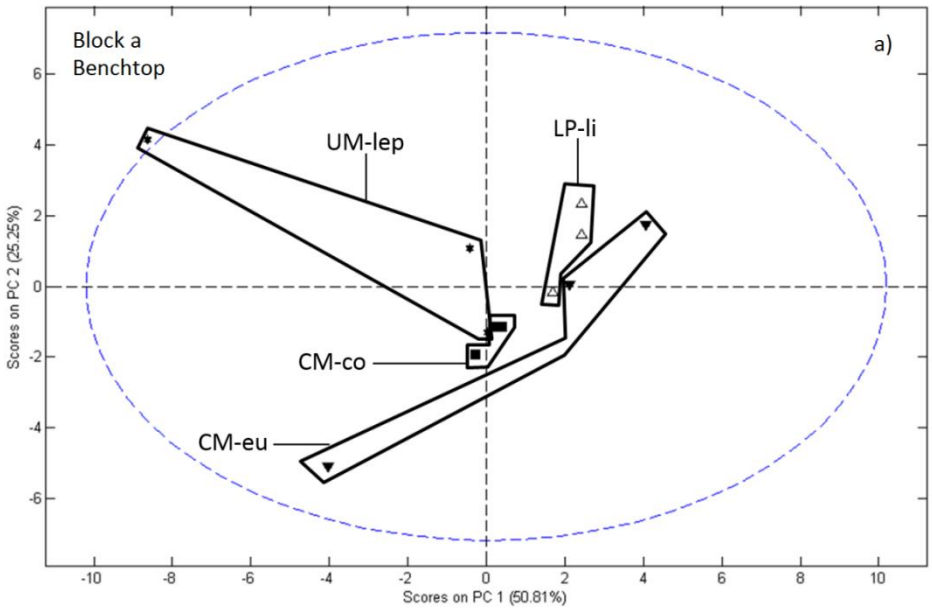


Figure 3.2.3a. Score plot obtained from PCA models built on NIRS data from samples collected on block *a* for benchtop instrument.

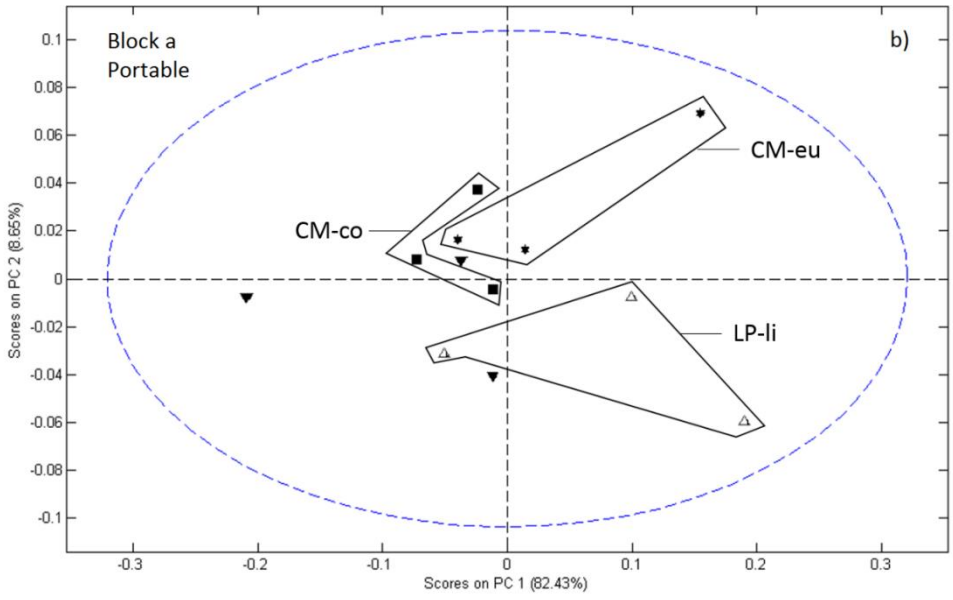


Figure 3.2.3b. Score plot obtained from PCA models built on NIRS data from samples collected on block *a* for portable instrument.

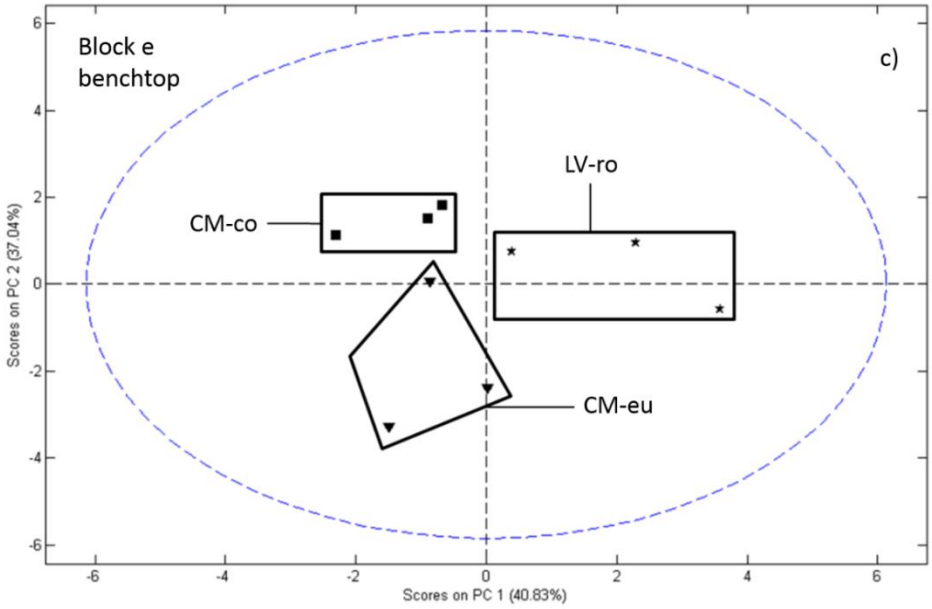


Figure 3.2.3c. Score plot obtained from PCA models built on NIRS data from samples collected on block *e* for benchtop instrument.

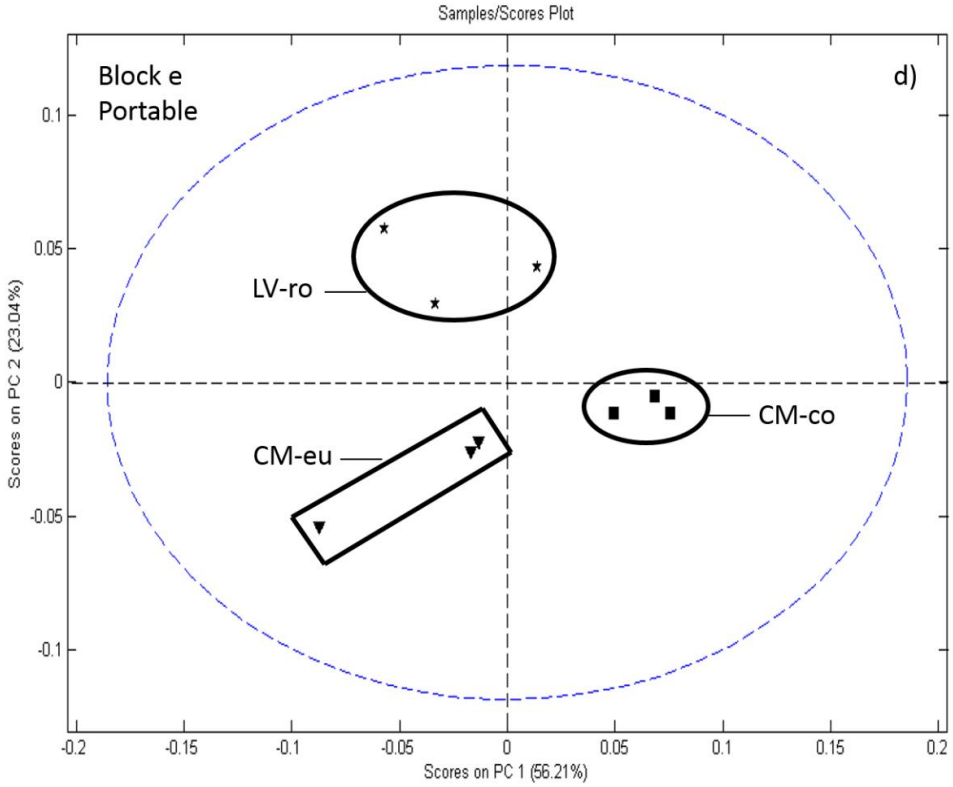


Figure 3.2.3d. Score plot obtained from PCA models built on NIRS data from samples collected on block *e* for portable instrument.

To further elucidate about the performance of both instruments to discriminate soils, regression analyses based on PLS-DA models were performed. The method was applied considering data from each block individually (individual models) and all the blocks (a global model) for both instruments (Table 3.2.3). A spectral window selection was performed according to the procedure described in the Materials and Methods section by calibrating models with all possible combinations of the five windows described. Models were calibrated for each instrument separately and considering all soils in all vineyard blocks. Only the calibration data was used and the best models were reported according to the minimum RMSECV. It was found for the portable instrument that the best spectral regions were 4 and 5. For the benchtop instrument the best spectral regions were 1, 3 and 4 (selected regions excludes region 2 encompassing the water combination band centered at circa 5200cm^{-1}). Both instruments included region 4 which may be an indicator that this particular region is the one that has most of the information related with the soil type. However, the benchtop instrument seems to provide more information than the portable one, since three different spectral regions were found to contain relevant information.

Models considering data obtained from each vineyard block and the global model were calibrated and optimized for the number of LVs considering the optimal spectral regions for each spectrometer. After the calibration step, the independent validation set was used to test the accuracy of the PLS-DA models. Overall, PLS-DA models yielded more percentage of correct predictions for soils when analyzing individual blocks than when analyzing the global model, both for benchtop and portable equipment (Table 3.2.3). Models developed with spectra obtained from the portable device exhibited a higher value of correct predictions when considering each block individually except for blocks *a* and *e*. Regarding the global model, the benchtop instrument was superior in terms of prediction accuracy. Predictions for the global models are further presented under the form of confusion matrices (Table 3.2.4). Globally, 70.9% and 75.2% correct prediction rates for the soil type were found for *in-situ* samples (portable) and dry samples (benchtop), respectively. The percentage values of correct predictions for each soil type are shown in Table 3.2.5. Correct prediction rates between 82 and 100% are obtained when soils in individual blocks are modelled. Most poor predictions involve the soil types LP-li, CM-eu and CM-co and are consistent in both instruments. These soils are present in several of the vineyard's blocks analyzed within this study, and therefore, were obtained from quite distant locations within the vineyard (9 sampling spots for CM-eu, 15 for CM-co and 9 for LP-li). This may indicate that there are small geographical variations within the area of the vineyard in terms of physical/chemical composition of the soils that may account for the

poor results obtained with these samples. This could only be validated by an extensive chemical/physical analysis. In a paper analyzing the variation of eutric cambisols (CM-eu) chemical properties based on altitudinal and geomorphologic zoning, Spârchez and co-workers found that slight variations occur in this soil and that these variations are based on several parameters. The authors further state that most chemical properties decrease with altitude [23].

Table 3.2.3. Results for the soil discrimination PLS-DA models considering individual blocks and globally, with respective number of LVs for benchtop and portable instruments.

Model	Block	Correct predictions rate (%)		Number of LVs		Variance (%)			
						Benchtop		Portable	
		Benchtop	Portable	Benchtop	Portable	X-block	Y-block	X-block	Y-block
Individual models	*a*	92.6	72.2	7	8				
	b	100	77.8	3	4				
	c	81.5	100	5	5				
	d	95.2	100	4	4				
	e	81.5	100	3	4				
Global model	All	75.2	70.9	10	10	97.8	39.0	99.2	40.9

Table 3.2.4. Confusion matrix for the soil discrimination model using the benchtop (75.2% of global correct predictions rate and 10 LVs) and portable (70.9% of global correct predictions rate and 10 LVs) equipment. All values are in %.

Predicted soil types		Real soil types							
		CM-eu	UM-lep	LP-li	CM-co	UM-cm	LV-ro	CM-dy	Sum
		benchtop/portable	benchtop/portable	benchtop/portable	benchtop/portable	benchtop/portable	benchtop/portable	benchtop/portable	benchtop/portable
CM-eu	benchtop/portable	13.7/6.8	0.0/09	0.0/0.9	0.9/3.4	0.9/0.0	0.0/1.7	0.0/1.7	15.4/15.4
UM-lep	benchtop/portable	0.0/0.9	2.6/3.4	1.7/0.9	0.9/0.0	0.0/0.0	0.0/0.0	0.0/0.0	5.1/5.1
LP-li	benchtop/portable	0.0/0.0	0.0/1.7	15.4/15.4	2.6/3.4	0.9/0.0	1.7/0.0	0.0/0.0	20.5/20.5
CM-co	benchtop/portable	0.9/2.6	0.0/2.6	7.7/6.0	22.2/23.9	0.9/0.9	4.3/0.0	0.0/0.0	35.9/35.9
UM-cm	benchtop/portable	0.0/0.0	0.0/0.0	0.0/0.0	1.7/0.0	6.0/7.7	0.0/0.0	0.0/0.0	7.7/7.7
LV-ro	benchtop/portable	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	10.3/10.3	0.0/0.0	10.3/10.3
CM-dy	benchtop/portable	0.0/1.7	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	5.1/3.4	5.1/5.1
Sum	benchtop/portable	14.5/12.0	2.6/8.5	24.8/23.1	28.2/30.8	8.5/8.5	16.2/12.0	5.1/5.1	100/100

Table 3.2.5. Percentage of correct predictions obtained for both instruments, based on the respective confusion matrices.

Correct predictions (%)		
Soil type	Benchtop	Portable
CM-eu	69.5	57.8
UM-lep	52.0	59.6
LP-li	78.5	74.6
CM-co	67.1	74.1
UM-cm	75.3	97.4
LV-ro	86.4	95.1
CM-dy	88.4	68.6

Regarding the use of different spectral regions in both instruments, it can be said that using region 1, 3 and 4 in the benchtop instrument did not improve the soils prediction classification accuracy. Thus, a portable NIR equipment able of collecting spectra in region 4 and 5 is enough to obtain the same results achieved with a benchtop instrument.
With the objective of understanding which part of the NIR signal used in the PLS-DA discrimination models is more important, the respective loadings were analyzed (Fig. 3.2.4a and 3.2.4b). Loadings provide an insight about specific wavenumbers involved in the soils discrimination. For simplicity, only the first two LVs were considered in the analysis. There are two distinct wavenumber regions fundamental for the discrimination of soils using the benchtop equipment. These zones are comprised at 4600-4000 cm^{-1} and 7300-7000 cm^{-1}. The first zone seems to be dominated by mineral compositions, namely 4630 cm^{-1} and 4529 cm^{-1} which are attributed to kaolin doublet from clay minerals; 4533 cm^{-1} is considered smectite and illite also from clay minerals and 4484 cm^{-1} is also smectite from clay minerals. The absorptions near 4619 cm^{-1} and 4537 cm^{-1} can be related to Al-OH bonds and at 4386 and 4396 cm^{-1} are ascribed to Fe-OH and aliphatic compounds respectively [24]. The second zone also has a strong dominance of minerals, but in this case shared with water bands. Kaolin doublet from clay minerals at 7067 and 7169 cm^{-1}, hydroxyl bounds at 7143cm^{-1} and finally water bounds at 7082 cm^{-1} [25]. Regarding the portable equipment, there seems to be only one wavenumber region that is fundamental for the discrimination of soils. That zone is comprised between 7300 and 7000 cm^{-1} in a similar fashion of the second zone for the benchtop equipment [26]. If the goal of the developed work was to confirm the presence of the aforementioned minerals, a

X-ray diffraction analysis should be performed. More details of this technique could be found in [27].

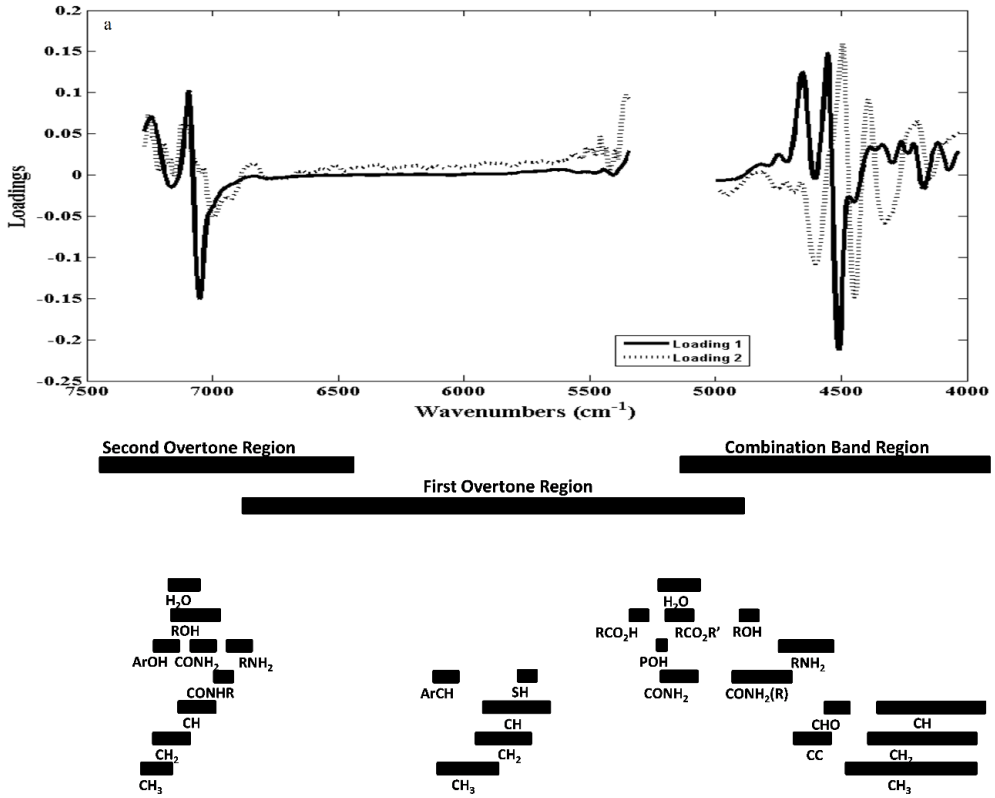


Figure 3.2.4a. Loadings plot obtained for the PLS-DA model discriminating between all soil types for benchtop instrument (X-block percentage of captured variance: 97.8% and Y-block percentage of captured variance: 39.0%).

In order to assess the distribution of correct predictions originated by the PLS-DA models, the number of correct predictions was repeatedly estimated following a bootstrapping strategy. The distributions of correct soil predictions are very similar when considering individual blocks or all blocks together for both portable and benchtop instruments (Fig. 3.2.5). All the blocks, when analyzed individually, gave correct predictions results higher than 90% with the exception of block “a” for both instruments. The distributions of the correct predictions for the global model when using both instruments are in the range of 70%. All these results in conjunction with the results shown in Table 3.2.3, corroborate that using a portable or benchtop equipment for soil type discrimination has no significant difference. The narrow spectral window as well as the lower resolution of the portable equipment did not seem to be an important disadvantage for the purpose of this study.

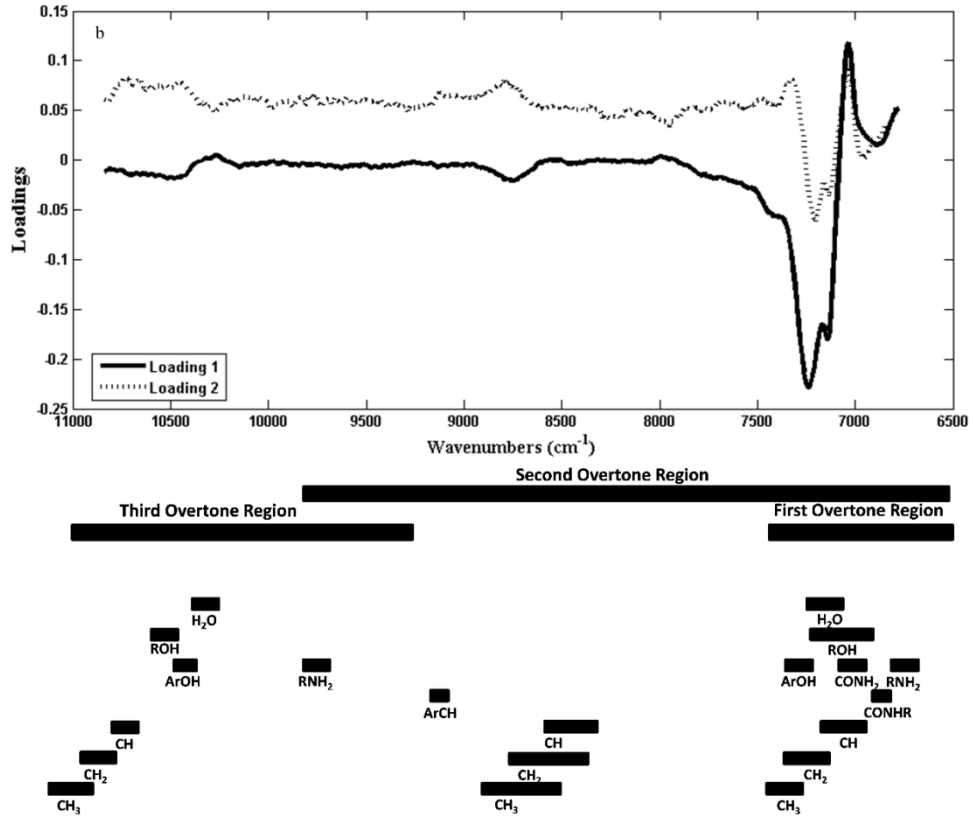


Figure 3.2.4b. Loadings plot obtained for the PLS-DA model discriminating between all soil types for portable instrument (X-block percentage of captured variance: 99.2% and Y-block percentage of captured variance: 40.9%).

The results obtained in this work reveal that there are no significant differences between using a portable or a benchtop equipment for discriminating soil types. Thus, NIRS can be used to give a fast, low-cost and accurate soil mapping in vineyards as well as any other agriculture cultivar and with this improve the efficiency in resource usage and define tailor-made strategies for the future.

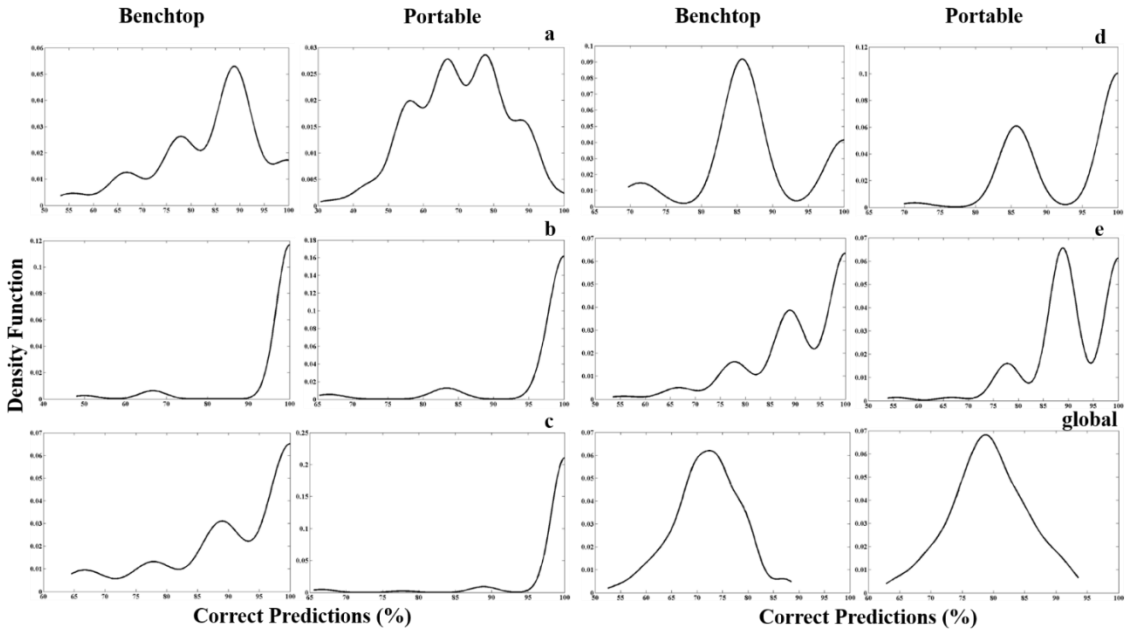


Figure 3.2.5. Distributions of correct soil predictions for all vineyard blocks individually (*a* to *e*) and the global model for benchtop and portable instruments (X-axis: Correct predictions; Y-axis: Density function).

3.2.4. Conclusions

Research in environmental monitoring, modelling and precision agriculture needs good quality and inexpensive soil data. Hence, development of more time- and cost-efficient methodologies for soil analysis is of vital importance. NIRS can be an essential part of these developments as a fast and accurate soil mapping technique of areas of particular and economical interest. It seems possible to apply NIRS as a tool for monitoring changes in soil composition as well as receiving a quick estimate of some soil chemical properties. Furthermore, NIRS could also be used in the field for swift soil mapping, leading to a more accurate and sustainable vineyard management, as well as any other agriculture cultivar, by improving efficiency in resource usage and defining tailor-made strategies for the future.

In this work, the ability of two NIRS instruments, a dispersive (portable) and a FT (benchtop) NIR spectrometers, to discriminate between specific vineyard soil types was compared with the objective of investigating the potential of NIRS as a rapid and low-cost technique to map vineyards soil both in the field and in the laboratory. The portable instrument was used for direct *in-situ* analysis of samples (no sample processing at all) and the benchtop was used after samples drying at the lab. A soil characterized vineyard

in the Dão Wine Region (Portugal) was employed in this study. The proposed spectral models were able to accurately differentiate distinct soil types with both instruments (around 90% of correct predictions considering soil samples within a vineyard block, and 74% of correct predictions considering samples collected on a 60 ha vineyard). Results demonstrated that the performance regarding soils discrimination of the portable spectrometer was comparable to the benchtop spectrometer. In some cases, the portable spectrometer presented higher correct soil predictions values. It was expected that the laboratory equipment, due to its higher resolution, spectral window and the fact that it was used under more controlled conditions would yield better results that the portable device. However, this work demonstrated that the use of portable NIR equipment for soil differentiation can be as reliable as benchtop equipment operating under well controlled conditions in the laboratory. This is of paramount importance for direct field analysis, providing a more time- and cost-efficient methodology for soil analysis.

3.2.5 References

1. Ben-Dor, E., D. Heller, and A. Chudnovsky, *A novel method of classifying soil profiles in the field using optical means.* Soil Science Society of America Journal, 2008. **72**(4): p. 1113-1123.
2. Soriano-Disla, J.M., et al., *The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties.* Applied Spectroscopy Reviews, 2014. **49**(2): p. 139-186.
3. Dos Santos, C.A.T., et al., *A review on the applications of portable near-infrared spectrometers in the agro-food industry.* Applied spectroscopy, 2013. **67**(11): p. 1215-1233.
4. Reeves, J., G. McCarty, and T. Mimmo, *The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils.* Environmental Pollution, 2002. **116**: p. S277-S284.
5. Rossel, R.A.V., et al., *On the soil information content of visible-near infrared reflectance spectra.* European Journal of Soil Science, 2011. **62**(3): p. 442-453.
6. Lang, R., et al., *Bioappearance and pharmacokinetics of bioactives upon coffee consumption.* Analytical and bioanalytical chemistry, 2013. **405**(26): p. 8487-8503.
7. Sudduth, K.A. and J.W. Hummel, *Soil organic-matter, cec, and moisture sensing with a portable nir spectrophotometer.* Transactions of the Asae, 1993. **36**(6): p. 1571-1582.
8. Acevedo, F., et al., *Spent coffee grounds as a renewable source of bioactive compounds.* Journal of Biobased Materials and Bioenergy, 2013. **7**(3): p. 420-428.
9. Baveye, P.C. and M. Laba, *Visible and near-infrared reflectance spectroscopy is of limited practical use to monitor soil contamination by heavy metals.* Journal of hazardous materials, 2015. **285**: p. 137-139.
10. Pessanha, M. and A. Graça. Increasing efficiency in production management of a wine enterprise: the project I.C.O.N.E. in *Proceedings of the 34th. World Congress of Vine and Wine*. 2011. Porto, Portugal.
11. FAO-WRBSR. Food and Agriculture Organisation of the United Nations. International Union of Soil Science (IUSS) Working Group World Reference Base. World Reference Base for Soil Resources. International soil classification system for naming soils and creating legends for soil maps. 1st ed. FAO, 2014. Rome.
12. Savitzky, A. and M.J. Golay, *Smoothing and differentiation of data by simplified least squares procedures.* Analytical chemistry, 1964. **36**(8): p. 1627-1639.

13. Joe Qin, S., *Statistical process monitoring: basics and beyond.* Journal of chemometrics, 2003. **17**(8-9): p. 480-502.
14. Barker, M. and W. Rayens, *Partial least squares for discrimination.* Journal of Chemometrics, 2003. **17**(3): p. 166-173.
15. Westerhuis, J.A., et al., *Assessment of PLS-DA cross validation.* Metabolomics, 2008. **4**(1): p. 81-89.
16. Preisner, O., et al., *Application of Fourier Transform Infrared Spectroscopy and Chemometrics for Differentiation of Salmonella enterica Serovar Enteritidis Phage Types.* Applied and Environmental Microbiology, 2010. **76**(11): p. 3538-3544.
17. Zhang, L. and M.J. McCarthy, *Measurement and evaluation of tomato maturity using magnetic resonance imaging.* Postharvest Biology and Technology, 2012. **67**: p. 37-43.
18. Bronick, C.J. and R. Lal, *Soil structure and management: a review.* Geoderma, 2005. **124**(1): p. 3-22.
19. Magalhães, L.M., et al., *Rapid assessment of bioactive phenolics and methylxanthines in spent coffee grounds by FT-NIR spectroscopy.* Talanta, 2016. **147**: p. 460-467.
20. Cozzolino, D., et al., *In situ measurement of soil chemical composition by near-infrared spectroscopy: A tool toward sustainable vineyard management.* Communications in soil science and plant analysis, 2013. **44**(10): p. 1610-1619.
21. Waiser, T.H., et al., *In situ characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy.* Soil Science Society of America Journal, 2007. **71**(2): p. 389-396.
22. Viscarra Rossel, R.A., et al., *In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy.* Geoderma, 2009. **150**(3–4): p. 253-266.
23. Spârchez, G., Dincă, L.C., Marin, G., Dincă, M., Enescu, R.E. 2014. The variation of eutric cambisols' chemical properties based on altitudinal and geomorphologic zoning. Environ. Eng. Manag. J.
24. Rossel, R.A.V. and T. Behrens, *Using data mining to model and interpret soil diffuse reflectance spectra.* Geoderma, 2010. **158**(1-2): p. 46-54.
25. Bishop, J.L., et al., *Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite-serpentines, chlorites and micas.* Clay Minerals, 2008. **43**(1): p. 35-54.
26. Bokobza, L., *Near infrared spectroscopy.* Journal of Near Infrared Spectroscopy, 1998. **6**(1): p. 3-17.

27. Kahle, M., M. Kleber, and R. Jahn, *Review of XRD-based quantitative analyses of clay minerals in soils: the suitability of mineral intensity factors.* Geoderma, 2002. **109**(3): p. 191-205.

The following work is ready for submission in a scientific journal:

Lopo, M., Soriano-Disla, J. M., Oliver, D. P., Páscoa, R. N., Janik, L., McLaughlin, M. J., & Lopes, J. A. Modelling of soil parameters with infrared spectroscopy in Australian vineyards: an instrument comparative study.

3.3. Modelling of soil parameters with infrared spectroscopy in Australian vineyards: an instrument comparative study

Abstract: Soil organic management and sustainable agriculture focus mainly on the growth of high-quality food and products, while preserving and trying to improve the fertility and quality of the soil. The characterization and evaluation (assessment, management, degradation, discrimination, etc.) of the environmental status of the soil is, therefore, crucial. This demand spurred the advent of easy-to-use, rapid and cost effective techniques to assess soil properties affected by changes in land use. Currently, soil analysis techniques are mainly based on laborious, expensive, time-consuming wet chemistry methods. Infrared (IR) spectroscopy is a rapid, non-destructive, cost-effective and reliable technique. A comparison was made of the performance of five different IR instruments – three Fourier-transform infrared (FT-IR) and two visible and near infrared (vis-NIR) – analysing the same soil samples from vineyards in the McLaren Vale wine region in South Australia. The objective was to assess the suitability of the instruments for the estimation of different soil properties, namely total nitrogen (TN), total organic carbon (TOC), pH, moisture and effective cation exchange capacity (eCEC). Partial least squares (PLS) models were first developed using the entire spectral range followed by models using specific wavelengths. Results revealed good models for three parameters (pH, moisture and eCEC), both in models using the whole wavelength and models constructed using just specific areas of the spectra, with R^2_p values between 0.51-0.68, 0.47-0.74 and 0.68-0.86 respectively. Models for TN using the whole spectra and specific areas exhibited poor results, probably due to the low concentrations in these soils (between 0.05-0.77%). Models to predict TOC using the whole spectra did not yield very satisfactory results, contrasting with models using specific regions of the spectra which had, for some instruments, very good results (R^2_p between 0.49-0.80). The performance of the spectrometers was not “consistent” (depending on the approach used), meaning that the best result for a given parameter using the whole spectra was not necessarily the same when specific regions of the spectra were used. This would suggest that, at least for the spectrometers used, there is no best equipment for all the different soil constituents analysed. However, the FT-IR instruments had an overall better performance than the vis-NIR spectrometers. Furthermore, models developed using specific regions of the spectra yielded better results than models constructed with the whole spectral range.

3.3.1. Introduction

Soil organic management is critical for life on earth, even more so now with the ever growing population of the planet. Sustainable agriculture today focuses on the growth of high-quality food and products, while preserving and trying to improve the fertility and quality of the soil. In light of this, the characterization and evaluation of the environmental status of the soil, discriminating between soil types, assessment, management, degradation, etc., is essential. Soil is a highly complex matrix consisting of organic and inorganic mineral matter, water, and gases [1]. Besides its key functions of food and energy production, soil also plays an important role in water regulation and carbon sequestration. Soils can be very heterogeneous: be it on a macroscopic scale in the landscape, both horizontally and vertically, occurring in distinguishable layers [2], but also on a microscopic scale where differences may occur over very short distances (e.g. moving from discrete mineral grain to soil microorganisms). Thus, no two soils are exactly alike [3]. Given the importance and availability of soils, there is a dire need for regular monitoring to assess soil status and detect changes that may occur so that suitable management can be implemented. It has been advocated that soil surveys should be performed at regional levels for agriculture and at a national level for the inventory of soil resources [4]. The information gathered could help in the improvement of future objectives such as proper management planning and sustainable land use. In the context of productive and sustainable agriculture there is a need to evaluate several soil properties such as pH, organic matter, nutrient and pollutant concentrations, among others. The development of the concept of Precision Agriculture, highlighting the importance of resource use efficiency [5], has also spurred the need for the development of fast, accurate, easy-to-use and cost effective techniques for soil analysis. Precision Agriculture reinforces the notion of supplying soils with their exact requirements at a high spatial resolution.

Consequently, there is a heavy demand for easy-to-use, rapid and cost effective techniques to assess soil properties affected by changes in land use. The advent of rapid techniques that provide complete information in a fast way is therefore of paramount importance. Current strategies for analysing soils and changes in properties over time are based on wet chemical methods involving liquid extractions and analysis of solutions by laboratory-based instruments, which are often laborious, expensive and time-consuming.

Over the past decades, various agricultural sensors have been used to determine soil properties [6]. Among those, infrared (IR) spectroscopy, based on mid infrared (MIR), near infrared (NIR), and visible and near infrared (vis-NIR), can constitute a valuable

solution when used in combination with chemometric tools. These technologies are rapid, reliable and non-destructive to the sample, relatively inexpensive and less-laborious when compared to wet chemistry methods. In addition, they use simple sample pre-treatment, and do not require the use of chemicals potentially harmful for the environment [7]. The use of IR spectroscopy as a method for discriminating soil types as well as determining different soil constituents is rapidly increasing [8]. The application of IR spectroscopy such as NIR in soil characterisation has been demonstrated, mainly for the determination of organic matter and other parameters such as pH, nitrogen (N), potassium (K) or phosphorous (P) [9], which are very important for crop maintenance and soil chemical fertility, as well as the identification of soil types based on soil constituents (organic compounds, carbonates, clay minerals, etc.) [10]. In soil science, MIR is still less used than NIR, probably because portable MIR instruments, until recently, were less readily available than equivalent NIR instruments. However, the use of MIR is gaining popularity, particularly due to the richness of spectral information gained from soils in that spectral range in comparison to the NIR [11]. There are not many studies that have evaluated soil properties in vineyards using IR spectroscopy. Cozzolino and co-workers [12] used a portable NIR instrument to predict several properties in soils from different wine regions. The authors obtained good results using partial least squares (PLS) with coefficients of determination (R^2) at the calibration level of: 0.74 for total nitrogen (TN); 0.81 for organic carbon (OC); 0.84 for electrical conductivity (EC) and 0.83 for pH. More recently, Lopo et al. [13] compared the performance of a benchtop NIR spectroscopy instrument with a portable one for soil classification. These authors wanted to assess if NIR spectroscopy could be used as a swift and accurate tool to map vineyard soils. Páscoa and colleagues [14], used a similar approach but with grapevine leaves. Using a portable NIR spectroscopy instrument, they were able to characterize soils based on direct *in-situ* measurements of vine leaves, enabling the estimation of soil variability in a fast, simple and precise method.

The objective of this study was to assess the performance of different Fourier-transform infrared (FT-IR) and vis-NIR instruments to determine a range soil properties, namely TN, total organic carbon (TOC), pH, moisture and effective cation exchange capacity (eCEC). Soil samples were collected from different vineyards in the McLaren Vale wine region (South Australia) and characterised by wet-chemistry laboratory analysis, followed by scanning with different FT-IR and vis-NIR instruments. To the best of our knowledge, it is the first time that the performance of IR instruments has been assessed using a broad number of samples from vineyards soils. The goal was to assess which instrument yielded better results predicting the various soil constituents and

consequently, determine which instrument proves a better investment from a practical point of view. Furthermore, a comparison of the instruments performance using the whole spectra and selected spectral regions was also undertaken.

3.3.2. Materials and Methods

3.3.2.1. Experimental site

A total of forty vineyard locations were selected across McLaren Vale (South Australia - 35° 14' 00.00" S 138° 31' 60.00" E) for soil sampling. McLaren Vale is one of the most important producing wine regions of Australia, repeatedly producing some of the finest wines in the world. The vineyards were chosen in conjunction with the McLaren Vale Grape Wine & Tourism Association to cover the range of soil types present in the region (Figure 3.3.1), as well as the different local viticultural management practices (conventional, organic, biodynamic). Out of the forty sites, some were under conventional management (including low input conventional), other were under organic management (both certified and uncertified), and a few were biodynamic.

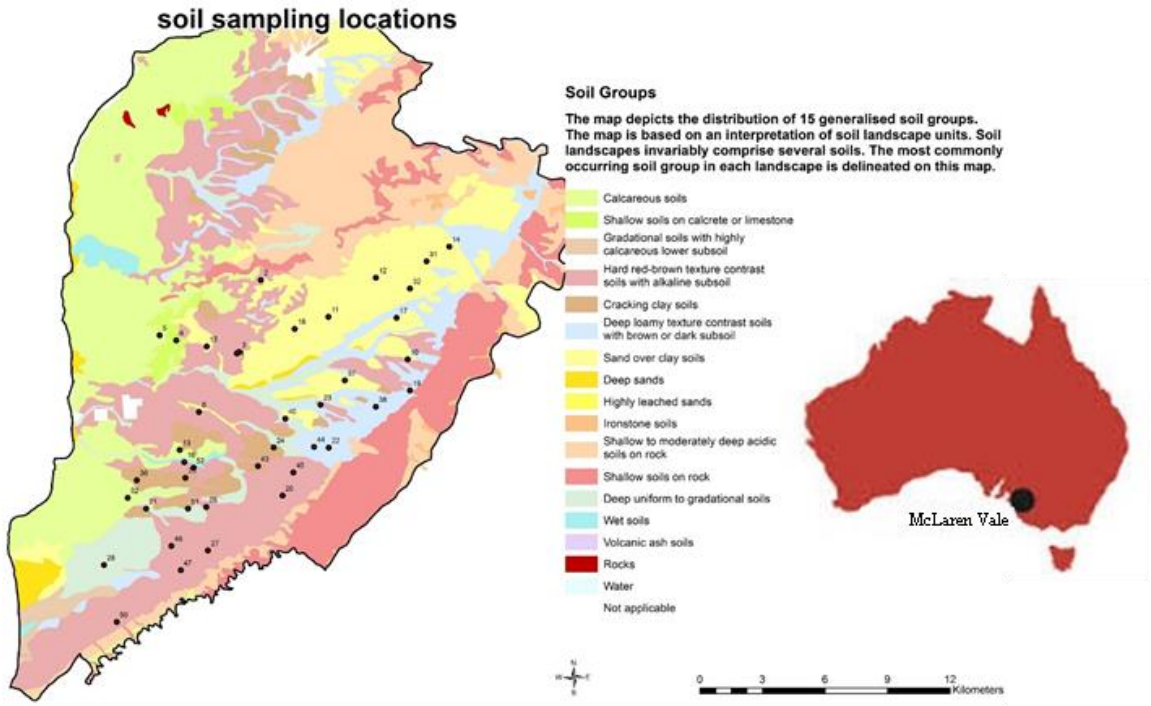


Figure 3.3.1. 1:50.000 soil map for McLaren Vale obtained from Primary Industries and Regions. South Australia (PIRSA).

3.3.2.2. Sample collection and processing

Samples were collected a few weeks after harvest in two subsequent years. For each sampled vineyard, soils were georeferenced using a differentially corrected global positioning system (dGPS; accurate to +/- 50 cm). The sampling location within the vineyard was selected as being representative of the block. At each sampling location, an “undervine” and a “midrow” sampling point were selected. The undervine sample was selected 20 cm towards the midrow from the dripper closest to the trunk; a midrow sample was selected in the centre of the midrow but in line with the undervine site. A hand-auger (30 cm diameter) was used to collect both the undervine and midrow samples, clods were sampled towards the centre of the hole and stored in rigid plastic jars for transportation to the laboratory for physical characterisation. Approximately 1 kg of soil was taken from the surface (0-10 cm) and sub-surface (35-45 cm) for both undervine and midrow locations (Table 3.3.1) making a total of 160 samples. Samples were immediately transported to the laboratory and stored at 4°C before chemical characterisation. Prior to spectral analysis, the samples were air-dried, sieved to <2 mm and homogenized, oven-dried at 40 °C for 12 h, and cooled in a desiccator. Further details of the soil collection and the vineyard sites are given in [15].

Table 3.3.1. Sampling locations in the different vineyards performed in two subsequent years at specific depths.

Location	Sampling Points at each location	Vineyard	Years	Depth
Undervine, Midrow	2	Chapel Hill	2	0-10; 35-45cm
Undervine, Midrow	4	Oliver's Taranga	2	0-10; 35-45cm
Undervine, Midrow	4	Paxton	2	0-10; 35-45cm
Undervine, Midrow	4	Gemtree	2	0-10; 35-45cm
Undervine, Midrow	4	Rosemount	2	0-10; 35-45cm
Undervine, Midrow	4	d’Arenberg	2	0-10; 35-45cm
Undervine, Midrow	2	Inkwell	2	0-10; 35-45cm
Undervine, Midrow	4	Cameron	2	0-10; 35-45cm

Table 3.3.1. **(cont.).** Sampling locations in the different vineyards performed in two subsequent years at specific depths.

Undervine, Midrow	4	Coriole	2	0-10; 35-45cm
Undervine, Midrow	2	Noons	2	0-10; 35-45cm
Undervine, Midrow	2	Leask Vineyards	2	0-10; 35-45cm
Undervine, Midrow	1	Bottin	1	0-10; 35-45cm
Undervine, Midrow	1	Leconfield	1	0-10; 35-45cm
Undervine, Midrow	2	Yangarra Estate	2	0-10; 35-45cm

3.3.2.3. Soil properties determination

Total carbon and nitrogen were determined by high temperature combustion in an atmosphere of oxygen using a Leco TruMAC. Carbon was converted to CO_2 and determined by infrared detection. Nitrogen was determined as N_2 by thermal conductivity detection [16]. Inorganic carbon was determined by reacting the sample with acid in a sealed container and measuring CO_2 released manometrically. Sufficient finely ground sample to contain no more than 0.8g $CaCO_3$ equivalent was weighed into a 250mL glass bottle, a tube containing 8mL 3*M* HCl and 3% ferrous chloride added and the bottle sealed. The contents were mixed intermittently during a 1 hour period and the pressure in the bottle measured by piercing the septum with a needle attached to a pressure transducer [17]. Total organic carbon was determined by difference between total carbon and inorganic carbon.

Soil pH and electrical conductivity (EC) were determined using a 1:5 soil/water extract. Air dried material (5 g) was shaken with 25 ml water for one hour and left to settle for 20 minutes. Electrical conductivity was determined and a subsample for chloride taken. pH in water was then determined before calcium chloride added, sample re-agitated, re-settled before reading for pH in calcium chloride. pH was measured using a Metrohm 815 Robotic Processor and EC was determined using a Radiometer CD230 [18-19].

Exchangeable cations and cation exchange capacity (CEC) were determined using NH_4Cl solution at either pH 7.0 or pH 8.5 dependent on soil pH. Non calcareous soils use extraction solution pH 7.0 and alkaline soils use extraction solution pH 8.5. Samples were pre-treated for soluble salts prior to extraction. Exchangeable cations, Ca, Mg, Na and K, were analysed by Flame Atomic Absorption Spectrometry. Cation exchange capacity

ammonium and chloride were analysed using Flow Injection Analyser [21]. Effective cation exchange capacity was determined by converting the exchangeable cations (Ca^{2+}, K^{+}, Mg^{2+} and Na^{+}) from mg/kg to cmol(+)/kg and adding together.

Table 3.3.2 shows the mean, standard deviation (SD), coefficient of variation (CV), and range for the chemical properties analysed.

Table 3.3.2. Descriptive statistics for chemical properties of the different soil parameters used in modelling.

Soil Property	Min	Max	Mean	SD	CV
Total nitrogen (%)	0.05	0.77	0.25	0.14	56.74
Total organic carbon (%)	0.46	8.01	2.58	1.33	51.52
pH	4.60	8.10	7.38	0.87	13.09
Moisture (%)	0.16	36.30	9.96	7.27	72.99
eCEC (meq/100g)	0.42	41.79	14.21	10.10	71.06

Notes. CV = (SD/mean) x 100; min, minimum; max, maximum; and SD, standard deviation.

3.3.2.4. Spectral acquisition

Spectra acquisition through diffuse reflectance mode was performed using five different spectrometers.

A) PerkinElmer *Spectrum-One*™ Fourier Transform Infrared spectrometer (PerkinElmer Inc., USA). The spectrometer was equipped with an extended range KBr beam-splitter, a high intensity ceramic source and a deuterium triglycine-sulfate (DTGS) detector. Spectra were scanned for 60 s over the frequency range of 7800 to 450 cm^{-1} at a resolution of 8 cm^{-1}, with the final spectrum being the average of approximately 60 scans. Background reference scans were carried out every hour using silicon carbide (SiC) discs (PerkinElmer. USA) with a reflectivity of 100%. Spectra were expressed in pseudo absorbance (*A*) units (where A = Log $Reflectance^{-1}$).

B) Alpha FT-IR spectrometer (Bruker, Germany). This portable instrument was equipped with a front reflectance accessory. Samples were deposited onto a Petri dish (10 mm depth and 63 mm diameter), on top of a stainless steel laboratory jack, and raised up to level with the front reflectance accessory, to simulate the scanning from the top as if the samples were rolled on a conveyor belt. Spectra were acquired for 15 s over the

frequency range of 7500-375 cm^{-1} at a resolution of 8 cm^{-1}. A gold reference background (Bruker, Germany) was used.

C) FlexScan FT-IR model 4200 (Agilent, USA). This instrument is a handheld spectrometer set up for DRIFT mode and equipped with a Michelson interferometer, zinc selenide beam splitter and thermoelectrically cooled DTGS detector. Spectra were acquired over the frequency range of 6000–650 cm^{-1}. Being a handheld instrument, portability is an important issue. In this case, the optical and electronic/battery components are split into two, where the electronics/battery is worn on the belt and linked by a communications cable to the optical component which is naturally held by hand. The instrument was positioned with the optical DRIFT component facing down and samples, placed in stainless steel cups (9 mm diameter, 3 mm depth), were lifted by compressed air against the optical aperture allowing a high reproducibility in sample presentation. A course-grained SiC reference disk was used as background and scanned every hour.

D) Spectral Evolution SM-3500 vis-NIR spectrometer (Spectral Evolution, USA). This handheld vis-NIR spectrometer has a crossed Czerny-Turner configuration with ruled gratings used as the dispersive element. The reflected energy enters the spectrometer and is collimated before being reflected off the gratings and refocused onto three detectors (range 28571-4000 cm^{-1}): a 512-element Si array (vis-NIR; up to 10000 cm^{-1}) and two Peltier cooled indium gallium arsenide (InGaAs) arrays of 256 elements each extending detection to 4000 cm^{-1}. Soil samples were prepared on a Petri dish of 10 mm depth and 63 mm diameter. The surface of the samples was tamped using a glass surface, and then brought to the contact probe by using a stainless steel laboratory jack, following the methodology proposed by [22]. Spectra were obtained with a high energy contact probe with external illumination (Spectral Evolution, USA) coupled to the instrument with a fixed fibre optic cable. The resolution was 3.5 nm, 10 nm and 7 nm for the 28571-10000 cm^{-1}, 10000-5263 cm^{-1} and 4761-4000 cm^{-1} spectral ranges, respectively. A total of 30 spectra were averaged, with a background white reference (Spectralon™, Spectral Evolution, USA) scanned every 30 minutes.

E) NIRScan Nano vis-NIR spectrometer (Texas Instruments, USA). This instrument is a miniature handheld spectrometer (weight ~ 100 g). The diffusively reflected light is congregated and focused through the input slit (25 µm wide by 1.69 mm tall). The light then strikes a reflective grating which disperses, in combination with a focusing lens, the light into the different wavelengths. The focusing lenses form an image

of the slit at the Digital Light Proccessing (DLP) micro-mirror device (DMD; (0.2-inch WVGA, 854 × 480 orthogonal pixel, NIR optimized). The energy reflected by the DMD is directed through the collection optics to the single pixel InGaAs detector (11111-5882 cm^{-1}). The spectra were acquired by placing the samples in close contact to a sapphire window which was illuminated with two integrated and angled tungsten infrared lamps. The background was scanned with the white reference (Spectralon™, Spectral Evolution, USA) every 30 minutes.

Each sample was measured in triplicate yielding a total of 480 spectra for each instrument. The average spectrum was considered for subsequent analysis.

3.3.2.5. Data analysis

All spectra were preprocessed with Savitzky–Golay filter (15-point filter size, second-order polynomial and first-order derivative) [23] and standard normal variate. Mean-centring was applied before principal component analysis (PCA) and partial least squares (PLS) modelling [24]. To extract common patterns from the infrared spectra and to assist in outlier detection, PCA was performed, whereas PLS was used to develop calibration models for soil parameters determination purposes. Chemometric analysis of the spectra was first performed using the whole spectral range of each instrument, minus noisy areas at each end of the spectra, followed by combinations of specific regions of the spectra to fine tune the models and extract the best performance from each instrument. The PLS models were developed considering the PLS-1 algorithm described in [25]. All PLS models were constructed considering approximately 70% of the available samples (calibration set) and then tested using the remaining samples (test set). Thus, the 160 spectra available for analysis were separated randomly into a calibration (110 spectra) and testing (50 spectra) sets. The optimal number of latent variables (LVs) was estimated, for each model, by leave-one-out cross-validation using only the calibration set and considering the root mean square error of cross validation (RMSECV) as the minimization criterion [24]. The PLS models performance was primarily evaluated with the RMSECV, and the root mean square error of prediction (RMSEP) was used to evaluate the performance of the selected model when the testing set was projected. Spectra were divided into different regions for each equipment (different detector range in each equipment), according to the major chemical/physical properties captured by IR spectra (Table 3). The best spectral regions were estimated, by testing for each spectrometer all possible combinations, and selected according to the lowest root mean square error (RMSE). Additional statistics such as the experimental versus predicted response and

coefficients of determination (R^2), as well as the range error ratio (RER) (Eq. 1) were also calculated to evaluate the predictive ability of the models.

RER=Soil parameter tested range/RMSEP (3.1)

Table 3.3.3. Spectra regions for each spectrometer according to the major chemical/physical properties captured by the spectra.

#	Bruker Alpha FT-IR	Perkin Elmer Frontier FT-IR	Agilent FlexScan FT-IR	Spectral Evolution SM-3500 Vis-NIR	Texas Instruments DLP NIRScan Nano Vis-NIR
			Regions		
1	1340-411 cm^{-1}	1340-422 cm^{-1}	1230-760 cm^{-1}	4716-4000 cm^{-1}	6309-5878 cm^{-1}
2	2200-1340 cm^{-1}	2200-1340 cm^{-1}	2055-1230 cm^{-1}	5464-4716 cm^{-1}	6849-6309 cm^{-1}
3	3100-2200 cm^{-1}	3100-2200 cm^{-1}	2640-2055 cm^{-1}	7518-5464 cm^{-1}	7246-6849 cm^{-1}
4	4000-3100 cm^{-1}	4400-3100 cm^{-1}	3800-2640 cm^{-1}	17241-7518 cm^{-1}	8620-7246 cm^{-1}
5	5350-4000 cm^{-1}	5350-4400 cm^{-1}	5200-3800 cm^{-1}	27855-17241 cm^{-1}	10672-8620 cm^{-1}
6	7458-5350 cm^{-1}	7782-5350 cm^{-1}			

All chemometric methods and spectra processing were performed using Matlab version 7.9 (MathWorks. Natick. MA) and PLS Toolbox version 5.5.1 (Eigenvector Research Inc., Wenatchee. WA).

3.3.3. Results and Discussion

3.3.3.1. Preliminary spectra analysis

The IR raw spectra obtained using the five instruments were markedly different visually. An example from a midrow, 35-45 cm deep sample from the Cameron vineyard is shown

in Figure 3.3.2 to illustrate those discrepancies. Scale-wise, peaks seemed to be more prominent in the MIR region (4000-400 cm^{-1}) (Figures 3.3.2a-c) which is known to provide an overall chemical profile of the soil and harbour fundamental vibrations of most soil materials [9], such as silicates, including clay minerals, carbonates, organic matter, and also other types of minerals (e.g. sulfates). The IR spectra of soils normally produced in the MIR region arise from the fundamental vibrations of the components present [3]. The most noticeable peaks for the sample depicted in Figure 3.3.2 (a, b and c) are around 800-700 cm^{-1} and 1400-1300 cm^{-1} which are known to correspond to quartz and carbonates, and the broad absorption in the 3695-3200 cm^{-1} region which is known to correspond to clay minerals [26]. Some strong depressions (or "valleys") can be seen around 1150-1000 cm^{-1} and 2000-1800cm^{-1} which normally indicate the presence of quartz [27, 28]. Peaks attributable to SOM can also be seen at 2920 cm^{-1} and 2850 cm^{-1} and around 1600 cm^{-1} and 1400 cm^{-1} [29, 30].

The NIR region extends between 12500–4000 cm^{-1} (1250-2500 nm) and consists of broad absorption bands, namely, overtones and combinations of the fundamental OH, NH and CH absorptions found in the MIR range. These NIR overtones and combination bands can overlap, making the spectra more difficult to characterize than in the MIR [30-32]. Even though NIR spectra do not appear to have the resolution and intensity of the MIR spectra, NIR penetrates deeper into the samples due to its more energetic radiation and reduced specular reflectance from sample particles surfaces and moisture films on the particle surfaces, allowing a swift scanning with little, or no previous sample preparation [3]; this is obviously a very important advantage for field work where highly variable particle sizes and moisture contents can occur. In general, NIR spectra contain absorbance bands mainly due to chemical bonds of C–H (alkanes, fats, oil), O–H (alcohol, water) and N–H (protein). Thus, the SOM and clay minerals are reflected accurately in the NIR spectra, but provide less information about the mineral content of the samples since they are not responsive to the Si-O bond. Other chemical bonds may exhibit overtone bands in the NIR region, but are generally weaker [4]. For the sample shown in Figure 3.3.2 (a-c), it is possible to observe in the NIR region (although not as noticeable as in the mid-IR) a few peaks in the 4500 cm^{-1} region, related to clay minerals like illite and smectite [33]; the slightly broader absorption around 5200 cm^{-1} is related to O–H stretch of water and vibrations of H–O–H [34]; the small occurrence near 7100 cm^{-1} is related to the first over-tone of the O–H stretch vibration in metal–O–H [35]. Regarding Figures 3.3.2d-e which encompass the vis-NIR regions of the electromagnetic spectrum, peaks can be seen in the 4500 and 5200 cm^{-1} regions (Figure 3.3.2d), as well as a slight indentation in the 7100 cm^{-1} region. It is also possible to note a marked undulation at the end of the

visible spectrum (around 25000 cm^{-1}). Reports in the literature indicate that absorption in the visible region is mostly related to minerals that contain iron such as: maghemite, haematite, lepidocrocite or goethite [36]. Soil organic matter is also known to have broad absorption bands in the visible region that are dominated by chromophores [4]. In Figure 2e, only a broad peak is visible in the 7100-7000 cm^{-1} region.

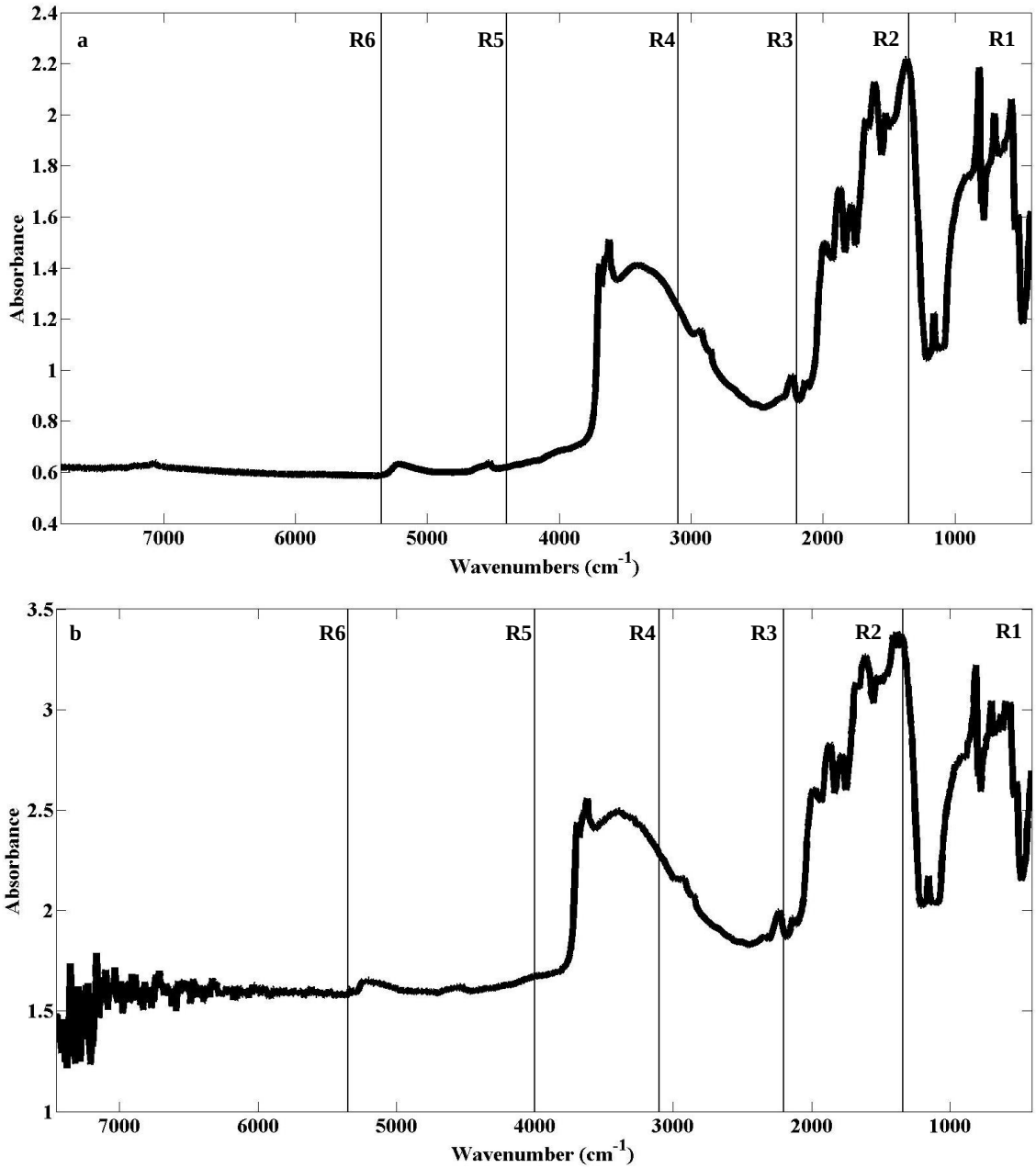


Figure 3.3.2. Raw spectra of a loamy soil sample for all five spectrometers used with respective spectral regions according to the major chemical/physical properties captured by the spectra. **a**-Perkin Elmer Frontier FT-IR; **b**-Bruker Alpha FT-IR; **c**-Agilent FlexScan FT-IR; **d**-Spectral Evolution SM-3500 Vis-NIR; **e**-Texas Instruments DLP NIRScan Nano Vis-NIR. **R1**-Spectral region one; **R2**-Spectral region two; **R3**-Spectral region three; **R4**-Spectral region four; **R5**-Spectral region five; **R6**-Spectal region 6.

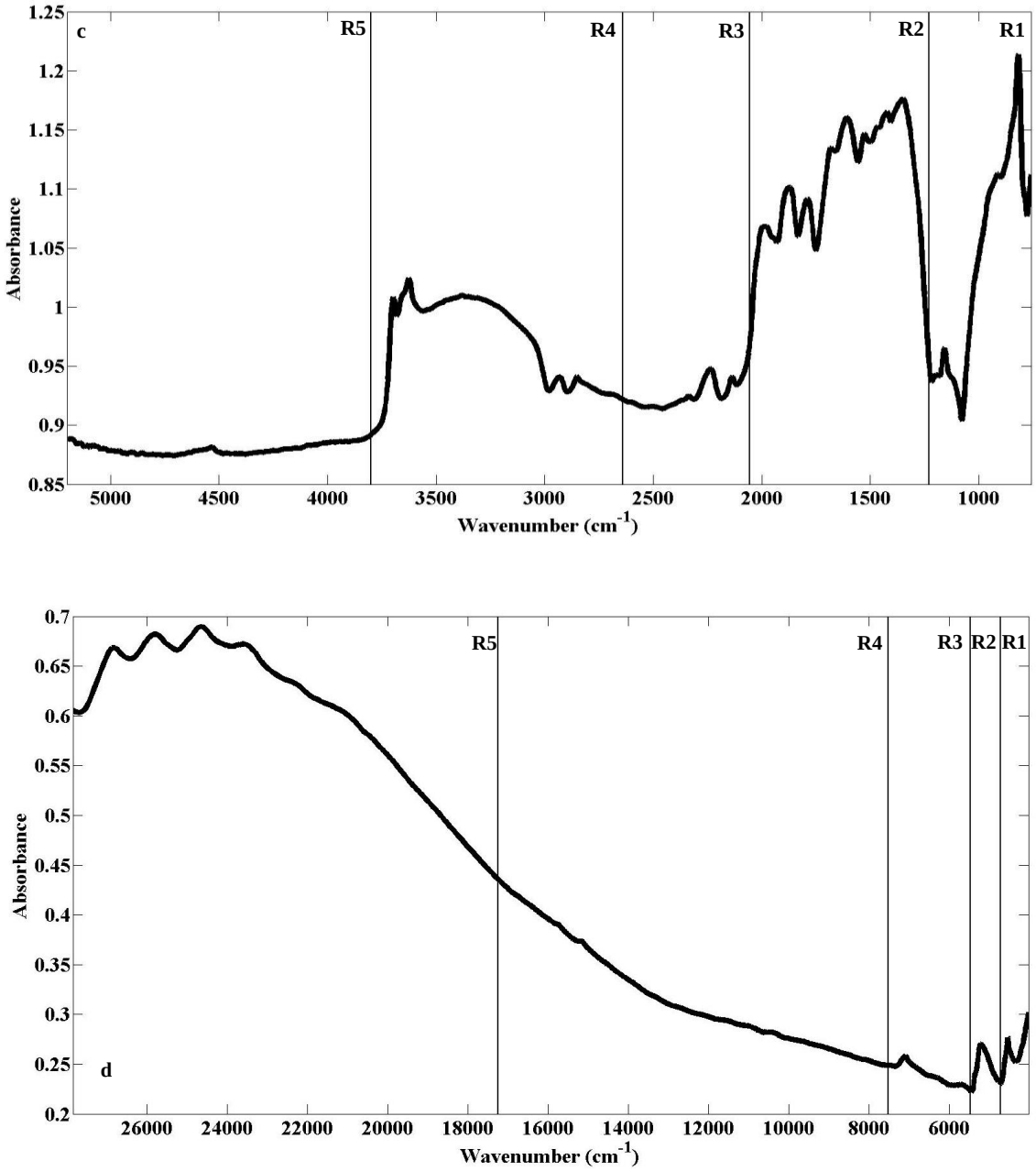


Figure 3.3.2. **(cont.).** Raw spectra of a loamy soil sample for all five spectrometers used with respective spectral regions according to the major chemical/physical properties captured by the spectra. **a**-Perkin Elmer Frontier FT-IR; **b**-Bruker Alpha FT-IR; **c**-Agilent FlexScan FT-IR; **d**-Spectral Evolution SM-3500 Vis-NIR; **e**-Texas Instruments DLP NIRScan Nano Vis-NIR. **R1**-Spectral region one; **R2**-Spectral region two; **R3**-Spectral region three; **R4**-Spectral region four; **R5**-Spectral region five; **R6**-Spectal region 6.

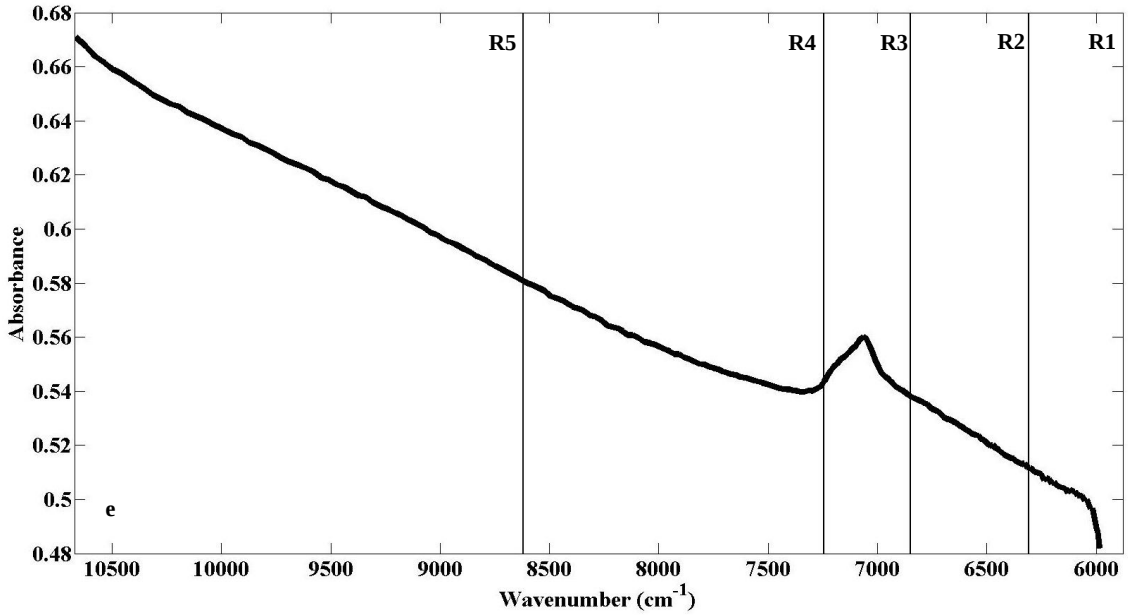


Fig. 3.3.2. **(cont.).** Raw spectra of a loamy soil sample for all five spectrometers used with respective spectral regions according to the major chemical/physical properties captured by the spectra. **a**-Perkin Elmer Frontier FT-IR; **b**-Bruker Alpha FT-IR; **c**-Agilent FlexScan FT-IR; **d**-Spectral Evolution SM-3500 Vis-NIR; **e**-Texas Instruments DLP NIRScan Nano Vis-NIR. **R1**-Spectral region one; **R2**-Spectral region two; **R3**-Spectral region three; **R4**-Spectral region four; **R5**-Spectral region five; **R6**-Spectal region 6.

3.3.3.2. Exploratory data analysis

Principal components analysis models of the spectral data were performed to identify outliers and also attempt to detect common patterns within the samples analysed. The entire wavelength range was considered, excluding noisy areas at both extremes of the spectra. This preliminary analysis did not reveal the presence of clear outliers or cluster formation: no clear groupings could be observed regarding the type of soil, sample location and sample depth (Figure 3.3.3).

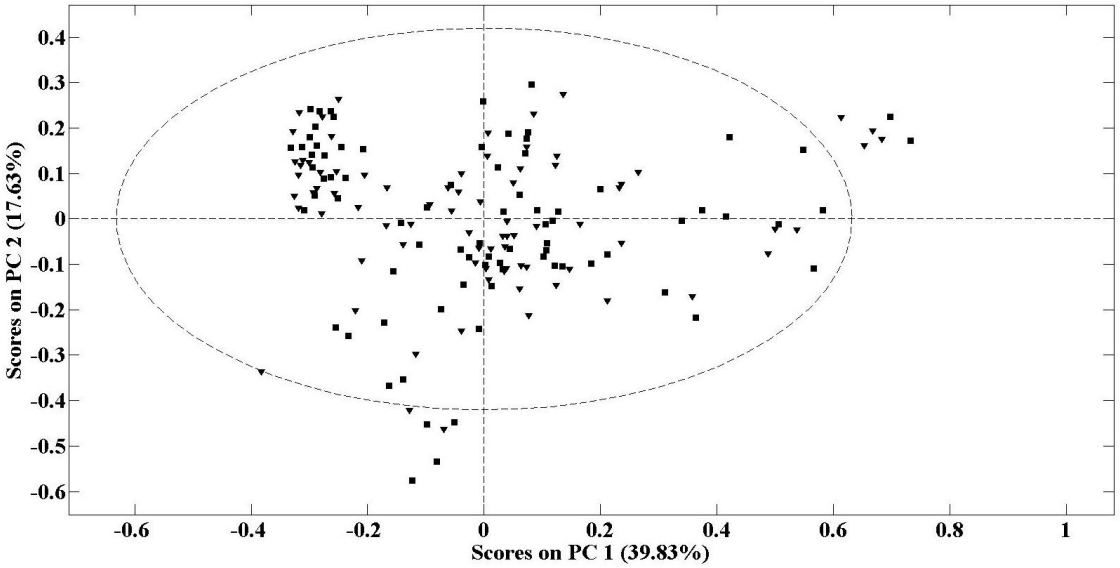


Figure 3.3.3. Score plot obtained from PCA models built on IR data collected with Agilent FlexScan FT-IR from soil samples collected at different locations (▼ Undervine; ■ Midrow).

3.3.3.3. Calibration

Calibration models using the whole spectra were initially developed for each instrument in order to obtain an unbiased comparison of their performance for each soil parameter. The resulting PLS models were then tested with the independent test set. A graphical representation of the experimental versus predicted parameter (eCEC in this case) values (cross-validation and prediction) is presented as an example in Figure 3.3.4. The coefficient of determination (R^2_p) and lowest error for all the instruments and parameters are shown in Table 3.3.4. From the obtained models, it is possible to conclude that for this specific set of data, the estimation of TOC and TN did not render very good models with any of the tested instruments: R^2_p values were quite low (0.28-0.67 and 0.32-0.49, respectively), with one notable exception for the benchtop PerkinElmer Frontier FT-IR which performed rather well in the estimation of TOC (R^2_p=0.67). R^2_p values for the remaining parameters revealed average results for pH (0.51-0.62), satisfactory results for moisture (0.47-0.71) and good results for eCEC (0.68-0.83). However, there does not seem to be a specific trend regarding the performance of each spectrometer. In other words, no single instrument exhibited the best result for all parameters. As expected, the PerkinElmer Frontier FT-IR had the highest R^2_p for most of the parameters (all except pH). Regarding the portable instruments, the Bruker Alpha FT-IR had the best result for TN, TOC and moisture; the Agilent FlexScan FT-IR for eCEC and the Spectral Evolution SM-3500 Vis-NIR for pH. Somewhat expected was the inferior performance of the Texas

Instruments DLP NIRScan Nano Vis-NIR due to its narrower spectral range, its main advantage being, obviously, the portability. Even though this instrument does not have the resolution of the others used in this study, it is small enough to be carried on a pocket or small pouch, making *in-situ* and on the fly measurements extremely easy.

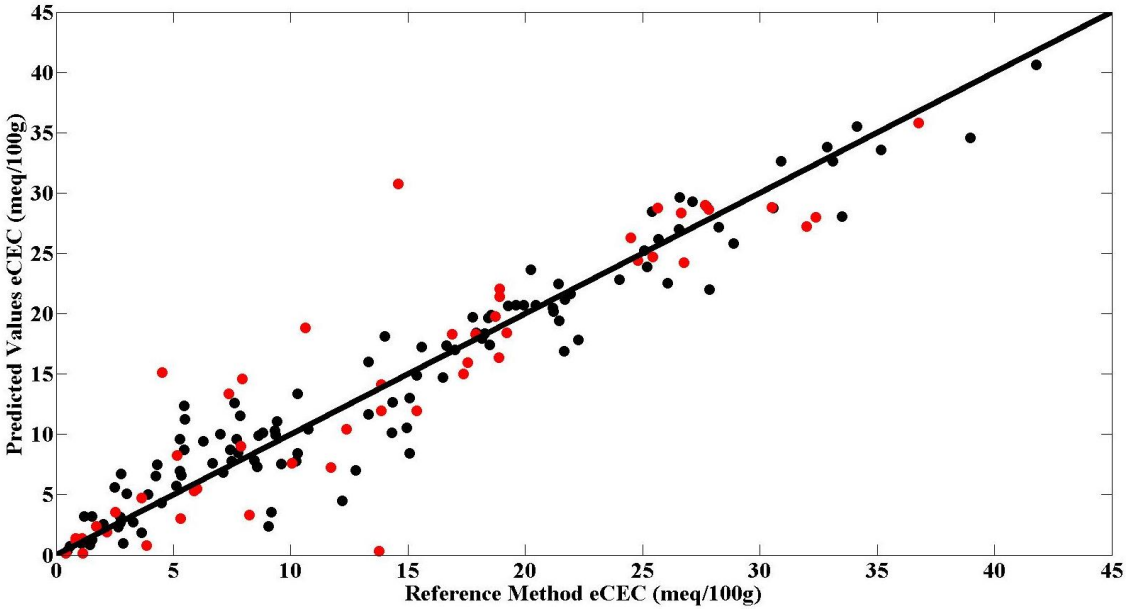


Figure 3.3.4. Comparison between Effective CEC determined with the reference method and the IR based method for the calibration samples (cross-validation predictions) and the independent test samples using the whole spectra (• - cross-validation; • - prediction) for the Agilent FlexScan FT-IR.

Table 3.3.4. PLS results for all the parameters and instruments analysed with respective number of LVs. R^2_C- coefficient of determination of calibration; R^2_{CV}- coefficient of determination of cross-validation; R^2_P- coefficient of determination of prediction; LV-latent variables.

		Bruker Alpha FT-IR		Perkin Elmer Frontier FT-IR		Agilent FlexScan FT-IR		Spectral Evolution SM-3500 Vis-NIR		Texas Instruments DLP NIRScan Nano Vis-NIR	
		Whole spectra	Specific regions	Whole spectra	Specific regions	Whole spectra	Specific regions	Whole spectra	Specific regions	Whole spectra	Specific regions
TN	LV	1	2	1	2	2	6	6	4	7	6
	RMSEC	0.09	0.09	0.11	0.11	0.10	0.07	0.08	0.1	0.08	0.08
	RMSECV	0.10	0.1	0.12	0.12	0.12	0.12	0.12	0.12	0.10	0.1
	REMSEP	0.13	0.12	0.09	0.09	0.10	0.09	0.14	0.14	0.14	0.13
	R^2_C	0.45	0.46	0.40	0.4	0.52	0.77	0.56	0.36	0.61	0.57
	R^2_{CV}	0.34	0.33	0.32	0.3	0.37	0.41	0.17	0.11	0.40	0.38
	R^2_P	0.45	0.53	0.49	0.52	0.41	0.54	0.34	0.45	0.32	0.39
	RER	5.45	5.78	4.84	4.98	4.46	4.94	5.06	5.03	4.97	5.25
TOC	LV	2	4	4	4	2	4	6	5	5	7
	RMSEC	0.75	0.75	0.83	0.87	0.93	0.8	0.08	0.09	0.98	0.91
	RMSECV	0.95	0.91	1.06	0.99	1.04	0.99	0.12	0.12	1.14	1.19
	REMSEP	1.16	1.01	0.58	0.46	0.86	0.52	0.13	0.13	1.36	1.28
	R^2_C	0.62	0.62	0.67	0.64	0.59	0.69	0.54	0.48	0.35	0.44
	R^2_{CV}	0.39	0.46	0.47	0.53	0.49	0.54	0.16	0.14	0.17	0.13
	R^2_P	0.48	0.64	0.67	0.8	0.39	0.76	0.39	0.49	0.28	0.53
	RER	5.31	6.12	5.86	7.49	3.98	6.57	5.24	5.50	4.54	4.81
pH	LV	3	3	11	10	11	10	5	4	11	4
	RMSEC	0.42	0.46	0.28	0.34	0.28	0.29	0.43	0.44	0.44	0.51
	RMSECV	0.58	0.6	0.48	0.48	0.50	0.46	0.51	0.5	0.55	0.54
	REMSEP	0.56	0.55	0.59	0.52	0.64	0.57	0.58	0.58	0.65	0.57
	R^2_C	0.79	0.75	0.91	0.86	0.90	0.9	0.75	0.74	0.77	0.69
	R^2_{CV}	0.59	0.57	0.73	0.73	0.71	0.75	0.66	0.67	0.65	0.65
	R^2_P	0.59	0.61	0.60	0.68	0.53	0.6	0.62	0.62	0.51	0.6
	RER	5.33	5.44	5.08	5.8	4.68	5.3	6.56	6.58	4.64	5.31
Moisture	LV	3	4	6	6	4	6	6	8	10	4
	RMSEC	3.95	3.78	3.61	3.64	4.04	3.49	3.45	3.56	4.16	4.97
	RMSECV	5.14	4.49	4.80	4.71	4.78	4.71	4.17	4.39	5.03	5.44
	REMSEP	4.70	4.34	4.42	4.31	4.27	4.24	4.55	4.03	5.39	5.09
	R^2_C	0.68	0.7	0.73	0.72	0.69	0.75	0.80	0.79	0.67	0.49
	R^2_{CV}	0.46	0.59	0.53	0.55	0.57	0.54	0.72	0.69	0.53	0.39
	R^2_P	0.67	0.73	0.71	0.73	0.65	0.74	0.47	0.6	0.50	0.6
	RER	7.00	7.57	7.44	7.63	8.44	7.76	5.64	6.37	6.69	6.46
eCEC	LV	4	5	11	10	7	10	5	5	5	5
	RMSEC	2.77	2.67	1.67	2.6	2.37	2.16	3.93	4.43	4.97	5.01
	RMSECV	4.25	3.69	3.26	3.65	3.37	3.45	4.51	5.08	5.39	5.39
	REMSEP	5.35	4.96	4.31	4.07	4.29	3.79	4.58	4.37	5.71	5.36
	R^2_C	0.92	0.93	0.97	0.93	0.94	0.95	0.86	0.82	0.75	0.74
	R^2_{CV}	0.82	0.86	0.89	0.87	0.88	0.88	0.81	0.77	0.70	0.7
	R^2_P	0.73	0.78	0.83	0.84	0.83	0.86	0.77	0.8	0.68	0.72
	RER	6.29	6.78	7.81	8.27	7.85	8.88	7.08	7.42	5.89	6.28

The following step was the fine tuning of the models using combinations of specific regions of the spectra, so as to optimise instrument performance.

3.3.3.4. Selection of the best spectral region

For the development of the models using just specific regions of the spectra, all the wavelength (Table 3.3.3) combinations were tested for each instrument. Results

highlighted the spectral windows yielding the lowest RMSE errors for each parameter and equipment (Figure 3.3.5). These models were then used to predict the test set. As expected, due to the different spectral ranges and sensitivities, no instrument had the same spectral window for a given parameter. The required number of latent variables for the optimisation of the predictive model were duly noted and the models were calculated (Table 3.3.4). A graphical representation of the experimental versus predicted soil parameter (eCEC in this case) values (cross-validation and prediction) using specific regions of the spectra is presented in Figure 3.3.6. By comparing these results with the ones using the whole spectra, it is possible to conclude that the R^2_p for all the parameters were higher in the models developed using only specific regions of the spectra, thus justifying the strategy implemented. By using the whole spectral range it is possible that some unwanted noise is introduced into the models as well as information not related to the specific parameter that is being estimated, which in turn may hide or impair the desired information. Again, models for TN did not exhibit very good results (R^2_P values: 0.39-0.54). TOC models, however (R^2_P values: 0.49-0.80) were good when combinations of different spectral regions were used, except for the Spectral Evolution SM-3500 Vis-NIR (R^2_P=0.49). Models for the remaining parameters performed quite well (Table 3.3.4). The results obtained with specific regions reinforce the importance of selecting the best spectral region for each analyte.

One of the reasons for such a disparity of results, particularly when using the whole spectral range, is believed to be connected with the range exhibited by the values obtained through laboratory analysis. Values for TN ranged from 0.05-0.77%; TOC: 0.46-8.01%; pH: 5.2-9.1; moisture: 0.16-36.3%; effective CEC: 0.42-41.79 meq/100g (Table 2). For instance, values for TN of some of the samples were so low (below the detection limit of the laboratory analysis) that a value of 0.05 (which is half of the detection limit value) was given for descriptive statistical purposes. Furthermore, TN values as well as TOC values for the samples used in this study seem to be rather low when compared to values of the same parameters found in the literature. For instance, [37] studied the effects of moisture content (MC) and texture on the prediction of OC and TN with vis-NIR spectroscopy, and reported values of TN between 0.9-3.1% and of OC between 9.4-35% [38] reported values of OC between 0.5-40.8% in their study, while [39] had a TN range in their soil samples of 0.6-2.8%. Regarding soil pH and eCEC values, [40] reported values between 4.4-10.1 and 1.4-51.8 meq/100g. For eCEC, [41] reported values between 21.58-37.99 meq/100g in a California rice field.

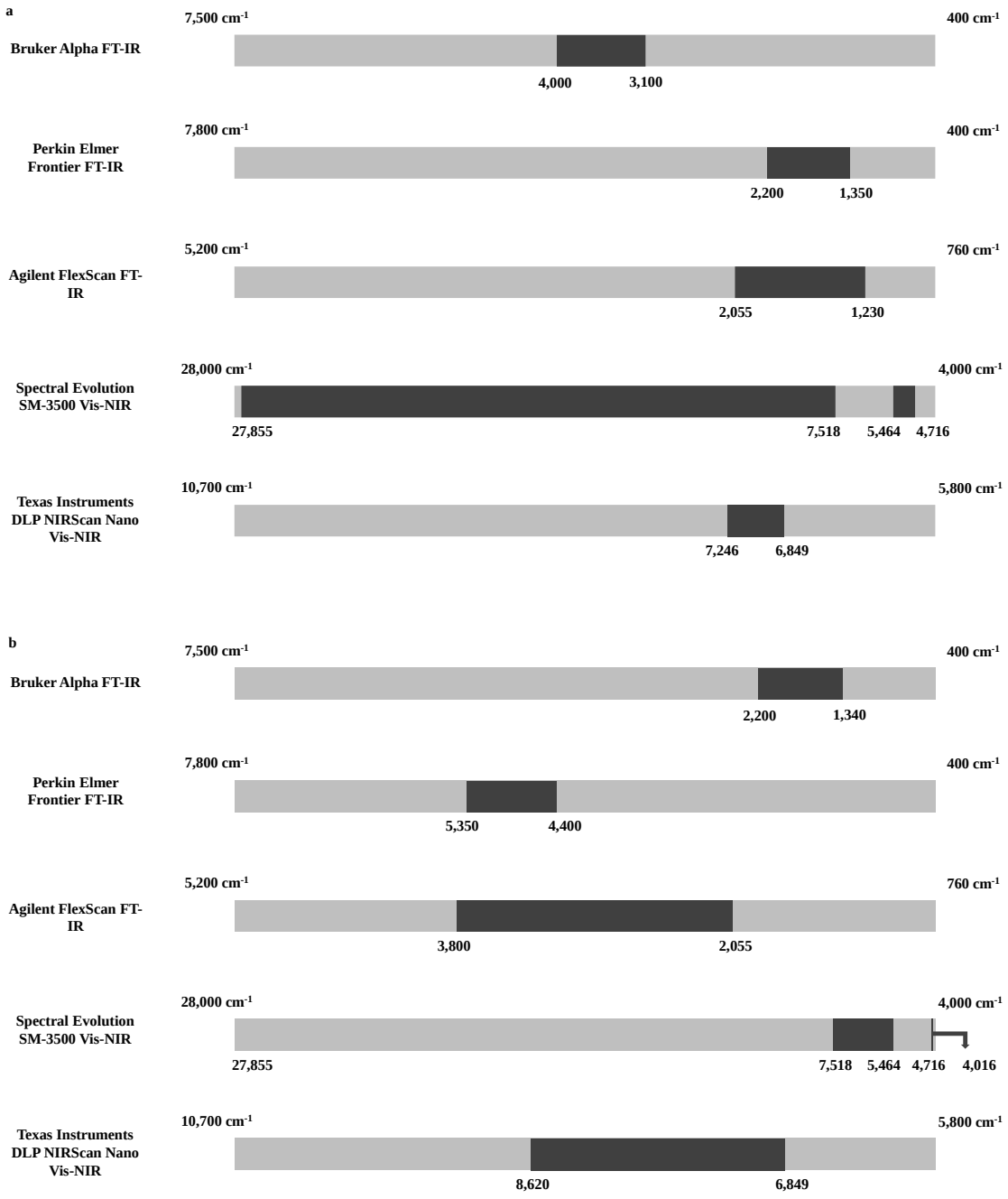


Figure 3.3.5. Regions identified as producing the lowest errors for each instrument and soil parameter analysed. **a**-TN; **b**-TOC; **c**-pH; **d**-Moisture; **e**-eCEC

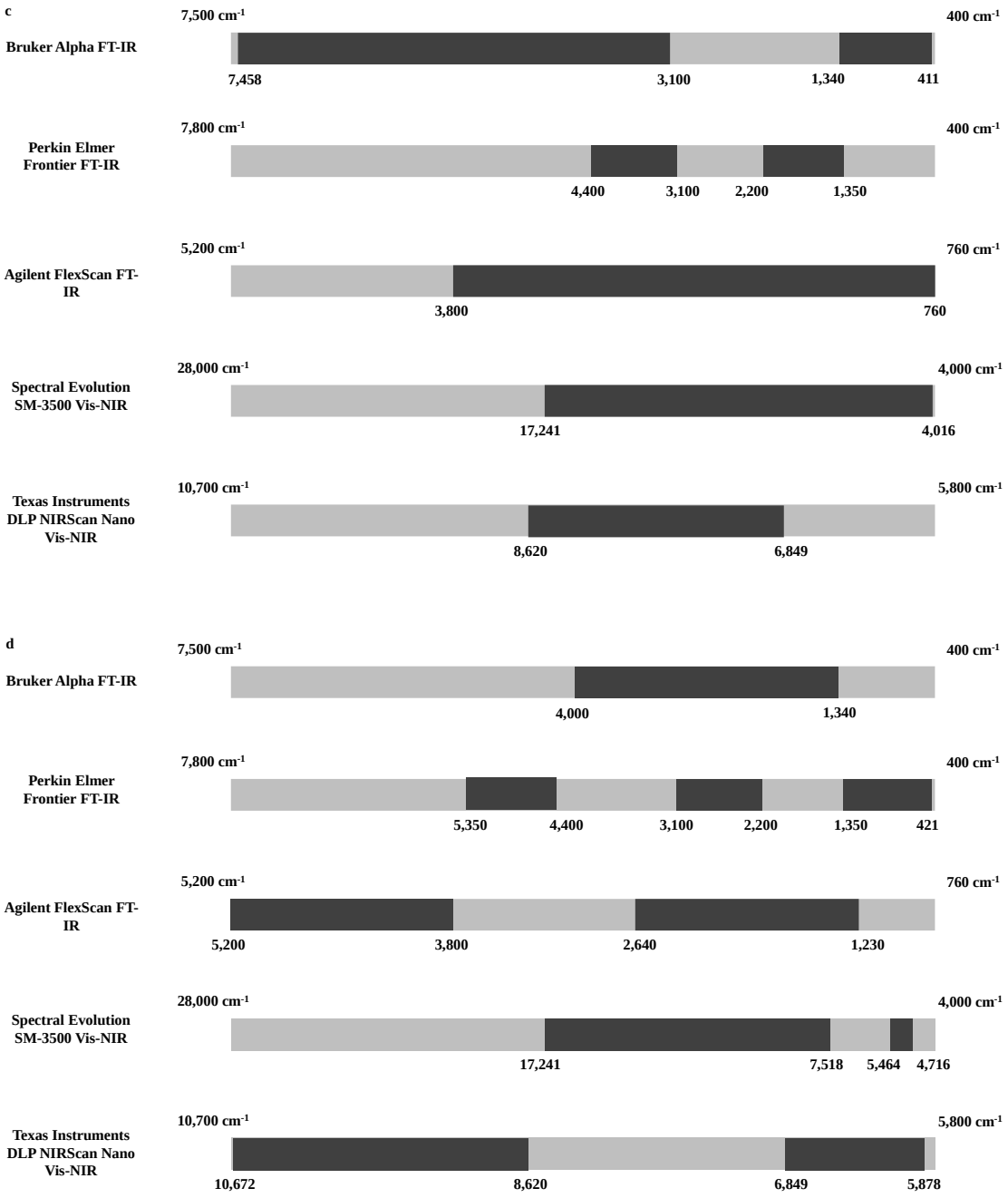


Figure 3.3.5. **(cont.).** Regions identified as producing the lowest errors for each instrument and soil parameter analysed. **a**-TN; **b**-TOC; **c**-pH; **d**-Moisture; **e**-eCEC

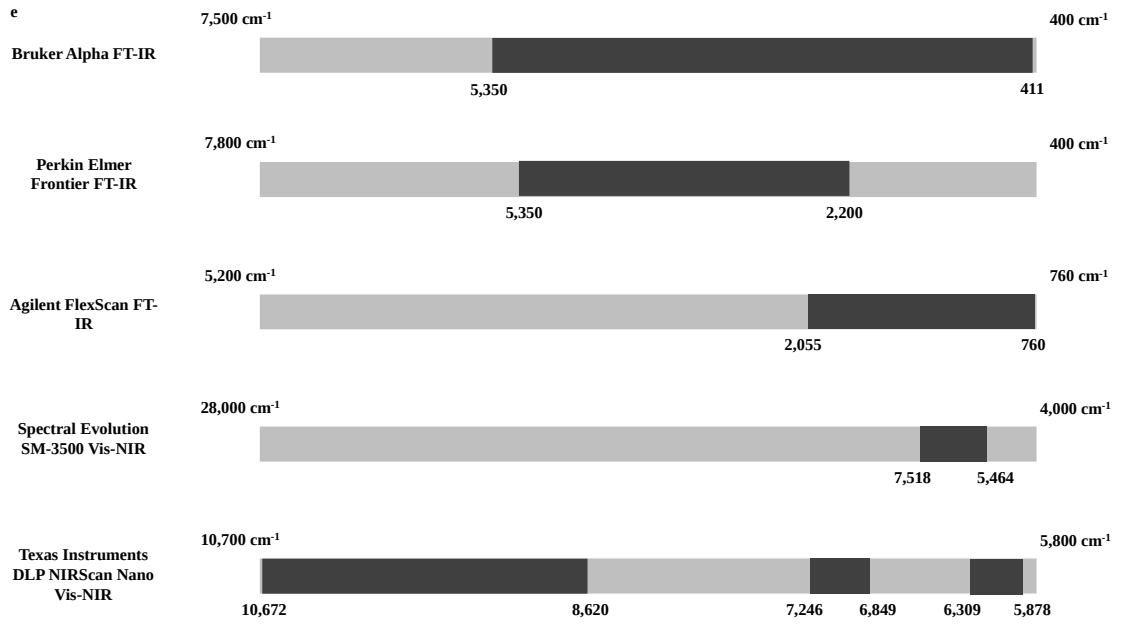


Figure 3.3.5. **(cont.).** Regions identified as producing the lowest errors for each instrument and soil parameter analysed. **a**-TN; **b**-TOC; **c**-pH; **d**-Moisture; **e**-eCEC

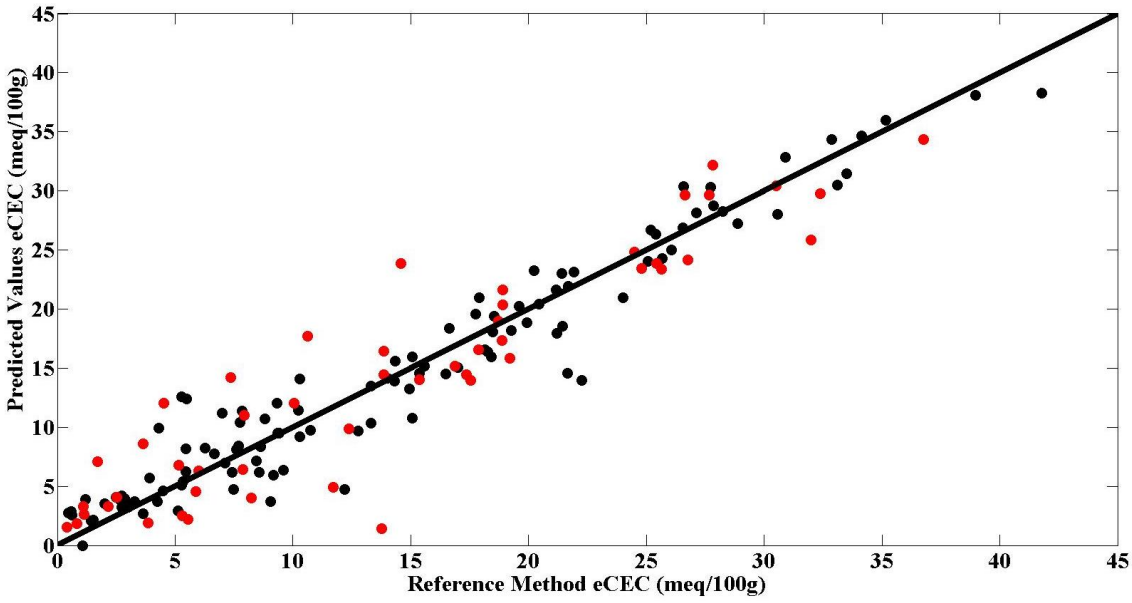


Figure 3.3.6. Comparison between Effective CEC determined with the reference method and the IR based method for the calibration samples (cross-validation predictions) and the independent test samples using specific regions of the spectra (• - cross-validation; • - prediction) for the Agilent FlexScan FT-IR.

Mixed results are found in the literature for the estimation of these parameters using IR spectroscopy. For instance Wijewardane and co-workers [40] reported quite good R^2_P values for total carbon (0.67) and very poor results for pH and eCEC (0.08 and 0.24, respectively) with vis-NIR. On the contrary, the study by Van Groenigen et al. [41] presented very poor results for total C (R^2_P= 0.01) and N (R^2_P= 0.20 and 0.02) and very good ones for eCEC (R^2_P= 0.83 and 0.56) with both NIR and DRIFT-MIR technologies, respectively. Soil's intricate complexity with superimposed layers of different soil types could explain this ambiguous behaviour. In fact, several studies mention the need for local calibrations when predicting soil analytes [42, 43]. From the results of the models developed using specific regions of the spectra it is possible to conclude that there is no definite "best instrument" for the estimation of all these soil parameters. For the portable instruments, the Bruker Alpha FT-IR seems to exhibit better results when the whole spectra is used for modelling, whereas the Agilent FlexScan FT-IR seems to provide better results when specific regions of the spectra are used, rivalling the results obtained with the benchtop PerkinElmer Frontier FT-IR. The surprising result was that the Bruker Alpha FT-IR performance when using specific regions of the spectra was worse than the other MIR instruments, despite having the same spectral range. One possible reason for this outcome could be the scanning methodology which may need some adjustment. The vis-NIR instruments performed poorly compared to the MIR instruments, but exhibited similar results between both instruments, with a slight advantage to the Spectral Evolution SM-3500 Vis-NIR. However, the extremely small size and portability of the Texas Instruments DLP NIRScan Nano Vis-NIR confer it a significant advantage.

Our results seem to indicate that out of the five different IR spectrometers used, there is no "best" IR equipment for the estimation of the different soil parameters tested. Furthermore, the performance of the different instruments is not consistent across the different spectral regions used, meaning that the best estimation of parameters does not necessarily occur with the same equipment when using the whole spectra or specific regions of the spectra. For instance, taking into consideration only the portable instruments, the Bruker Alpha FT-IR had the best R^2_P for moisture (0.67) in the models when the whole spectra was used, whereas in the models where specific regions of the spectra were used, the best result was obtained with the Agilent FlexScan FT-IR (R^2_P =0.74). The benchtop Perkin Elmer Frontier FT-IR was expected to yield the best results due to its broader spectral range and higher sensitivity but surprisingly that was not the case for the models developed using specific areas of the spectra, where it only had the best estimation for TOC and pH.

3.3.4. Conclusions

In this study, five different IR instruments were tested for the estimation of different soil parameters from samples collected on several vineyards in the McLaren Vale wine region (South Australia). Preliminary PLS analysis using the whole spectra revealed good results for the estimation of pH (R^2_p: 0.51-0.62), moisture (R^2_p: 0.47-0.71) and effective CEC (R^2_p: 0.68-0.83) and weaker results for the estimation of TN and TOC (with the exception of the benchtop Perkin Elmer Frontier FT-IR), which exhibited R^2_p values were between (0.32-0.49) and (0.28-0.67) respectively. This trend continued in the models developed using specific regions of the spectra, except with TOC which exhibited quite good results (R^2_P values: 0.49-0.80). The poor results for TN are probably due to the low amounts of N present in the samples collected. The instrument with the best result for a given parameter using the whole spectra was not necessarily the same when specific regions of the spectra were modelled, leading to the conclusion that, at least for the spectrometers used, there is no best equipment for predicting the different soil constituents analysed. However, all the FT-IR instruments exhibited better results than the vis-NIR instruments, both when the whole spectra and specific regions of the spectra were used. The portable Agilent FlexScan FT-IR and benchtop Perkin Elmer Frontier FT-IR had similar performance and better results when specific regions of the spectra were used, whereas the Bruker Alpha FT-IR had the best performance for the portable instruments when the whole spectra was used. Both vis-NIR instruments performed similarly, with the Spectral Evolution SM-3500 Vis-NIR presenting slightly better models. On the other hand, the extreme portability of the Texas Instruments DLP NIRScan Nano Vis-NIR is appealing.

This study demonstrated that there are IR portable instruments capable of results which are of the same quality as those obtained with benchtop spectrometers, allowing for reliable swift, *in-situ* soil analysis.

3.3.5 References

1. Skjemstad, J.O., et al., Characterization of soil organic matter by solid-state 13C NMR spectroscopy in Driven by nature: plant litter quality and decomposition, E. G. Gadish and K. E. Giller, Editor. 1997, CAB international: CAB International, Wellington, UK. p. 253–271.
2. Weil, R.R. and N. Brady, The nature and properties of soils. Pearson Education Ltd., USA, 2002. **5**: p. 359-387.
3. Robertson, A.J., C. Shand, and E. Perez-Fernandez, The application of Fourier transform infrared, near infrared and X-ray fluorescence spectroscopy to soil analysis. 2016.
4. Stenberg, B., et al., Chapter five-visible and near infrared spectroscopy in soil science. Advances in agronomy, 2010. **107**: p. 163-215.
5. Robert, P.C., W.H. Thompson, and D. Fairshild. Soil specific anhydrous ammonia management system. in Symposium on Automated Agriculture for the 21st century. 1991.
6. Gebbers, R. and V.I. Adamchuk, Precision agriculture and food security. Science, 2010. **327**(5967): p. 828-831.
7. Rossel, R.V., et al., Using a legacy soil sample to develop a mid-IR spectral library. Soil Research, 2008. **46**(1): p. 1-16.
8. Madhavan, D.B., et al., Mid-infrared spectroscopy for rapid assessment of soil properties after land use change from pastures to Eucalyptus globulus plantations. Journal of environmental management, 2016. **175**: p. 67-75.
9. Aranda, V., et al., Agro-environmental characterization of semi-arid Mediterranean soils using NIR reflection and mid-IR-attenuated total reflection spectroscopies. Vibrational Spectroscopy, 2014. **74**: p. 88-97.
10. Du, C. and J. Zhou, Evaluation of soil fertility using infrared spectroscopy: a review. Environmental Chemistry Letters, 2009. **7**(2): p. 97-113.
11. Du, C., R. Linker, and A. Shaviv, Identification of agricultural Mediterranean soils using mid-infrared photoacoustic spectroscopy. Geoderma, 2008. **143**(1): p. 85-90.
12. Cozzolino, D., et al., In situ measurement of soil chemical composition by near-infrared spectroscopy: A tool toward sustainable vineyard management. Communications in soil science and plant analysis, 2013. **44**(10): p. 1610-1619.
13. Lopo, M., et al., Classification of Vineyard Soils Using Portable and Benchtop Near-Infrared Spectrometers: A Comparative Study. Soil Science Society of America Journal, 2016. **80**(3): p. 652-661.

14. Páscoa, R., et al., Exploratory study on vineyards soil mapping by visible/near-infrared spectroscopy of grapevine leaves. Computers and Electronics in Agriculture, 2016. **127**: p. 15-25.
15. Edwards, J. Setting benchmarks and recommendations for management of soil health in Australian viticulture. Final Report to Australian Grape and Wine Authority Project Number DPI1101, July 2014.
16. Matejovic I. Determination of Carbon and Nitrogen in samples of various soils by the dry combustion. Communications in Soil Science and Plant Analysis, 1997. **28:** p. 1499-1511.
17. Rayment, G.E and Lyons, D.J. Carbonates by pressure change - transducer (method 19B2). In 'Soil Chemical Methods - Australasia'. (CSIRO publishing, Melbourne), 2011a. p. 420-422.
18. Rayment, G.E and Lyons, D.J. pH of 1:5 soil/water suspension (method 4A1). In 'Soil Chemical Methods - Australasia'. (CSIRO publishing, Melbourne), 2011b. p. 38-39.
19. Rayment, G.E and Lyons, D.J. pH of 1:5 soil/0.01M calcium chloride extract – following method 4A1 (method 4B4). In 'Soil Chemical Methods - Australasia'. (CSIRO publishing, Melbourne), 2011c. p. 41.
20. Rayment, G.E and Lyons, D.J. Electrical conductivity (EC) of 1:5 soil/ water extract (method 3A1). In 'Soil Chemical Methods - Australasia'. (CSIRO publishing, Melbourne), 2011d. p. 20-22.
21. Rayment, G.E and Lyons, D.J. (2011e). Exchangeable bases and CEC (method 15D2). In 'Soil Chemical Methods - Australasia'. (CSIRO publishing, Melbourne), 2011e. p. 318-321.
22. Dor, E.B., C. Ong, and I.C. Lau, Reflectance measurements of soils in the laboratory: Standards and protocols. Geoderma, 2015. **245**: p. 112-124.
23. Savitzky, A. and M.J. Golay, Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry, 1964. **36**(8): p. 1627-1639.
24. Naes, T., et al., Interpreting PCR and PLS solutions, in A User-Friendly Guide to Multivariate Calibration and Classification. 2004, NIR Publications: Chichester, UK. p. 39-54.
25. Geladi, P. and B.R. Kowalski, partial least-squares regression - a tutorial. Analytica Chimica Acta, 1986. **185**: p. 1-17.
26. Madejová, J., FTIR techniques in clay mineral studies. Vibrational spectroscopy, 2003. **31**(1): p. 1-10.

27. Janik, L.J. and J. Skjemstad, Characterization and analysis of soils using mid-infrared partial least-squares. 2. Correlations with some laboratory data. Soil Research, 1995. **33**(4): p. 637-650.
28. Soriano-Disla, J., et al., Prediction of the concentration of chemical elements extracted by aqua regia in agricultural and grazing European soils using diffuse reflectance mid-infrared spectroscopy. Applied geochemistry, 2013. **39**: p. 33-42.
29. Janik, L.J., R.H. Merry, and J. Skjemstad, Can mid infrared diffuse reflectance analysis replace soil extractions? Animal Production Science, 1998. **38**(7): p. 681-696.
30. Soriano-Disla, J.M., et al., The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Applied Spectroscopy Reviews, 2014. **49**(2): p. 139-186.
31. Rossel, R.V., et al., Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, 2006. **131**(1): p. 59-75.
32. Reeves, J.B., Near-versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: where are we and what needs to be done? Geoderma, 2010. **158**(1): p. 3-14.
33. Rossel, R.V., et al., In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma, 2009. **150**(3): p. 253-266.
34. Clark, R.N., et al., High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research: Solid Earth, 1990. **95**(B8): p. 12653-12680.
35. Whiting, M.L., L. Li, and S.L. Ustin, Predicting water content using Gaussian model on soil spectra. Remote Sensing of Environment, 2004. **89**(4): p. 535-552.
36. Sherman, D.M. and T.D. Waite, Electronic spectra of Fe3+ oxides and oxide hydroxides in the near IR to near UV. American Mineralogist, 1985. **70**(11-12): p. 1262-1269.
37. Kuang, B. and A.M. Mouazen, Non-biased prediction of soil organic carbon and total nitrogen with vis–NIR spectroscopy, as affected by soil moisture content and texture. Biosystems engineering, 2013. **114**(3): p. 249-258.
38. Chang, C.-W., D.A. Laird, and C.R. Hurburgh Jr, Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. Soil Science, 2005. **170**(4): p. 244-255.
39. Xie, H., et al., Predicting soil organic carbon and total nitrogen using mid-and near-infrared spectra for Brookston clay loam soil in Southwestern Ontario, Canada. Canadian Journal of Soil Science, 2011. **91**(1): p. 53-63.

40. Wijewardane, N.K., Y.F. Ge, and C.L.S. Morgan, Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. Geoderma, 2016. **267**: p. 92-101.
41. Van Groenigen, J., et al., NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. Plant and Soil, 2003. **250**(1): p. 155-165.
42. Lee, K.S., et al., Calibration methods for soil property estimation using reflectance spectroscopy. Transactions of the ASABE, 2010. **53**(3): p. 675-684.
43. Lucà, F., et al., Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. Geoderma, 2017. **288**: p. 175-183.

The following work is intended for future submission in a scientific journal:

Lopo, M., Soriano-Disla, J. M., Oliver, D. P., Páscoa, R. N., Janik, L., McLaughlin, M. J., & Lopes, J. A. Infrared spectroscopy suitability for the prediction of important soil properties for vine's growth and soils discrimination in Australian vineyards.

3.4. Infrared spectroscopy suitability for the prediction of important soil properties for vine's growth and soils discrimination in Australian vineyards

Abstract: The preservation of agricultural soils is paramount for the human population welfare. Recently, the demands of consumers have raised the concept of precision agriculture with the objective of designing tailor-made strategies for specific cultivars. Soil analysis is traditionally based on wet chemistry laboratory methods which are expensive, time-consuming and laborious. Infrared (IR) spectroscopy using either mid infrared (MIR) or near infrared (vis-NIR) is a rapid, non-destructive, cost-effective and reliable technique for the monitoring of several chemical and biological parameters. In this work, the potential of IR spectroscopy employing portable instruments to estimate a broad range of different soil chemical parameters in different Australian vineyards was evaluated. Phosphorous (P), potassium (K), sulphur (S), Conductivity, pH ($CaCl_2$), Calcium carbonate, chloride, exchangeable cations: calcium (exch. Ca), potassium (exch. K), magnesium (exch. Mg), sodium (exch. Na) and exchangeable sodium percentage (ESP) were estimated using a Fourier-transform infrared (FT-IR) spectrometer and a vis-NIR one. Furthermore, it was investigated if this technology was able to accurately classify different soil types and thus enable the possible development of a robust soil mapping. Results revealed partial least squares (PLS) R^2 values for parameters ranging from 0.27 to 0.99 using both MIR and NIR instruments, R^2 values below 0.20 were not considered and are not shown. Partial least squares discriminant analysis (PLS-DA) also demonstrated that IR spectroscopy was able to accurately differentiate vineyard soil types. Using the FT-IR spectrometer 74.9% of correct predictions were obtained regarding the analysed soils, whereas using the vis-NIR spectrometer to classify the same soils yielded 67.2% of correct predictions. Results using MIR were in almost every parameter better than using NIR. These results point out that IR spectroscopy can be a useful tool in the estimation of important soil parameters as well as differentiate specific soil types. This technology can therefore play an important part in the management of a vineyard, but also any other agricultural plantation.

3.4.1. Introduction

The main objective in every agricultural endeavour is to maximise the yield and quality of crops in a non-depleting and sustainable way. With the availability of arable fields diminishing ever-so constantly and the needs of a population permanently increasing, it is imperative that preventive steps should be taken. The preservation and protection of agricultural fields with appropriate, long term management plans is therefore paramount. The general population demand and raised standards as a consumer, led to several adaptations from the producers and to the introduction of tailored-made strategies in every field of agriculture. Soil, no doubt, plays an important part in every agricultural production. The assessment of its status regarding specific needs and goals as well as the understanding of its dynamics are undoubtedly essential. Soil analysis, traditionally, is based on wet chemistry methods which are time-consuming, laborious and for the most part, almost prohibitively expensive. Infrared (IR) spectroscopy using either mid infrared (MIR) or near infrared (vis-NIR) is a rapid, non-destructive, cost-effective and reliable technique for the monitoring of several chemical and biological parameters. This methodology requires no sample preparation, no chemical reagents and is highly adaptable to automated and *in situ* measurements [1]. The MIR wavelength region (4000-400 cm^{-1}) is characterized by fundamental molecular vibrations associated with particular chemical functional groups (e.g. aliphatic, amidic, aromatic and carboxylic) Several soil properties (chemical, physical and biological) have been successfully predicted using this technology, including organic carbon (OC), inorganic C, total nitrogen (TN), pH, electric conductivity (EC), sand, clay and microbial biomass [2-4]. In recent years, vis-NIR reflectance spectroscopy has also been shown to be a useful technique for the measurement of various soil properties and has proven to be a rapid and convenient means to analyse many soil constituents at the same time [5]. A combination of spectroscopy and chemometric tools such as principal components analysis (PCA) and partial least squares (PLS) are some of the most common multivariate data analysis methods and have been used to demonstrate the usefulness of vibrational spectroscopy in this field [3, 4, 6, 7]. The aforementioned authors have studied several soil properties using MIR and NIR spectroscopy and have reported promising results both at calibration and validation stages. For instance, Waruru and colleagues [4] analysed several soil properties – cation exchange capacity (CEC), plasticity index (PI), linear shrinkage (LS), coefficient of linear extensibility (COLE), among others – with NIR spectroscopy and obtained coefficients of determination (R^2) ≥ 0.70. [6] used PLS regression to predict the contents of TN, OC, potassium (K), sulphur (S), phosphorous (P), pH, and exchange

capacity (EC) of soils in several Australian vineyards and obtained results (R^2) of 0.74 for TN, 0.92 for S, 0.81 for OC, 0.70 for K, 0.84 for EC, 0.83 for pH, and 0.69 for P. However these results were not confirmed with a validation set. In another study, Zornoza et al. [7], evaluated the potential of NIR spectroscopy to predict various physical, chemical and biochemical properties in Mediterranean soils. These authors, obtained validation results with R^2 values greater than 0.90 for OC, Kjeldahl nitrogen (N), soil moisture, CEC, among others; values between 0.81-0.90 for exchangeable calcium (exch. Ca) and magnesium (Mg), water soluble carbon, etc.; and values lower than 0.66 for EC, pH, exch. P and exch. sodium (Na). Using both MIR and vis-NIR spectra of soil samples, Vohland and co-workers [3] predicted the contents of several soil properties such as OC, pH and N, firstly using the full spectra with R^2_{CV} values of 0.60, 0.72 and 0.37 respectively for vis-NIR and 0.78, 0.75 and 0.77 for MIR. The authors followed their study by the application of a specific variable selection procedure and improved their results. R^2 values for OC, pH and N using vis-NIR spectra were 0.74, 0.81 and 0.58 respectively and 0.91, 0.92 and 0.92 for the same parameters using MIR spectroscopy.

Soil classification is also capital for a thorough understanding of the all the variables affecting an agricultural field. The conventional soil characterization survey is normally performed through extensive field observations followed by laboratory analysis [8]. These chemical analysis are no-doubt effective, but due to practical limitations (costs, time) are generally only applied to a limited number of samples and restricted to specific areas of the field, even during comprehensive soil characterisation. The generated data will have little or no representative information on the spatial variability of soils across a large region [2]. For this reason, the need for quick, reliable, and cost-effective techniques for *in situ* soil analysis is unquestionably urgent. IR spectroscopy applied to soil classification is yet somewhat of an undisclosed novelty. Only a handful of studies using this technology for the aforementioned purposed exist. Lopo et al. [9] analysed the potential of NIR spectroscopy to map vineyard soils both in the field and in the laboratory with successful soil identification rates between 75 to 100%. Using MIR spectroscopy, Linker [10] successfully classified several Israeli agricultural soils, achieving percentages of correct classifications close to 100%. Utilising an indirect approach, Páscoa and colleagues [11] were able to discriminate different soil types through vis-NIR spectroscopy by the scanning of grapevine leaves on several vineyards. The analysis of the leaves spectra led to an accurate mapping of the studied vineyards soils. On a slightly different note, Awiti and team [12] were able, using IR spectroscopy discriminant analysis, to accurately classify different soil types according to their degree of degradation. These authors used this approach to assess the health of agricultural soils in a sub-Saharan

African region. In another study, Mouazen et al. [13] collected different soil samples from Belgium and northern France with different texture classes and were able, through vis-NIR spectroscopy, to classify soils into four different textural groups with 81.8% of correct classifications. To note that most of the cited works were not performed *in-situ*, with samples acquired in the field and spectra collected in the laboratory. There are not many studies that address *in-situ* soil analysis using IR spectroscopy.

The main objective of this study was to investigate the potential of IR spectroscopy employing portable instruments to estimate a broad range of different soil chemical parameters such as phosphorous (P), potassium (K), sulphur (S), Conductivity, pH ($CaCl_2$), Calcium carbonate, chloride, exchangeable cations: calcium (exch. Ca), potassium (exch. K), magnesium (exch. Mg), sodium (exch. Na) and exchangeable sodium percentage (ESP), simulating their in-filed applicability. A Fourier-transform infrared (FT-IR) spectrometer and a vis-NIR spectrometer were tested and compared. A second goal was to understand if IR spectroscopy is able to accurately classify different soil types and thus enable the possible development of a robust soil mapping. This approach may provide an advantageous alternative to currently existing methods for soil analysis and also a better understanding, from the producer's point of view, for the selection of soils best "matching" specific characteristics. Furthermore, it will permit a spectral case definition of specific soils, leading to better targeting of management interventions.

3.4.2. Materials and Methods

3.4.2.1. Samples collection and preparation

A total of 160 soil samples were acquired from forty vineyards in the McLaren Vale (South Australia -35° 14' 00.00" S 138° 31' 60.00" E), one of the most important wine producing regions of Australia and of the world. The sampling locations were selected over a wide geographical area with the objective to cover the broad range of soil types present in the region (Figure 3.4.1). At each location (vineyard), several sampling points were chosen and an "undervine" and a "midrow" sampling point were selected. The undervine sample was collected 20 cm away from the grapevine trunk while the midrow sample was collected in the centre of the path between two rows of plants, in line with the undervine sample. Approximately 1 kg of soil was taken from two different depths (0-10 cm and 35-45 cm) at each sampling point (Table 3.4.1). The samples were stored in rigid plastic jars for transportation before physical and chemical characterisation. Prior to spectral

acquisition, samples were air-dried, oven-dried at 40 °C for 12 h, cooled in a desiccator, sieved to <2 mm and homogenized. Further details of the soil collection and the vineyard sites are given in [14].

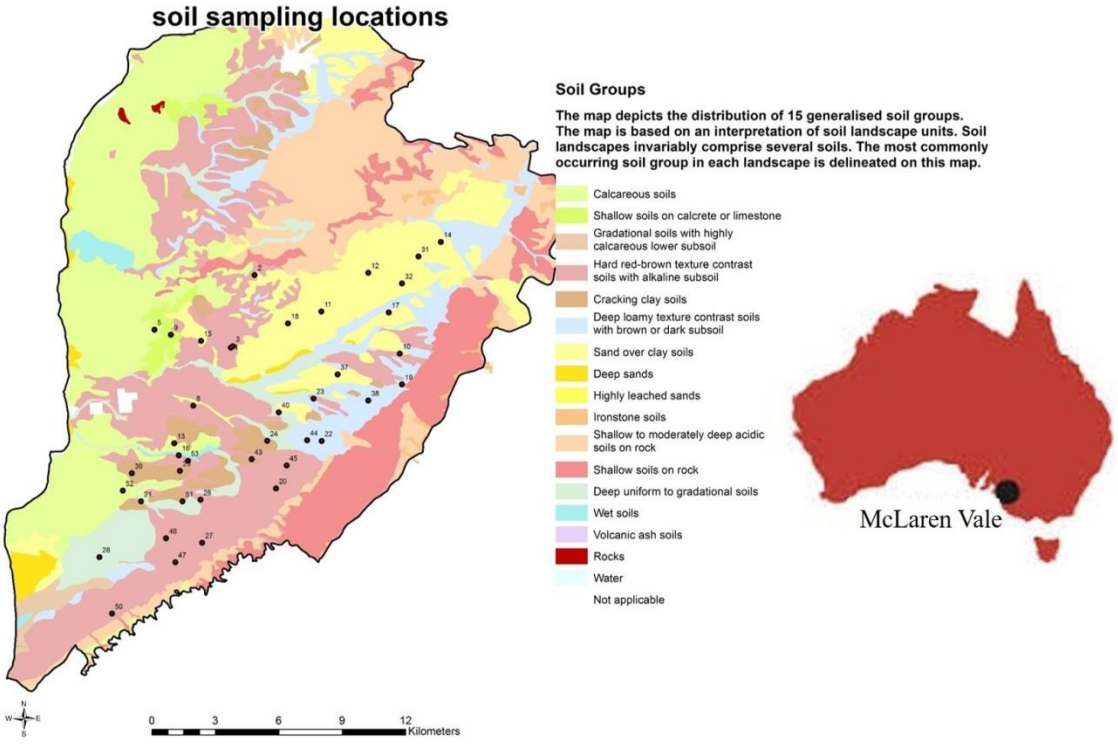


Figure 3.4.1. 1:50.000 soil map for McLaren Vale obtained from Primary Industries and Regions. South Australia (PIRSA).

Table 3.4.1. Sampling locations in the different vineyards performed in two subsequent years at specific depths.

Vineyard	Sampling Points at each location	Location	Soil type[1]	Depth	Years
Chapel Hill	2	Undervine, Midrow	6	0-10; 35-45cm	2
Oliver's Taranga	4	Undervine, Midrow	4, 6	0-10; 35-45cm	2
Paxton	4	Undervine, Midrow	1, 3	0-10; 35-45cm	2
Gemtree	4	Undervine, Midrow	4	0-10; 35-45cm	2
Rosemount	4	Undervine, Midrow	2, 3	0-10; 35-45cm	2
d'Arenberg	4	Undervine, Midrow	6	0-10; 35-45cm	2

Table 3.4.1. **(cont.).** Sampling locations in the different vineyards performed in two subsequent years at specific depths.

Inkwell	2	Undervine, Midrow	1	0-10; 35-45cm	2
Cameron	4	Undervine, Midrow	3, 4	0-10; 35-45cm	2
Coriole	4	Undervine, Midrow	4, 5, 6	0-10; 35-45cm	2
Noons	2	Undervine, Midrow	5	0-10; 35-45cm	2
Leask Vineyards	2	Undervine, Midrow	4	0-10; 35-45cm	2
Bottin	1	Undervine, Midrow	7	0-10; 35-45cm	1
Leconfield	1	Undervine, Midrow	7	0-10; 35-45cm	1
Yangarra Estate	2	Undervine, Midrow	6	0-10; 35-45cm	2

[1] 1 – Calcareous soils; 2 – Shallow soils on concrete or limestone; 3 – Gradational soils with highly calcareous lower subsoil; 4 – Hard red-brown texture contrast soils with alkaline subsoil; 5 – Deep loamy texture contrast soils with brown or dark subsoil; 6 – Sand over clay soils; 7 – Deep uniform to gradational soils.

3.4.2.2. Soil chemical analysis

Soil pH and EC were determined using a 1:5 soil/water extract. Air dried material (5g) was shaken with 25ml water for one hour and left to settle for 20 minutes. pH was firstly determined in water before calcium chloride was added. Samples were agitated and re-settled before reading for pH in calcium chloride. Electrical conductivity was determined and a subsample for chloride taken. EC was determined using a Radiometer CD230 and pH was measured using a Metrohm 815 Robotic Processor [15-17].

Exchangeable cations were determined using NH_4Cl solution at either pH 7.0 or pH 8.5 depending on soil pH. Samples were pre-treated for soluble salts prior to extraction. Non calcareous soils use extraction solution pH 7.0 and alkaline soils use extraction solution pH 8.5. Exchangeable cations, Ca, Mg, Na and K, were analysed by Flame Atomic Absorption Spectrometry. Chloride was analysed by flow injection analyser [18].

Table 3.4.2 shows the range, mean, standard deviation (SD) and coefficient of variation (CV) for the chemical properties analysed.

Table 3.4.2. Descriptive statistics for chemical properties of the different soil parameters used in modelling.

Soil Property	Min	Max	Mean	SD	CV
P (mg/Kg)	1.00	321.00	51.88	58.82	113.38
K (mg/Kg)	0.55	1561.00	340.73	312.96	91.85
S (mg/Kg)	1.50	1282.00	36.95	116.89	316.37
Conductivity (dS/m)	0.01	1.08	0.18	0.17	93.31
pH ($CaCl_2$)	4.60	8.10	6.68	0.87	13.09
Calcium Carbonate (%)	0.17	19.02	1.90	3.98	209.42
Chloride (mg/Kg)	2.00	847.60	92.59	137.68	148.69
exch. Ca (meq/100g)	0.25	33.18	10.60	8.18	77.18
exch. K (meq/100g)	0.01	6.93	0.85	0.99	116.64
exch. Mg (meq/100g)	0.05	9.81	2.28	2.03	88.81
exch. Na (meq/100g)	0.02	9.24	0.48	1.23	254.49
ESP (%)	0.13	57.32	4.44	10.33	232.85

Notes. CV = (SD/mean) x 100; min, minimum; max, maximum; and SD, standard deviation.

3.4.2.3. Spectral acquisition

Two hand-held IR instruments were used for the acquisition of spectra. The first instrument was a FlexScan FT-IR model 4200 (Agilent, USA) spectrometer equipped with a Michelson interferometer, zinc selenide beam splitter and thermoelectrically cooled DTGS detector. In this instrument, the optical and electronic/battery components are split into two, the electronics/battery and the optical component, connected by a communications cable. For spectral acquisition, the optical component was positioned facing down, samples were placed in stainless steel cups (3 mm depth , 9 mm diameter) and lifted by compressed air against the optical aperture allowing a high reproducibility in sample presentation. Spectra were acquired through a DRIFT accessory over the frequency range of 6000–650 cm^{-1}, a course-grained SiC reference disk was used as background and scanned every hour.

The second instrument was a miniature NIRScan Nano vis-NIR (Texas Instruments, USA) spectrometer. The spectra were acquired by placing the samples in close contact to a sapphire window which was illuminated with two integrated and angled tungsten infrared lamps. The background was scanned with the white reference (Spectralon™, Spectral Evolution, USA) every 30 minutes. In this instrument the diffusively reflected light is congregated and focused through the input slit (25 µm wide by 1.69 mm tall), the light then strikes a reflective grating which disperses, in combination

with a focusing lens, the light into the different wavelengths. The focusing lenses form an image of the slit at the Digital Light Proccessing (DLP) micro-mirror device (DMD; (0.2-inch WVGA, 854 × 480 orthogonal pixel, NIR optimized). The energy reflected by the DMD is directed through the collection optics to the single pixel InGaAs detector (11111-5882 cm^{-1}).

Spectra were measured in triplicate, making a total of 480 spectra for each instrument. For each sample, the average spectrum was considered for modelling.

3.4.2.4. IR spectra modelling

The acquired spectra was analysed using principal component analysis (PCA), partial least squares (PLS) and partial least squares discriminant analysis (PLS-DA). PCA was primarily used to detect outliers and perform an exploratory data analysis on the spectra, followed by the development of calibration models for soil parameters determination through PLS [19] and PLS-DA for soil type discrimination [20]. Before the application of PCA, PLS and PLS-DA, spectra were subjected to mean-centring as this is a requirement of the three methods [21]. Several pre-processing methods were tested for each soil parameter with Savitzky–Golay filter (15-point filter size, second-order polynomial, and first-order derivative) and (15-points filter size, second order polynomial, and second-order derivative) [22] followed by the application of standard normal variate (SNV) being the most common. For both PLS and PLS-DA models, data were randomly divided to form a calibration (70% of the available samples) and a test set (remaining 30%).

The PLS models were developed considering the PLS-1 algorithm described by [21]. For PLS, the optimal number of latent variables (LVs) was estimated, for each model, by leave-one-out cross-validation using only the calibration set and minimising the root mean square error of cross validation (RMSECV) [19]. The root mean square error of prediction (RMSEP) was used to evaluate the performance of the optimised model when the testing set was projected. Spectra were divided into five different regions for each equipment according to the major chemical/physical properties captured by IR spectra. Due to the different ranges of both spectrometers, the regions were different for each instrument. For the FT-IR instrument, the regions were 1230-760 cm^{-1} (Region 1), 2055-1230 cm^{-1} (Region 2), 2640-2055 cm^{-1} (Region 3), 3800-2640 cm^{-1} (Region 4) and 5200-3800 cm^{-1} (Region 5). For the vis-NIR instrument, the regions were 6309-5878 cm^{-1} (Region 1), 6849-6309 cm^{-1} (Region 2), 7246-6849 cm^{-1} (Region 3), 8620-7246 cm^{-1} (Region 4), and 10672-8620 cm^{-1} (Region 5). The best spectral region was estimated by testing all possible combinations, and selected according to the lowest root mean square

error (RMSE). Additional statistics such as coefficients of determination (R^2), experimental versus predicted response as well as the range error ration (RER) (Eq. 1) were also calculated to evaluate the predictive ability of the models.

RER=Parameter range/RMSEP (3.1)

The PLS-DA models were developed using the PLS-2 algorithm. Sample blocks division was made randomly while ensuring the same proportion between soil types was present in the calibration and test sets to avoid unbalanced classes (soil types) across sets [23]. As for PLS, the optimal number of LVs was estimated by leave-one-out cross-validation using only the calibration set. The test set results were then projected onto each optimized PLS-DA model and soil predictions were expressed as confusion matrices [24]. Confusion matrices compare each known soil type with the corresponding IR spectroscopy prediction. The objective of these matrices is not only to estimate the number of correctly predicted samples but also to define which samples are being incorrectly predicted, identifying the most similar soil types in terms of IR spectroscopy. Entries are expressed as percentages [23].

All chemometric methods and spectra processing were performed using Matlab version 7.9 (MathWorks. Natick. MA) and PLS Toolbox version 5.5.1 (Eigenvector Research Inc., Wenatchee. WA).

3.4.3. Results and Discussion

3.4.3.1. PCA Modelling

To analyse the IR spectral variability and identify common patterns within the different soil samples as well as detect outliers, PCA models were developed, for each instrument, using the entire wavelength range except noisy areas at both ends of the spectra (Figure 3.4.2). Soil differentiation using both instruments was not possible considering only two components. A detailed case-by-case analysis was made combining other components but no further discrimination was achieved. However, it should be mentioned that PCA maximises the variance criterion and no information about soils is provided. These results suggested the use of a more adequate analysis for soil discrimination, namely a supervised classification method such as PLS-DA that will be further discussed.

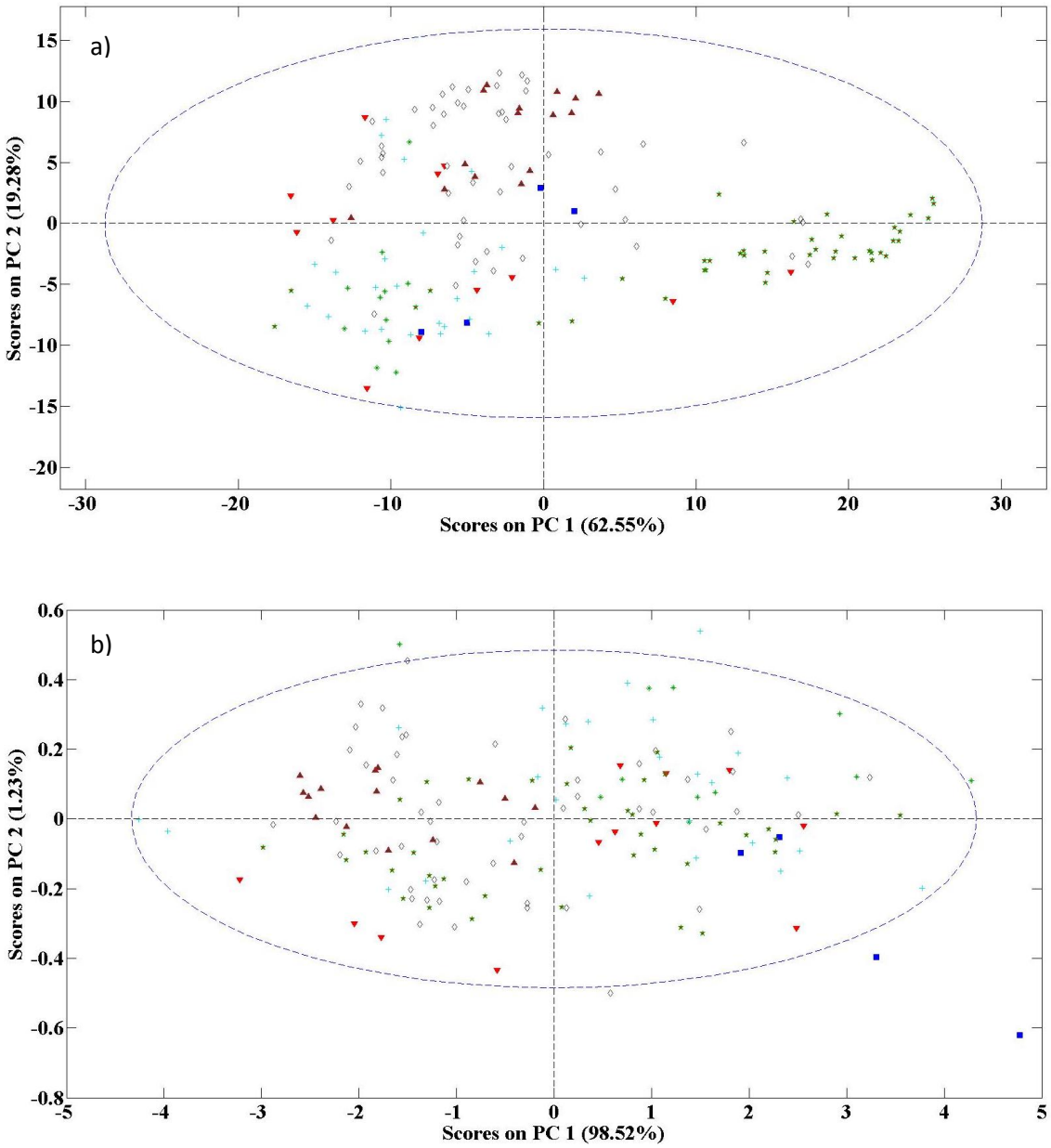


Figure 3.4.2. Score plots obtained from PCA models built on IR data obtained with both instruments from samples of different soil types (▼-1; ■-2; +-3; ◊-4; ▲-5;★-6; *-7).
a) Agilent Exoscan FT-IR; b) Texas Instruments DLP NIRScan Nano Vis-NIR.

3.4.3.2. Parameter determination

PLS models for the different chemical parameters analysed provided very different results. Pre-processing methods varied according to analyte and so did the spectral region gathering more information regarding that specific component, but that was to be expected. For the MIR equipment, Savitzky–Golay filter (15-point filter size, second-order polynomial, and first-order derivative) followed by SNV was the best pre-processing method for calcium carbonate, chloride and ESP; SNV was the best pre-processing option

for P, K, conductivity, pH and exch. Mg. Exch. Ca yielded better results with spectra just mean centred. The remaining parameters (S, exch. K and exch. Na) had R^2_P below 0.20 and therefore are not shown. Regarding the vis-NIR instrument, Savitzky–Golay filter (15-point filter size, second-order polynomial, and first-order derivative) followed by SNV was the best pre-processing method for pH, while Savitzky–Golay filter (15-point filter size, second-order polynomial, and second-order derivative) followed by SNV was the best option for calcium carbonate and ESP. For conductivity, the best pre-processing was SNV, and for K, chloride, exch. Ca and exch. Mg using only mean centre provided the best results. All the other parameters (P, S and exch. K) exhibited R^2_P values below 0.20. Exch. Na had an R^2_P value above 0.20, but its R^2 in calibration and cross-validation were so low that it was not taken into consideration. Calibration models were used, taking in consideration the lowest RMSECV, to pinpoint the best spectral regions for each instrument and each parameter. The best spectral combinations for each parameter are better shown in table format and can be seen in table 3.4.3. Each instrument has its own spectral range and therefore, very few parameters have the same spectral window between instruments or even within the same instrument. The only exceptions are pH and exch. Ca in the MIR equipment, but that is probably nothing more than coincidental. After the optimisation (in terms of best spectral regions, pre-processing and number of latent variables) for both instruments, the validation set was projected. Results from the predictive models (table 4) revealed R^2_P values above 0.60 for three parameters with both instruments. pH, exch Ca and exch. Mg presented R^2_P values of 0.72/0.61, 0.93/0.88, 0.77/0.64 for the MIR and NIR spectrometer respectively. Regarding the remaining parameters, the MIR instrument had a very good result for calcium carbonate (R^2_P = 0.99) contrasting with the poor result obtained with the NIR instrument (0.38) and a good result for K (R^2_P = 0.68). The MIR instrument also had an average result of R^2_P = 0.59 for P. The remaining parameters (conductivity, chloride and ESP) had poor predictive results with R^2_P values of 0.32, 0.36 and 0.34 respectively. The NIR instrument at first glance revealed an average predictive capacity for conductivity (R^2_P = 0.50), but the R^2 in calibration and cross-validation were in the order of 0.20 and therefore, this cannot be considered an average result. The other parameters exhibited poor results, besides the aforementioned parameters (pH, exch Ca and exch. Mg). Soil parameters K, calcium carbonate, chloride, exch. Na and ESP yielded R^2_P results of 0.42, 0.38, 0.37, 0.28 and 0.27 respectively. A graphical representation of the experimental versus predicted soil parameter (exch. Ca) values (cross-validation and prediction) for both instruments is presented in Figure 3.4.3 as an example.

Table 3.4.3. Spectral regions yielding better parameters prediction for both instruments.

Parameter	Agilent Exoscan FT-IR	Texas Instruments DLP NIRScan Nano Vis-NIR
P	2640-760cm^{-1}	
K	3800-264; 2055-760cm^{-1}	10672-7246; 6849-6309cm^{-1}
S		
Conductivity	5200-3800; 2055-760cm^{-1}	10672-8620; 7246-6309cm^{-1}
pH ($CaCl_2$)	2640-2055cm^{-1}	8620-6309cm^{-1}
Calcium carbonate	2640-1230cm^{-1}	8620-6849cm^{-1}
Chloride	3800-760cm^{-1}	6849-5878cm^{-1}
Exch Ca	2640-2055cm^{-1}	7246-5878cm^{-1}
Exch K		
Exch Mg	5200-3800; 2640-1230cm^{-1}	10672-7246cm^{-1}
Exch Na		10672-8620; 7246-6849cm^{-1}
ESP	3800-2640cm^{-1}	7246-6309cm^{-1}

Table 3.4.4. PLS results for both instruments and all the parameters analysed with respective number of LVs. LV – latent variables; R^2_C – coefficient of determination of calibration; R^2_{CV} – coefficient of determination of cross-validation; R^2_P – coefficient of determination of prediction. N/A – data not available.

Parameter		Agilent Exoscan FT-IR	Texas Instruments DLP NIRScan Nano Vis-NIR
P	LV	7	N/A
	RMSEC	45.21	N/A
	RMSECV	52.03	N/A
	RMSEP	35.06	N/A
	R^2_C	0.45	N/A
	R^2_{CV}	0.30	N/A
	R^2_P	0.59	N/A
	RER	7.22	N/A
K	LV	7	6
	RMSEC	163.94	223.09
	RMSECV	196.57	237.48
	RMSEP	240.53	309.00
	R^2_C	0.65	0.34
	R^2_{CV}	0.50	0.26
	R^2_P	0.68	0.42
	RER	6.42	4.99
Conductivity	LV	4	4
	RMSEC	0.14	0.15
	RMSECV	0.15	0.16
	RMSEP	0.13	0.12
	R^2_C	0.33	0.21
	R^2_{CV}	0.22	0.15
	R^2_P	0.32	0.50
	RER	6.68	7.74
pH ($CaCl_2$)	LV	5	8
	RMSEC	0.53	0.49
	RMSECV	0.56	0.57
	RMSEP	0.43	0.51
	R^2_C	0.66	0.71
	R^2_{CV}	0.62	0.61
	R^2_P	0.72	0.61
	RER	7.82	6.62
Calcium carbonate	LV	4	3
	RMSEC	0.68	3.47
	RMSECV	1.05	3.80
	RMSEP	0.53	2.79
	R^2_C	0.97	0.30
	R^2_{CV}	0.94	0.18
	R^2_P	0.99	0.38
	RER	24.98	4.70

Table 3.4.4. **(cont.).** PLS results for both instruments and all the parameters analysed with respective number of LVs. LV – latent variables; R^2_C – coefficient of determination of calibration; R^2_{CV} – coefficient of determination of cross-validation; R^2_P – coefficient of determination of prediction. N/A – data not available.

Chloride	LV	6	5
	RMSEC	91.14	121.44
	RMSECV	135.82	135.31
	RMSEP	110.78	110.45
	R^2_C	0.58	0.25
	R^2_{CV}	0.12	0.09
	R^2_P	0.36	0.37
	RER	6.54	6.56
Exch. Ca	LV	9	10
	RMSEC	2.24	3.13
	RMSECV	2.75	4.27
	RMSEP	2.51	2.82
	R^2_C	0.93	0.85
	R^2_{CV}	0.89	0.74
	R^2_P	0.93	0.88
	RER	10.11	9.02
Exch. Mg	LV	3	10
	RMSEC	1.14	1.03
	RMSECV	1.20	1.41
	RMSEP	1.16	1.46
	R^2_C	0.63	0.69
	R^2_{CV}	0.59	0.46
	R^2_P	0.77	0.64
	RER	8.43	6.70
Exch. Na	LV	N/A	2
	RMSEC	N/A	1.34
	RMSECV	N/A	1.37
	RMSEP	N/A	0.75
	R^2_C	N/A	0.02
	R^2_{CV}	N/A	0.00
	R^2_P	N/A	0.28
	RER	N/A	4.21
ESP	LV	5	3
	RMSEC	8.48	9.30
	RMSECV	9.66	9.87
	RMSEP	8.64	9.28
	R^2_C	0.28	0.14
	R^2_{CV}	0.11	0.05
	R^2_P	0.34	0.27
	RER	6.22	5.79

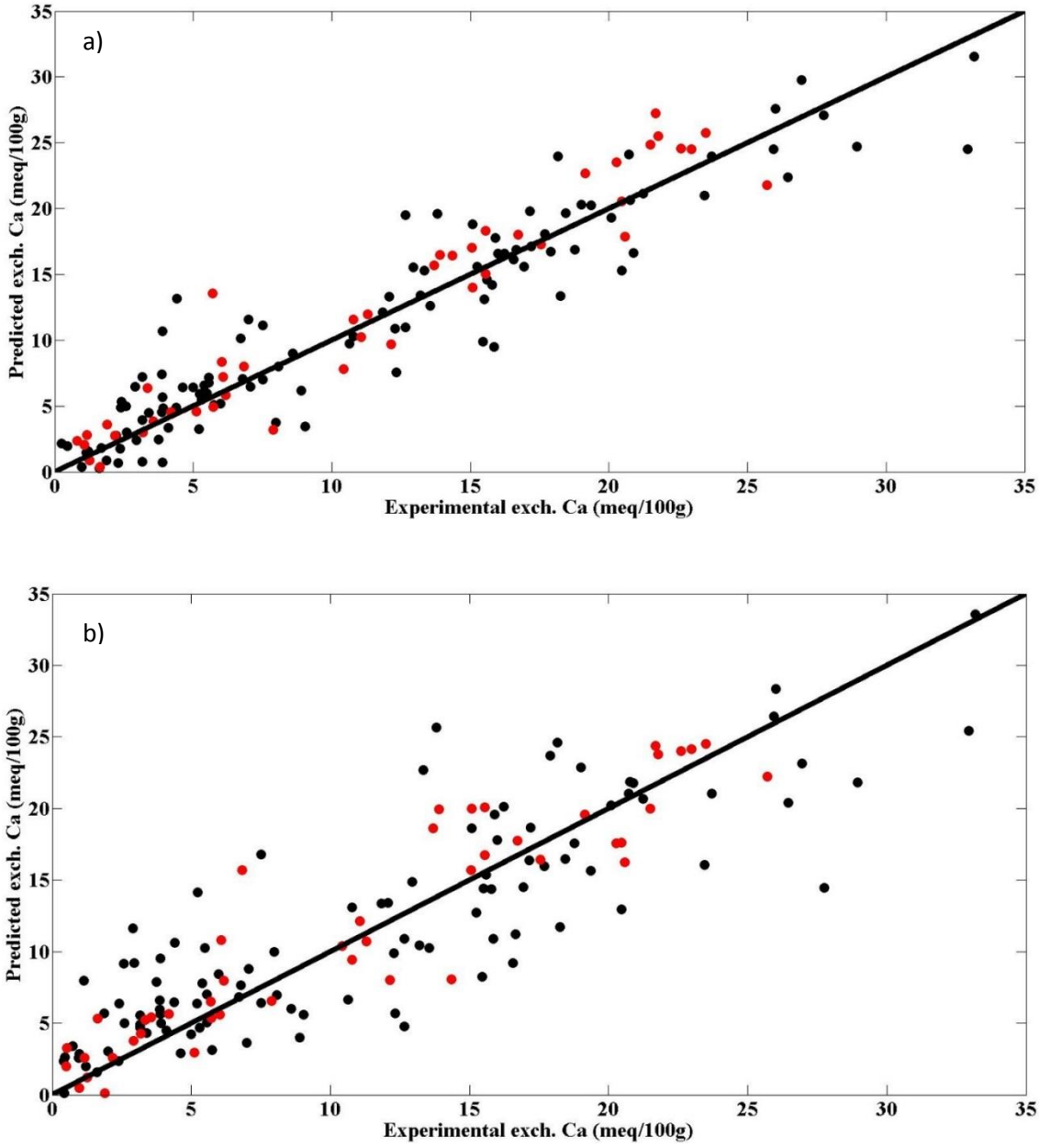


Figure 3.4.3. Comparison between exch. Ca determined with the reference method and the IR based method for the calibration samples (cross-validation predictions) and the independent test samples using specific regions of the spectra (• - cross-validation; • - prediction) for both instruments.
a) Agilent Exoscan FT-IR; b) Texas Instruments DLP NIRScan Nano Vis-NIR.

Soil nutrients such as Ca, Na, K, P, and Mg attract much interest due to their direct relationship to plant status and health, it is therefore natural that proficient means to

monitor their availability should be the constant subject of study. Good correlation results between these soil nutrients and IR spectra in the literature are often rare, but there are exceptions, for instance, Janik et al. [25] revealed good results analysing Ca and Mg using MIR spectroscopy, with R^2_C values of 0.89 and 0.76 respectively. Similar positive results had Shepherd and Walsh [26] with the same parameters, but using NIR spectroscopy (R^2_P of 0.88 and 0.81). Results regarding exch. K were more modest, with Janik et al [25] reporting a R^2_C of 0.33 with MIR and Chang and co-workers [27] an R^2_{CV} of 0.55 using NIR. Even though these results are not very good, they are still better than the ones in this study which, for both instruments, yielded an R^2_P lower than 0.20 (data not shown). Results for P could not be more unequal and different from the ones obtained in this study, with Janik et al [25] obtaining R^2_C values of 0.07 using MIR and Daniel et al. [28], R^2_C values of 0.81 using NIR. Results regarding other parameters such as calcium carbonate and pH ($CaCl_2$) are similar to those found in this study. Janik and colleagues [29] had very good results for calcium carbonate with MIR (R^2_C – 0.95) while Ben-Dor and Banin [30] had good results using NIR (R^2_P – 0.69), which are similar to the ones obtained in this study (even though the R^2_P values for calcium carbonate using NIR were lower). The estimation of pH ($CaCl_2$) had very similar results as those found in the literature. For instance, Janik et al [25] reported an R^2_C of 0.67 using MIR and Chang and co-workers [27] obtained an R^2_{CV} value of 0.56. Some of the cited studies, however, do not use a validation sample set, only give calibration and/or cross-validation results.

It is also quite obvious that the MIR instrument performed better than its NIR counterpart. It had a higher number of good prediction results and the percentage of these predictions was generally higher than the values predicted with the NIR instrument. In general, published papers regard MIR spectroscopy superior to NIR for the estimation of soil parameters. McCarty and team [31] obtained better results for the estimation of parameters such as TN, pH and soil carbon content (total, organic and inorganic carbon) using MIR than when NIR was used. On a similar note, Reeves and colleagues [32] reported worse results when using NIR for the determination of soil carbon pools than when MIR was used. Furthermore, Bellon-Maurel and McBratney [33] concluded in their review about carbon soil stock assessment using NIR and MIR spectroscopy that MIR spectroscopy generally presents lower prediction errors (10 to 40% lower) than NIR. There are, however, studies that support that MIR is not necessarily better than NIR for the estimation of soil properties. For instance, Ludwig et al. [34] found when analysing the composition of organic matter in soil and litter that the predictions using MIR were generally not superior to those of NIR. Similarly, Michel and co-workers [35] estimated the content of carbon and nitrogen in artificial soil mixtures and concluded that both MIR and

NIR are equally able to predict the contents of these analytes. Aranda et al. [36] used MIR and NIR spectroscopy to characterise semi-arid Mediterranean soils and determined that these soils were clearly identified when either MIR or NIR were used, but that the best results were obtained when combining both techniques.

3.4.3.3. Soils discrimination

The performance of both instruments was also assessed in terms of the discrimination between different soil types. This was performed using regression analyses based on PLS-DA models. Firstly, a spectral window selection was performed according to the same procedure used for PLS analysis. Models were calibrated for each instrument separately considering all soils in all sampling points. Only the calibration data were used, and the best models are reported according to the minimum RMSECV (tables 3.4.5 and 3.4.6). Overall, discrimination rates of 74.9% and 67.2% were obtained for the MIR and NIR instruments respectively, indicating that both instruments are well capable of discriminating different soil types (but MIR spectroscopy yielded better results). It was found that the best spectral regions for soil discrimination with the MIR instrument were 5200-2640cm^{-1} and 2055-760cm^{-1} and that the best pre-processing method was Savitzky–Golay filter (15-point filter size, second-order polynomial, and first-order derivative) followed by SNV. These regions are characterised by OH absorption peaks and soil components such as quartz, kaolinite, smectite and calcite [5]. Regarding the NIR instrument, the spectra was just mean centred and the spectral region yielding better results was 7246-6849cm^{-1}. This zone corresponds to CH_3, CH_2, CH, ROH and ArOH absorption bands (third overtone region) and is related to minerals such as clay [37].

Table 3.4.5. Confusion matrix for the soil discrimination model using the Agilent Exoscan FT-IR (74.9% of correct predictions using 15 latent variables).

Predicted soil type[1]	Actual soil type						
	1	2	3	4	5	6	7
1	**3.8**	0.0	0.4	1.9	0.0	1.0	0.6
2	0.0	**2.1**	0.0	1.5	0.0	0.0	0.3
3	0.2	0.1	**13.0**	0.1	0.4	0.5	1.2
4	0.9	0.9	0.4	**22.9**	3.0	2.1	0.5
5	0.0	0.0	0.3	2.3	**6.6**	0.0	0.4
6	0.3	0.9	0.9	0.4	0.5	**21.5**	0.5
7	0.3	0.0	1.3	0.5	0.7	0.0	**5.0**

[1] 1 – Calcareous soils; 2 – Shallow soils on concrete or limestone; 3 – Gradational soils with highly calcareous lower subsoil; 4 – Hard red-brown texture contrast soils with alkaline subsoil; 5 – Deep loamy texture contrast soils with brown or dark subsoil; 6 – Sand over clay soils; 7 – Deep uniform to gradational soils.

There are not many studies performing soil discrimination using IR spectroscopy, but results by Lopo and colleagues [9] in the discrimination of different soil types, revealed similar percentages of correct predictions (around 70%) and also that one of the best spectral regions for soil discrimination was 7300-6700cm^{-1} which is in accordance with the results found in this study.

3.4.4. Conclusions

In this study, two IR instruments, one MIR and one NIR were used to predict twelve soil properties from soil samples collected in several Australian vineyards and to discriminate the existing soil types. Chemometric analysis using PLS revealed that for each parameter and instrument, there was a specific pre-processing method. Parameters such as S, exch. K and exch. Na had results of R^2_P below 0.20 and were not considered when the MIR instrument was used, the same happening for P, S and exch. K with the NIR instrument. Very good correct predictions were obtained for pH, exch Ca and exch. Mg using both instruments. Regarding the remaining parameters, there was no particular trend, calcium carbonate had a very good predictive result with the MIR instrument whereas with the NIR instrument the result was poor. A good result was obtained for K (R^2_P = 0.68) and an

average one for P when using MIR. The results for these last two parameters with the NIR spectrometer were much different, with K having below average results and P yielding an R^2_P much lower (data not shown). Also, the MIR instrument presented better results than the NIR instrument in almost every parameter.

Soil classification was also attempted using both spectrometers with MIR and NIR exhibiting, through PLS-DA, good discriminative capacity for the soil types analysed. The MIR instrument revealed 74.9% of soil correct predictions while the result with the NIR instrument was slightly lower with 67.2% of correct predictions, but still good.

To the best of our knowledge it is the first time that a single study combines MIR and NIR analysis on the estimation of so many different soil parameters as well as dealing with soil type classification.

The results in this study showed that IR spectroscopy can constitute a viable means for the analysis of several important soil parameters and thus constitute an added value for the monitoring and development of specific agricultural cultivars. Furthermore, it was shown that this technology can also be used for swift soil classification, opening a myriad of possibilities, from the generation of accurate soil maps, to the adjustment of a specific cultivar in a specific soil type depending on the producer aims and objectives.

3.4.5 References

1. P Leone, A., et al., *Prediction of soil properties with PLSR and vis-NIR spectroscopy: application to mediterranean soils from Southern Italy.* Current Analytical Chemistry, 2012. **8**(2): p. 283-299.
2. Soriano-Disla, J.M., et al., *The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties.* Applied Spectroscopy Reviews, 2014. **49**(2): p. 139-186.
3. Vohland, M., et al., *Determination of soil properties with visible to near-and mid-infrared spectroscopy: effects of spectral variable selection.* Geoderma, 2014. **223**: p. 88-96.
4. Waruru, B.K., et al., *Rapid estimation of soil engineering properties using diffuse reflectance near infrared spectroscopy.* biosystems engineering, 2014. **121**: p. 177-185.
5. Rossel, R.V., et al., *Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties.* Geoderma, 2006. **131**(1): p. 59-75.
6. Cozzolino, D., et al., *In situ measurement of soil chemical composition by near-infrared spectroscopy: A tool toward sustainable vineyard management.* Communications in soil science and plant analysis, 2013. **44**(10): p. 1610-1619.
7. Zornoza, R., et al., *Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils.* Soil Biology and Biochemistry, 2008. **40**(7): p. 1923-1930.
8. Ben-Dor, E., D. Heller, and A. Chudnovsky, *A novel method of classifying soil profiles in the field using optical means.* Soil Science Society of America Journal, 2008. **72**(4): p. 1113-1123.
9. Lopo, M., et al., *Classification of Vineyard Soils Using Portable and Benchtop Near-Infrared Spectrometers: A Comparative Study.* Soil Science Society of America Journal, 2016. **80**(3): p. 652-661.
10. Linker, R. *Soil classification via mid-infrared spectroscopy*. in *International Conference on Computer and Computing Technologies in Agriculture*. 2007. Springer.
11. Páscoa, R., et al., *Exploratory study on vineyards soil mapping by visible/near-infrared spectroscopy of grapevine leaves.* Computers and Electronics in Agriculture, 2016. **127**: p. 15-25.

12. Awiti, A.O., et al., *Soil condition classification using infrared spectroscopy: A proposition for assessment of soil condition along a tropical forest-cropland chronosequence.* Geoderma, 2008. **143**(1): p. 73-84.
13. Mouazen, A.M., et al., *Classification of soil texture classes by using soil visual near infrared spectroscopy and factorial discriminant analysis techniques.* Journal of near infrared spectroscopy, 2005. **13**(4): p. 231.
14. Edwards, J. Setting benchmarks and recommendations for management of soil health in Australian viticulture. Final Report to Australian Grape and Wine Authority Project Number DPI1101, July 2014.
15. Rayment, G.E and Lyons, D.J. pH of 1:5 soil/water suspension (method 4A1). In 'Soil Chemical Methods - Australasia'. (CSIRO publishing, Melbourne), 2011b. p. 38-39.
16. Rayment, G.E and Lyons, D.J. pH of 1:5 soil/0.01M calcium chloride extract – following method 4A1 (method 4B4). In 'Soil Chemical Methods - Australasia'. (CSIRO publishing, Melbourne), 2011c. p. 41.
17. Rayment, G.E and Lyons, D.J. Electrical conductivity (EC) of 1:5 soil/ water extract (method 3A1). In 'Soil Chemical Methods - Australasia'. (CSIRO publishing, Melbourne), 2011d. p. 20-22.
18. Rayment, G.E and Lyons, D.J. (2011e). Exchangeable bases and CEC (method 15D2). In 'Soil Chemical Methods - Australasia'. (CSIRO publishing, Melbourne), 2011e. p. 318-321.
19. Naes, T., et al., *Interpreting PCR and PLS solutions*, in *A User-Friendly Guide to Multivariate Calibration and Classification*. 2004, NIR Publications: Chichester, UK. p. 39-54.
20. Barker, M. and W. Rayens, *Partial least squares for discrimination.* Journal of Chemometrics, 2003. **17**(3): p. 166-173.
21. Geladi, P. and B.R. Kowalski, *PARTIAL LEAST-SQUARES REGRESSION - A TUTORIAL.* Analytica Chimica Acta, 1986. **185**: p. 1-17.
22. Savitzky, A. and M.J. Golay, *Smoothing and differentiation of data by simplified least squares procedures.* Analytical chemistry, 1964. **36**(8): p. 1627-1639.
23. Preisner, O., et al., *Application of Fourier Transform Infrared Spectroscopy and Chemometrics for Differentiation of Salmonella enterica Serovar Enteritidis Phage Types.* Applied and Environmental Microbiology, 2010. **76**(11): p. 3538-3544.
24. Borrelli, P., et al., *Assessment of the cover changes and the soil loss potential in European forestland: First approach to derive indicators to capture the ecological*

impacts on soil-related forest ecosystems. Ecological Indicators, 2016. **60**: p. 1208-1220.

25. Janik, L.J., R.H. Merry, and J. Skjemstad, *Can mid infrared diffuse reflectance analysis replace soil extractions?* Animal Production Science, 1998. **38**(7): p. 681-696.
26. Shepherd, K.D. and M.G. Walsh, *Development of reflectance spectral libraries for characterization of soil properties.* Soil science society of America journal, 2002. **66**(3): p. 988-998.
27. Chang, C.W., et al., *Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties.* Soil Science Society of America Journal, 2001. **65**(2): p. 480-490.
28. Daniel, K., N. Tripathi, and K. Honda, *Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand).* Soil Research, 2003. **41**(1): p. 47-59.
29. Janik, L.J. and J. Skjemstad, *Characterization and analysis of soils using mid-infrared partial least-squares. 2. Correlations with some laboratory data.* Soil Research, 1995. **33**(4): p. 637-650.
30. Ben-Dor, E. and A. Banin, *Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties.* Soil Science Society of America Journal, 1995. **59**(2): p. 364-372.
31. McCarty, G., et al., *Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement.* Soil Science Society of America Journal, 2002. **66**(2): p. 640-646.
32. Reeves III, J.B., et al., *Can Near or Mid-Infrared Diffuse Reflectance Spectroscopy Be Used to Determine Soil Carbon Pools?* Communications in Soil Science and Plant Analysis, 2006. **37**(15-20): p. 2307-2325.
33. Bellon-Maurel, V. and A. McBratney, *Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils–Critical review and research perspectives.* Soil Biology and Biochemistry, 2011. **43**(7): p. 1398-1410.
34. Ludwig, B., et al., *Use of mid-infrared spectroscopy in the diffuse-reflectance mode for the prediction of the composition of organic matter in soil and litter.* Journal of Plant Nutrition and Soil Science, 2008. **171**(3): p. 384-391.
35. Michel, K., et al., *Use of near-and mid-infrared spectroscopy to distinguish carbon and nitrogen originating from char and forest-floor material in soils.* Journal of Plant Nutrition and Soil Science, 2009. **172**(1): p. 63-70.

36. Aranda, V., et al., *Agro-environmental characterization of semi-arid Mediterranean soils using NIR reflection and mid-IR-attenuated total reflection spectroscopies.* Vibrational Spectroscopy, 2014. **74**: p. 88-97.
37. Bishop, J.L., et al., *Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite-serpentines, chlorites and micas.* Clay Minerals, 2008. **43**(1): p. 35-54.

The following work was adapted from the article:

Páscoa, R. N. M. J., Lopo, M., dos Santos, C. T., Graça, A. R., & Lopes, J. A. (2016). Exploratory study on vineyards soil mapping by visible/near-infrared spectroscopy of grapevine leaves. *Computers and Electronics in Agriculture*, *127*, 15-25

3.5. Exploratory study on vineyards soil mapping by visible/near-infrared spectroscopy of grapevine leaves

Abstract: This work demonstrates the possibility of discriminating vineyard soils through the non-destructive and in-situ visible/near infrared monitoring of leaves. A portable Vis/NIR spectrometer was applied for monitoring *in-situ Vitis vinifera* leaves in vineyards of two wine regions in Portugal in the maturation period. Leaves reflectance spectra of different grapevine varieties planted in different vineyard locations (distinct soil taxonomic types) were analysed by principal component analysis and partial least squares discriminant analysis. Soil discriminant models based on leaves Vis/NIR spectra yielded for both vineyards approximately 95% correct soil taxonomic predictions. This methodology was then applied to monitor all plants within a 0.3ha vineyard block in the Dão vineyard resulting in a highly detailed soil taxonomic map built exclusively from leaves Vis/NIR spectra. A comparison with the existing soil map proved that the NIR spectrocopy based estimation was not only extremely reliable and accurate but also a lot more detailed than pedological soil maps. Even though further studies are needed, namely in different maturation stages and other geographical regions, to ensure reliability of this technique, results in this work showed that it can be used as an additional auxiliary tool for obtaining vineyard soil maps. Its main advantages over pedological reference procedures are speed and cost efficiency analysis.

3.5.1. Introduction

The emergence of precision viticulture is based on the need for a more detailed monitoring and control of the quantity and quality of the production in a vineyard. Precision viticulture involves methodologies able of site-specific monitoring and management [1] and is based on the assumption that productivity within vineyard blocks can be highly variable. Soils play a major role in this variability [2] together with other factors such as latitude, topography, climate (temperature, frost, humidity, rainfall, sun exposure), diseases (powdery mildew, downey mildew, phylloxera), grapevine variety and viticulture practices [3]. Vineyard soil characterization is of the utmost importance since soils play a major role in vine behaviour, grape quality and ultimately wine sensory properties [4]. Since vineyard soils have a strong influence over grapevine leaves, knowing their physical characteristics and chemical composition is very important for an efficient management [5]. This characterization is an interesting support tool for the cultivation/replantation of new vineyards with the objective of increasing the quality and yield of produced grapes. However, the actual situation shows that in most vineyards around the world, winemakers still do not take into account this particular knowledge due the lack of customary resentment that new technologies normally encounter in a traditional industry. Due to lack of information, fear of investment in new technologies and deeply rooted habits, vineyards are usually divided into blocks, each block having typically a single grape variety. Additionally, in many situations, soils with different characteristics may coexist in the same block. With this type of configuration, the quality of grapes harvested from the same block will not be the same. Since each vine variety has specific growth requirements, the same variety planted in different soils will grow differently. This is demonstrated by grapes analytical parameters and yield. Consequently, there is a demand for the development of faster, more reliable and cost effective techniques able to characterize vineyard soils that can be used efficiently as a viticulture support tool. Near infrared spectroscopy (NIRS) is a candidate for this purpose since it presents all the aforementioned characteristics. Near infrared radiation is influenced by combinations and overtones of fundamental vibrational transitions, essentially of C-H, N-H, O-H and S-H bonds [6]. A major advantage of NIRS is the possibility to obtain spectral fingerprints without sample processing. Another important feature of this technique is that it can be incorporated in remote sensing devices or used to obtain multispectral images [7]. Spectral measurements can be made in plant leaves, either at-line in laboratory or in-situ at the vineyard, or by remote devices (remote sensing).

The determination of chemical elements of vineyard soils by NIRS has been reported in several recent studies. Salazar et al. [8] evaluated cobalt contamination in vineyard soils with NIRS demonstrating the feasibility of the method for detecting concentrations over 1 g-Co kg^{-1} of soil sample. Cozzolino and team [9] measured total nitrogen, organic carbon, potassium, sulphur, phosphorous, pH and electric conductivity *in-situ* with a portable NIRS instrument in three different wine regions in Australia. The method was able to discriminate soils and a good linearity for all elements/properties was obtained except for potassium and organic carbon. The first study attempting to use Vis/NIRS directly in grapevine leaves was performed by Steele and colleagues[10] with the objective of estimating the anthocyanin content. The best results were obtained using the green and red regions for the anthocyanin reflectance index and green, red and NIRS regions for the modified anthocyanin reflectance index. In 2012, two commercial optical devices (GreenSeeker RT100 and Crop Circle) were tested for detecting different levels of grapevine downy mildew symptoms in Cabernet Franc leaves. However, this method was not applied *in-situ* [11]. Off-line collected hyperspectral images of grapevine leaves in the range of 380 to 1028 nm were used to classify Tempranillo, Grenache and Cabernet Sauvignon varieties [12]. NIRS as a remote sensing tool has been extensively used for the characterization of spatial distribution of vine vegetation, often through the estimation of the NDVI index [13-15]. Mazzetto et al. [1] tested a remote sensing system for monitoring disease (*Plasmoparaviticola*) spreading in vineyards. This system included two GreenSeeker RT100 sensors, a DGPS receiver and ultrasonic sensors. The obtained results were comparable to the real vine phytosanitary status [1]. The determination of iron deficiency chlorosis, through carotene and anthocyanin pigments content estimated with hyperspectral imaging data, allowed the design of maps pinpointing specific harvesting regions that have the desired wine properties [16]. Remote sensing was used to estimate leaves anthocyanin content using Vis/NIR hyperspectral imaging [17]. More recently, Ciraolo and co-workers [18] used multispectral images from different spectral regions (visible and NIR regions) to map the evapotranspiration of leaves on a Sicilian (Italy) vineyard. Saiz-Rubio and Rovira-Mas [19] mounted a UV/VIS/NIR camera on a conventional tractor to estimate vine vigour achieving the best results using the NIR spectral region. On a similar note, Kodaira and Shibusawa [20] used VIS-NIR reflectance spectroscopy to estimate soil properties (including moisture, organic matter, pH, electrical conductivity, among others). Stamatiadis et al. [21] also used device to estimate biomass production, pruning weight, yield, Brix, phenolic and anthocyanin contents.

This work intended to investigate the potential of Vis/NIR spectroscopy to characterize soils based on direct in-situ measurements of vine leaves, enabling the

estimation of soil variability maps with a superior resolution when compared with currently existing methods. This approach involves only in-situ and non-destructive analyses. For this task, a portable Vis/NIR spectrometer spanning a 350-2500 nm range was used to monitor several blocks in two vineyards of two Portuguese wine regions (Dão and Alentejo). This approach is intended to provide another decision support tool for winemakers, enabling delimitation of harvesting zones and planning of grapevines plantation or replantation. To the best of our knowledge this is the first time Vis/NIR spectroscopy is used on vineyard leaves with the purpose of mapping soil taxonomic types.

3.5.2. Material and Methods

3.5.2.1. Vineyards monitoring

Two vineyards, property of SOGRAPE VINHOS SA, in two different wine regions in Portugal were selected: *Quinta dos Carvalhais* (Mangualde, 40.556721-7.787247) in the Dão Wine Region (centre of Portugal) and *Herdade do Peso* (Vidigueira, 38.141579-7.677813) in the Alentejo Wine Region (south of Portugal) (Fig. 3.5.1). Soils in these vineyards were previously characterized through pedological methods and named according to the international soil classification system (*IUSS*, 2014). Table 3.5.1 compiles the existing taxonomic types of soils in both vineyards together with soil texture information for the first two horizons (*IUSS*, 2014). Vineyards are divided in numbered blocks, each containing a single *Vitis vinifera* cultivar. Several monitoring spots were identified with the objective of analysing leaves spectra variability within the same grape variety, planted on different soil taxonomic types. The rationale for vineyard spots selection was based on the existence in the vineyards of grapes of the same variety grown on soils with distinct characteristics (4 varieties in *Herdade do Peso* and 5 varieties in *Quinta dos Carvalhais*). For these varieties, sampling spots were defined in locations which have only one soil taxonomic type (Fig. 3.5.1). Table 3.5.2 summarizes the characteristics of the selected monitoring spots: 14 spots in *Quinta dos Carvalhais* and 15 spots in *Herdade do Peso*. At each defined sampling spot, a total of twenty leaves in five different plants were monitored (four monitored leaves per plant). Monitoring in *Herdade do Peso* (with an average temperature of 15-17.5°C, precipitation 500-800 mm and no significant altitude shift) and *Quinta dos Carvalhais* (with an average temperature of 14-16°C, precipitation 1100-1600 mm and no significant altitude shift) was made on 2012-07-29 and on 2012-08-31, respectively (ripening stage shortly after the veraison period). A

total of 300 and 280 spectra were collected in *Herdade do Peso* and *Quinta dos Carvalhais*, respectively (Table 3.5.2). Additionally, three leaves of all plants of a 0.3ha block in *Quinta dos Carvalhais* were monitored (total of 1200 spectra). The goal was the validation of this methodology, allowing the generation of a detailed soil map of that block, based exclusively on Vis/NIR spectral measurements.

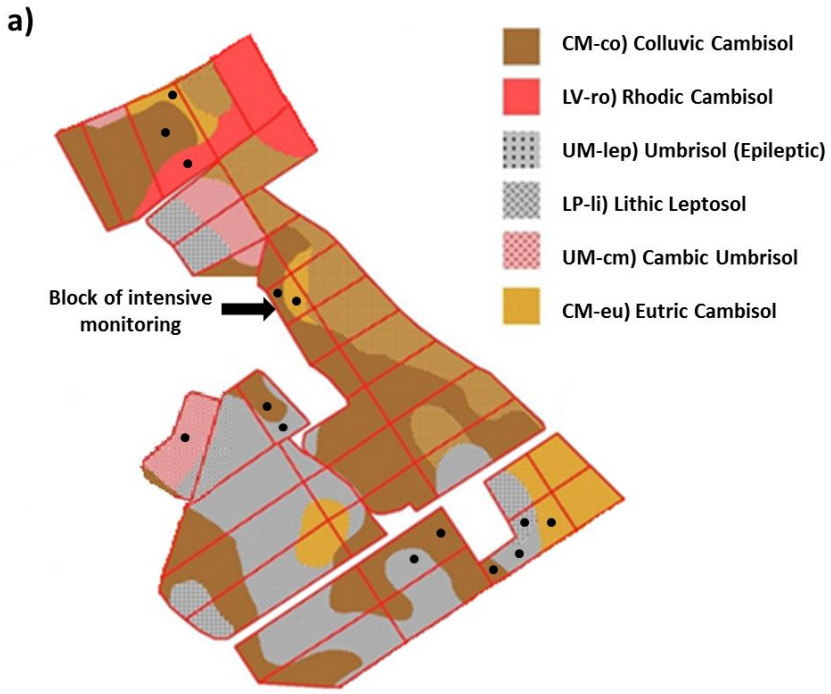


Figure 3.5.1. Maps of the monitored vineyards with the representation of soil types estimated through the application of pedology methods (a-*Quinta dos Carvalhais*, b-*Herdade do Peso*).

b)

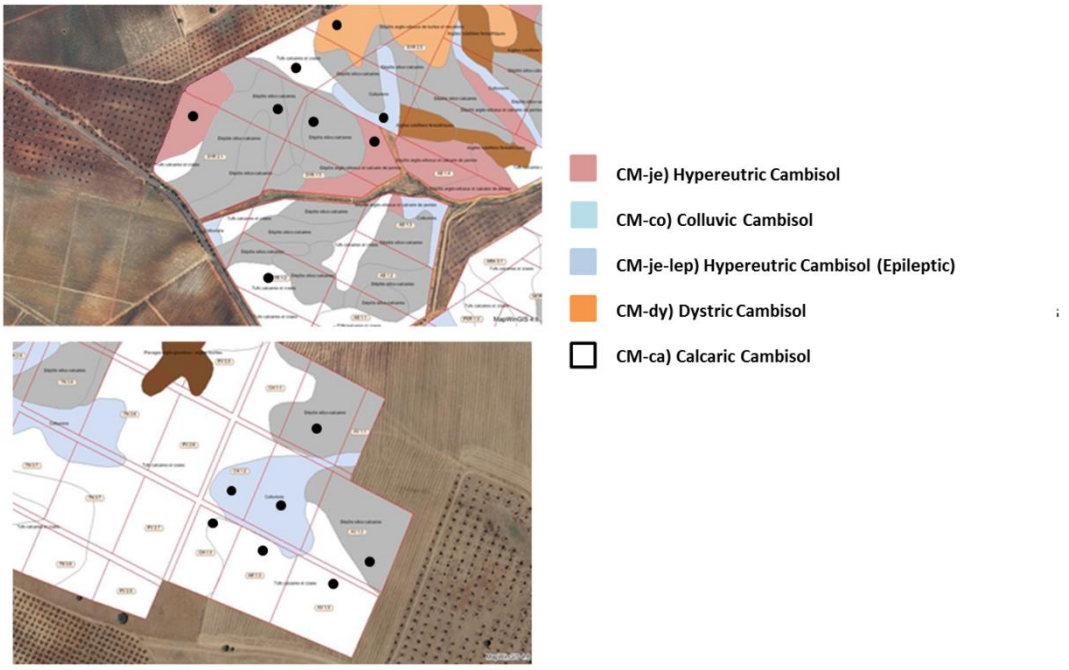


Fig. 3.5.1. **(cont.)**. Maps of the monitored vineyards with the representation of soil types estimated through the application of pedology methods (a-*Quinta dos Carvalhais*, b-*Herdade do Peso*).

Table 3.5.1. Vineyards soil types and textural analysis.

Vineyard	Soil type			
	FAO-WRBSR	FAO-WRBSR-2014 Code	Soil texture	
			Horizon 1 0-30 cm	Horizon 2 30-80 cm
Herdade do Peso	Calcaric Cambisol	CM-ca	Loam	Silt loam
	Hypereutric Cambisol (Epileptic)	CM-je-lep	Sandy clay loam	Sandy clay loam
	Hypereutric Cambisol	CM-je	Clay	clay
	Colluvic Cambisol	CM-co	Sandy clay loam	Sandy clay loam
	Hypereutric Cambisol	CM-je	Clay	Silty clay loam
	Dystric Cambisol	CM-dy	Sandy clay loam	Sandy clay loam
Quinta dos Carvalhais	Colluvic Cambisol	CM-co	Sandy loam	Sandy loam
	Eutric Cambisol	CM-eu	Sandy loam	Sand
	Lithic Leptosol	LP-li	Sandy loam	Sandy loam
	Umbrisol (Epileptic)	UM-lep	Sandy loam	Loamy sand
	Cambic Umbrisol	UM-cm	Sandy loam	Sandy loam
	Rhodic Luvisol	LV-ro	Sandy loam	Sandy loam

Table 3.5.2. Summary of varieties, soil types and monitoring spots in the two vineyards.

Varieties	Soil type[1]	Number of monitoring spots	Number of spectra	Varieties	Soil type[1]	Number of monitoring spots	Number of spectra
Vineyard 1 *(Herdade do Peso)*				Vineyard 2 *(Quinta dos Carvalhais)*			
Syrah	CM-je-lep	2	40	*Verdelho*	UM-lep	1	20
	CM-ca	2	40		LP-li	1	20
	CM-co	1	20		CM-eu	1	20
	CM-je	2	40		CM-co	1	20
	CM-dy	1	20	*Encruzado*	CM-eu	1	20
	---	---	---		CM-co	1	20
Arinto	CM-je-lep	1	20	*Jaen*	UM-cm	1	20
	CM-ca	1	20		CM-eu	1	20
	CM-co	1	20		CM-co	1	20
Antão Vaz	CM-je-lep	1	20	*Alvarinho*	UM-lep	1	20
	CM-ca	1	20		CM-co	1	20
Chardonnay	CM-ca	1	20	*Semillon*	UM-lep	1	20
	CM-co	1	20		CM-co	1	20
	---	---	---		LV-ro	1	20
Total	5	15	300	Total	6	14	280

[1] CM-ca (Calcaric cambisol); CM-je-lep (Hypereutric Cambisol Epileptic); CM-je (Hepereutric Cambisol); CM-co (Colluvic Cambisol); CM-dy (Dystric Cambisol); CM-eu (Eutric Cambisol); LP-li (Lithic Leptosol); UM-lep (Umbrisol Epileptic); UM-cm (Cambic Umbrisol); LV-ro (Rhodic Luvisol).

3.5.2.2. Spectral acquisition

Near infrared leaves spectra (Fig. 3.5.2) were acquired using a FieldSpec 4 Wide-Res (ASD Inc, Boulder, CO) in diffuse reflectance mode spanning the 350-2500 nm range. The spectral resolution and sampling interval were 3 and 1.4 nm for the 350-1050 nm spectral range and 10 and 2 nm for the 1000-2500 nm range, respectively. The system incorporated a reflectance contact probe (Hi-Brite, ASD Inc., Boulder, CO) with a measurement surface area equivalent to a 10 mm diameter circle, enclosing a halogenous light source (Fig. 3.5.2). A background was taken every hour with a Spectralon® disk (ASD Inc., Boulder, CO). Leaves were measured in diffuse reflectance mode directly in the plant (*in-situ*) with no cleaning process involved. Average size leaves located at one shoot of the first arm were selected. Spectral measurements were performed approximately at the leaf centre between 7am and 11am. All measurements were made in the same day at each vineyard. All spectral measurements were performed in triplicate and the average considered for further processing. Leaves spectra were pre-processed with a Savitzky–Golay filter (15-points filter size, second order polynomial, and second-order derivative) [22].

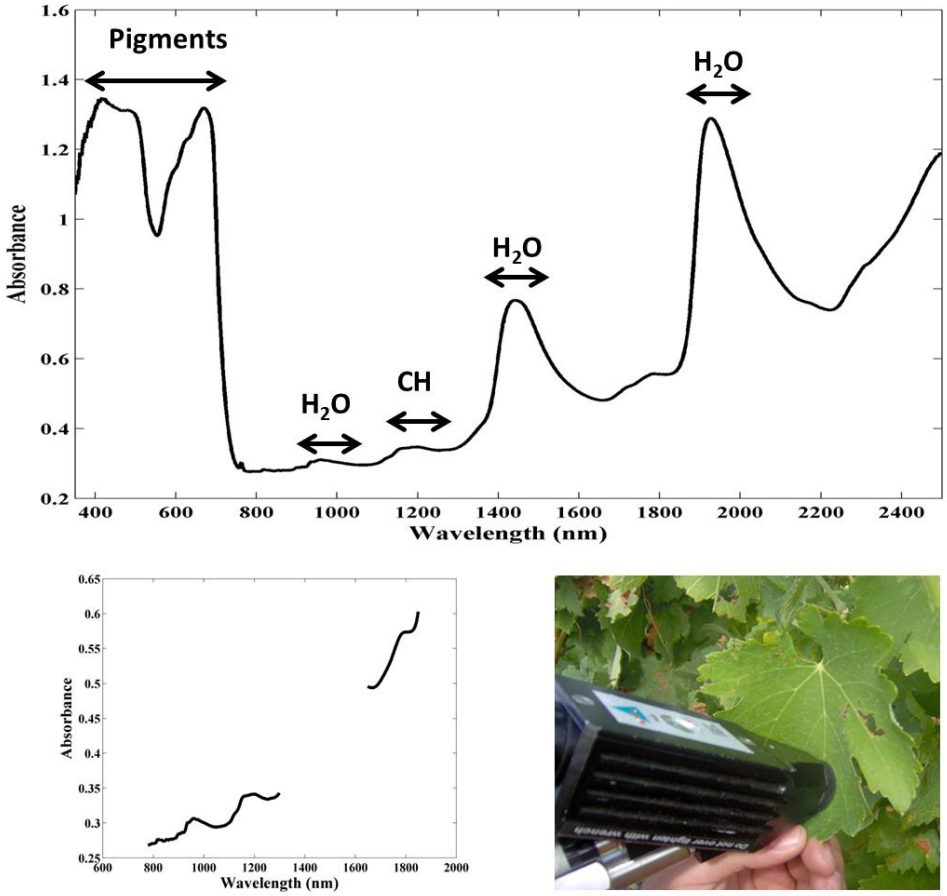


Figure 3.5.2. Vis/NIR spectrum of a grapevine leaf with indication of the major bands (at top), Vis/NIR spectrum of the zone used in the models (at bottom left) and picture of the probe operation in contact with the leaf acquiring a spectrum (bottom right).

3.5.2.3. Vis/NIR spectra modelling

Principal component analysis (PCA) [23] was used to perform exploratory data analysis on leaves spectra and to detect outliers. Partial least squares discriminant analysis (PLS-DA) [24] was the selected method to develop soils taxonomic type discrimination models. For PLS-DA models, the available data were divided to form a calibration (70% of the available samples) and a test (remaining 30%) set. The division was made considering blocks of 15 contiguous spectra in order to avoid overfitting problems (e.g., preventing the existence of spectra collected from nearby leaves both in the calibration and test sets) [25]. Sample blocks division was made randomly but ensuring that the same proportion between soils is present in calibration and test sets to avoid unbalanced classes in both sets [26]. The optimal number of latent variables (LVs) was

estimated by leave-one-block-out cross-validation (contiguous blocks of 15 samples) using only the calibration set [23].Two different models will be developed for the two vineyards. To optimize the NIR wavelengths to consider in PLS-DA models, a strategy for wavelength selection based on an exhaustive search over the entire spectral range available (350-2500 nm) was adopted (Fig. 3.5.3). All contiguous wavelength windows with a resolution of 12 nm were tested. For each spectral window, a PLS-DA model was cross-validated using the above mentioned cross-validation strategy (leave-one-block-out) considering only the calibration data. Cross-validation results are provided as a surface chart where each pixel corresponds to the cross-validated % of correct predictions for the corresponding wavelength window: y-axis is the start wavelength and x-axis is the final wavelength. The spectral zones with the highest % of correct predictions are indicated by blue regions whereas red regions indicate the poorest models (Fig. 3.5.3). Note that by combining multiple windows, model predictions might be further improved but the goal here was to identify specific contiguous wavelength windows for soil differentiation. After optimizing the wavelength region, models for the two vineyards are calibrated using the selected wavelength region and corresponding optimal number of latent variables. The test set is then projected onto the optimized PLS-DA model and soil predictions are expressed as confusion matrices. Confusion matrices compare each known soil with the corresponding NIRS prediction and entries are expressed as percentages. PLS-DA loadings were also analysed in order to understand which specific wavenumbers are more important for soils discrimination. Before applying PCA and PLS-DA the spectral sets were mean centred.

All chemometric methods and spectra processing were performed using Matlab version 7.9 (MathWorks, Natick, MA) and the PLS Toolbox version 5.5.1 (Eigenvector Research Inc., Wenatchee, WA).

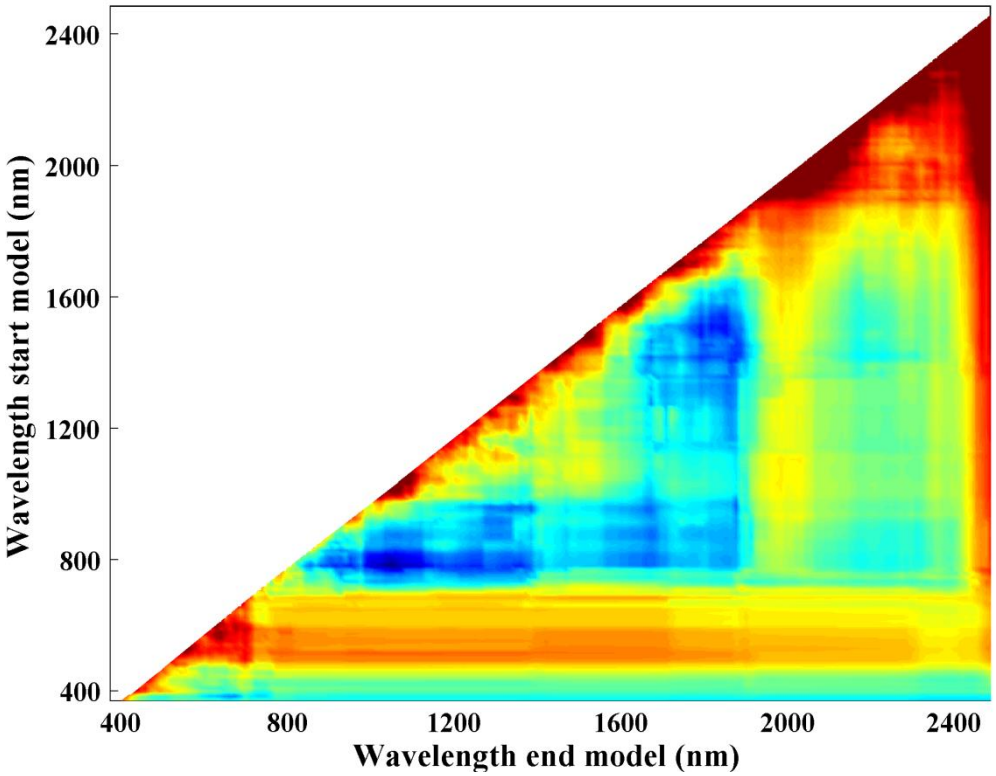


Figure 3.5.3. RMSECV obtained for PLS-DA models using different spectral windows for all varieties considered in this work. The corresponding % of correct predictions goes from 5.9% (red color) to 83.6% (blue color).

3.5.3. Results and Discussion

The obtained leaves Vis/NIR spectra show the typical patterns found previously (Porra et al. 1989) (Fig. 3.5.2). Chlorophyll a (around 419 and 663 nm) and b (around 454 and 646 nm) and carotenoids (around 450 nm) pigments are responsible for peaks in the visible region [27, 28]. Other major bands in the NIR region are due to O-H (e.g., water) and C-H bonds (combinations and overtones).

3.5.3.1. Exploratory data analysis of leaves spectra

To analyse the NIR spectral variability obtained for leaves of the same variety grown in soils with different properties, PCA models considering each grapevine variety individually were developed. A total of 9 PCA models were developed: four models for *Herdade do Peso* (four grapevine varieties) and five models for *Quinta dos Carvalhais* (five grapevine varieties). The score plots of the first two principal components (first principal component against second principal component) of all the PCA models (each

vine variety individually) are shown in Fig. 3.5.4. For most varieties, a simple scatter plot between the first and second principal components is enough to verify that samples cluster according to the taxonomic type of soil. This is less evident for example for Syrah (Fig. 3.5.4a) due to the existence of five different soils. Differentiation in this case was not possible considering only two components. A detailed case-by-case analysis was made considering combinations of other components but no further discrimination was achieved and therefore we kept this analysis restricted to the two major components. It should also be mentioned that PCA maximized the variance criterion and no information about soils is provided. Nevertheless, results appear promising indicating that there are spectral features of leaves that can be correlated with the taxonomic type of soil. It is known that leaves metabolites vary significantly over the vegetative cycle, especially during the ripening period [29]. Thus, measurements should be made on a relatively short period of time and with the same weather conditions for spectra to be comparable (ideally performing measurements on the same day). Further studies are needed to identify which period of the year would be the most suitable for leaves spectral analysis used in soil taxonomic type discrimination.

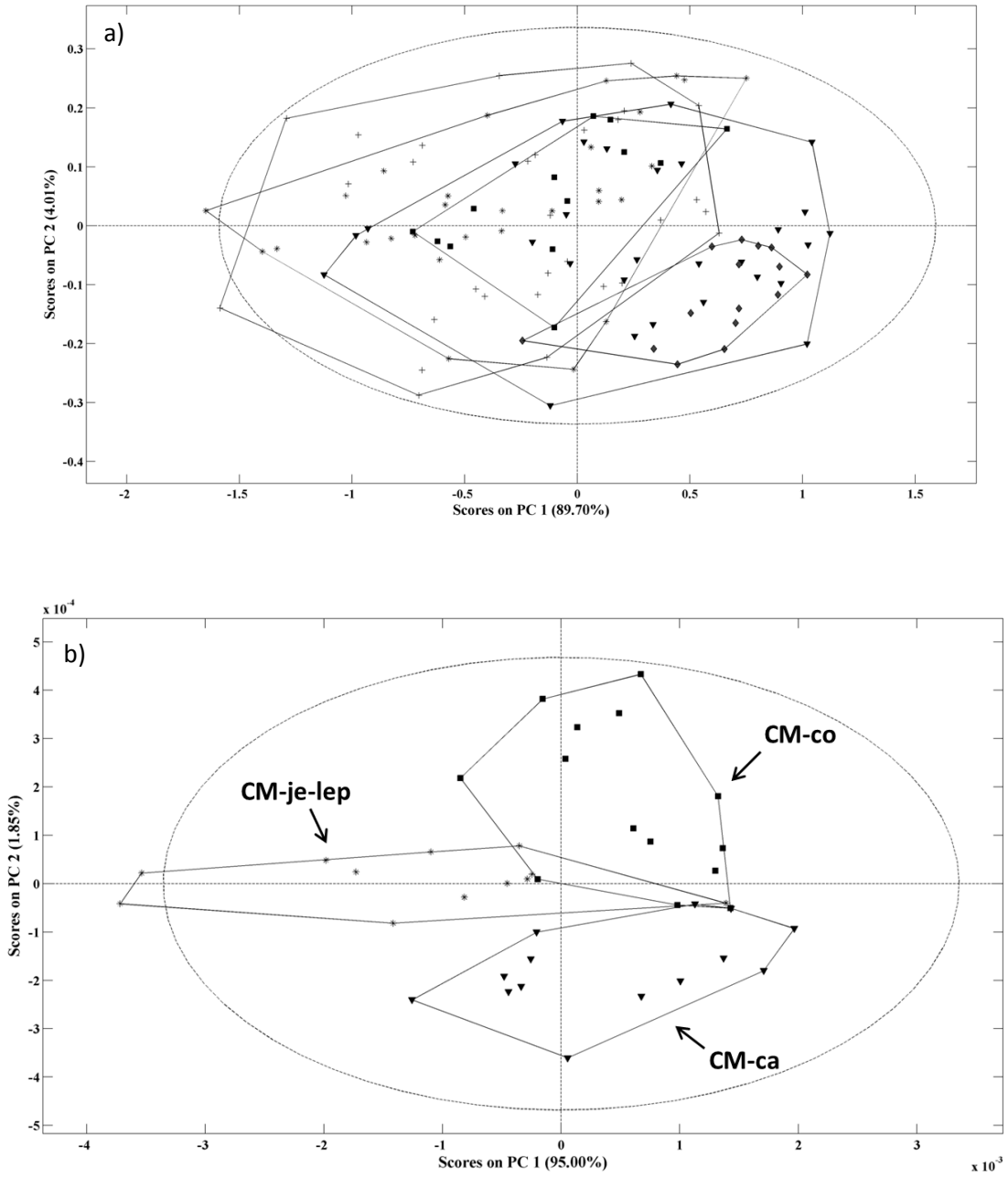


Figure 3.5.4. Score plots obtained by PCA of grapevine leaves spectra in the 800 to 1375 nm and 1475 to 1850 nm regions, considering different varieties (a:*syrah*, b:*arinto*,c:*antão vaz*, d:*chardonnay*, e:*verdelho*, f:*encruzado*, g:*jaen*, h:*alvarinho*, i:*semillon*). Markers are defined accordingly to the soil type (▼ CM-ca: Calcaric cambisol; ❋ CM-je-lep: Hypereutric Cambisol Epileptic; + CM-je: Hepereutric Cambisol; ■ CM-co: Colluvic Cambisol; ◆ CM-dy: Dystric Cambisol; × CM-co: Colluvic Cambisol; ◇ CM-eu: Eutric Cambisol; □ LP-li: Lithic Leptosol; ● UM-lep: Umbrisol Epileptic; ○ UM-cm: Cambic Umbrisol; ● LV-ro: Rhodic Luvisol).

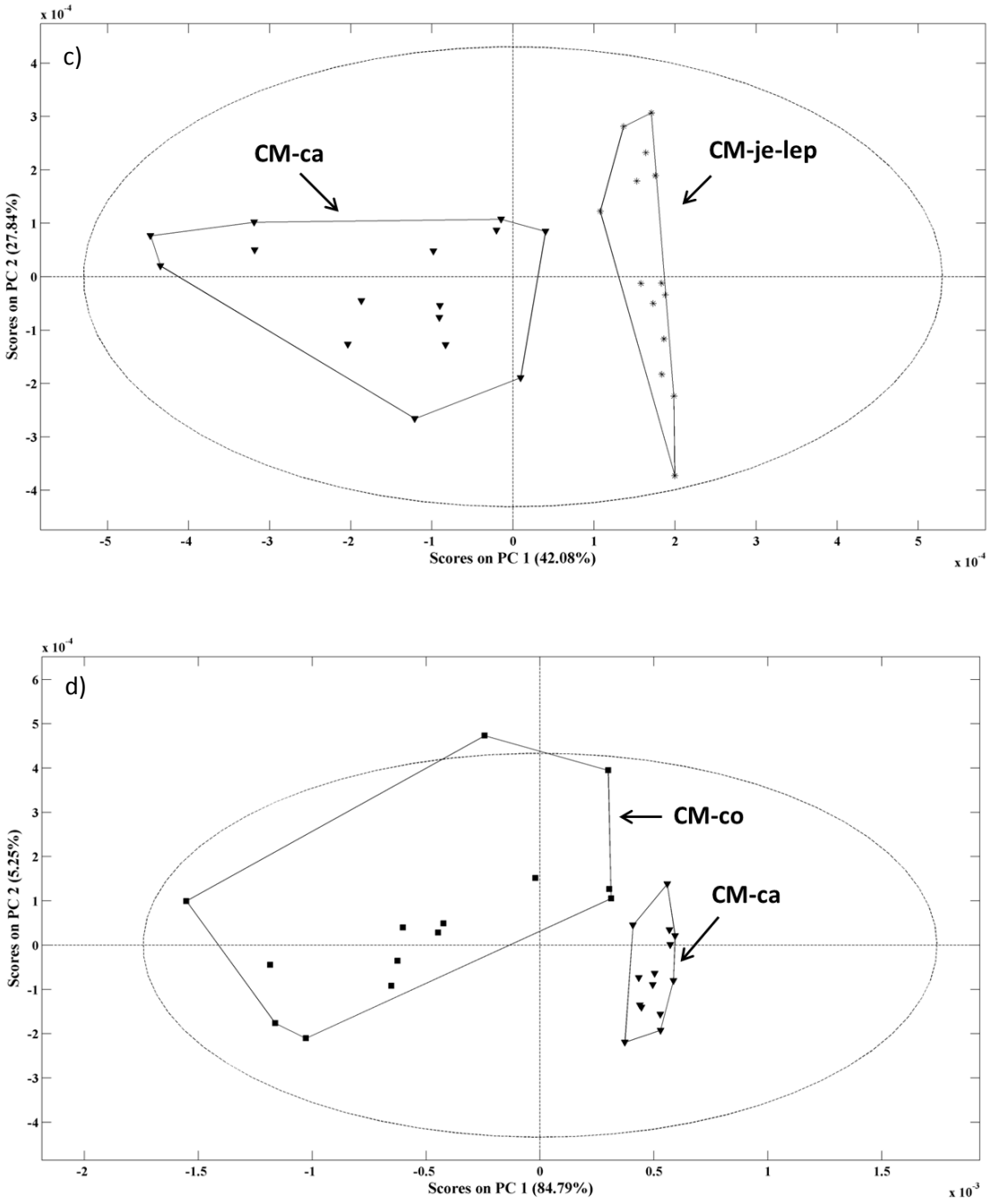


Figure 3.5.4. **(cont.).** Score plots obtained by PCA of grapevine leaves spectra in the 800 to 1375 nm and 1475 to 1850 nm regions, considering different varieties (a:*syrah*, b:*arinto*,c:*antão vaz*, d:*chardonnay*, e:*verdelho*, f:*encruzado*, g:*jaen*, h:*alvarinho*, i:*semillon*). Markers are defined accordingly to the soil type (▼ CM-ca: Calcaric cambisol; ✱ CM-je-lep: Hypereutric Cambisol Epileptic; + CM-je: Hepereutric Cambisol; ■ CM-co: Colluvic Cambisol; ◆ CM-dy: Dystric Cambisol; × CM-co: Colluvic Cambisol; ◇ CM-eu: Eutric Cambisol; □ LP-li: Lithic Leptosol; ● UM-lep: Umbrisol Epileptic; ○ UM-cm: Cambic Umbrisol; ● LV-ro: Rhodic Luvisol).

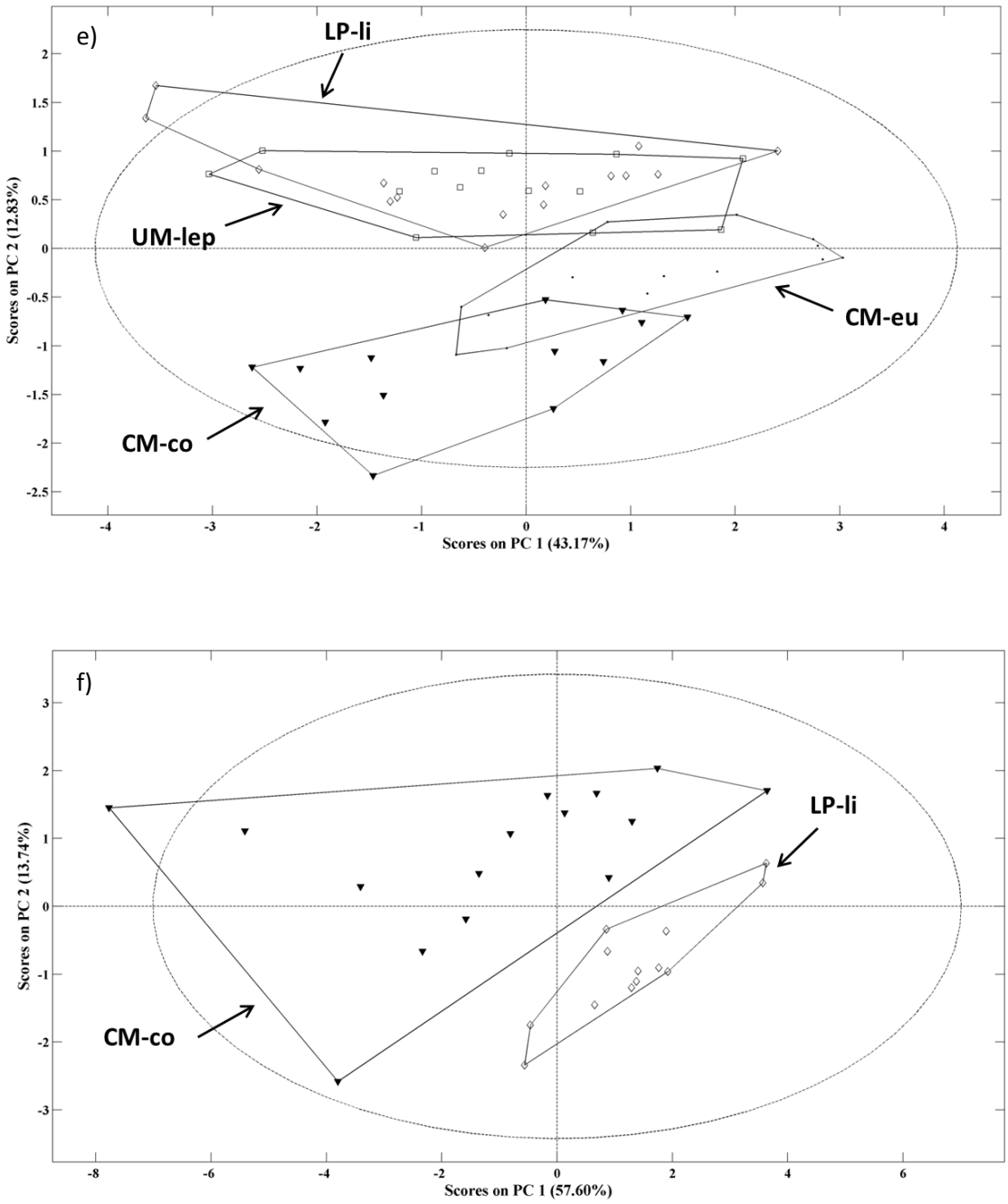


Figure 3.5.4. **(cont.).** Score plots obtained by PCA of grapevine leaves spectra in the 800 to 1375 nm and 1475 to 1850 nm regions, considering different varieties (a:*syrah*, b:*arinto*,c:*antão vaz*, d:*chardonnay*, e:*verdelho*, f:*encruzado*, g:*jaen*, h:*alvarinho*, i:*semillon*). Markers are defined accordingly to the soil type (▼ CM-ca: Calcaric cambisol; ✱ CM-je-lep: Hypereutric Cambisol Epileptic; + CM-je: Hepereutric Cambisol; ■ CM-co: Colluvic Cambisol; ◆ CM-dy: Dystric Cambisol; × CM-co: Colluvic Cambisol; ◇ CM-eu: Eutric Cambisol; □ LP-li: Lithic Leptosol; ● UM-lep: Umbrisol Epileptic; ○ UM-cm: Cambic Umbrisol; ● LV-ro: Rhodic Luvisol).

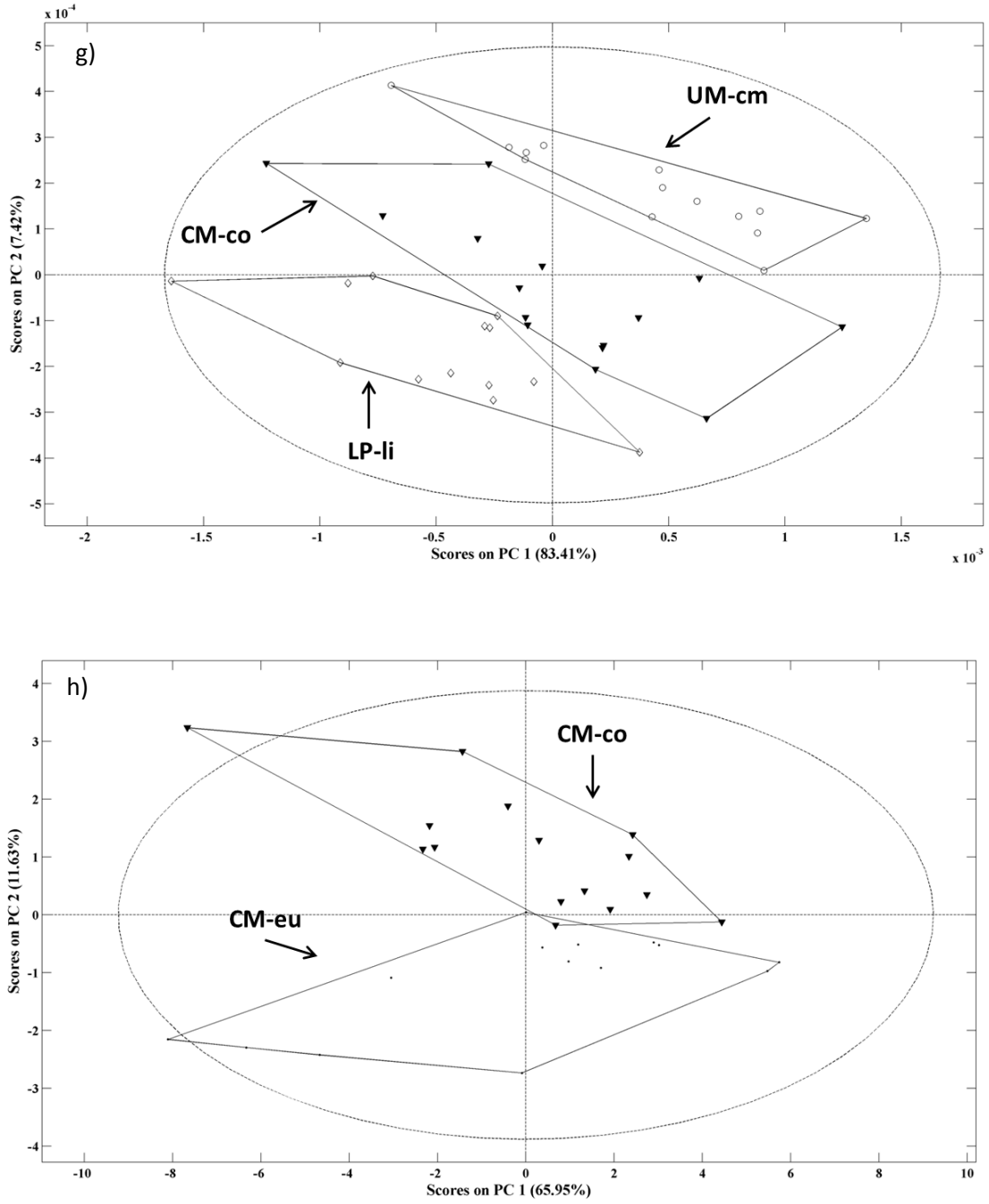


Figure 3.5.4. **(cont.).** Score plots obtained by PCA of grapevine leaves spectra in the 800 to 1375 nm and 1475 to 1850 nm regions, considering different varieties (a:*syrah*, b:*arinto*,c:*antão vaz*, d:*chardonnay*, e:*verdelho*, f:*encruzado*, g:*jaen*, h:*alvarinho*, i:*semillon*). Markers are defined accordingly to the soil type (▼ CM-ca: Calcaric cambisol; ❋ CM-je-lep: Hypereutric Cambisol Epileptic; + CM-je: Hepereutric Cambisol; ■ CM-co: Colluvic Cambisol; ◆ CM-dy: Dystric Cambisol; × CM-co: Colluvic Cambisol; ◇ CM-eu: Eutric Cambisol; □ LP-li: Lithic Leptosol; ● UM-lep: Umbrisol Epileptic; ○ UM-cm: Cambic Umbrisol; ● LV-ro: Rhodic Luvisol).

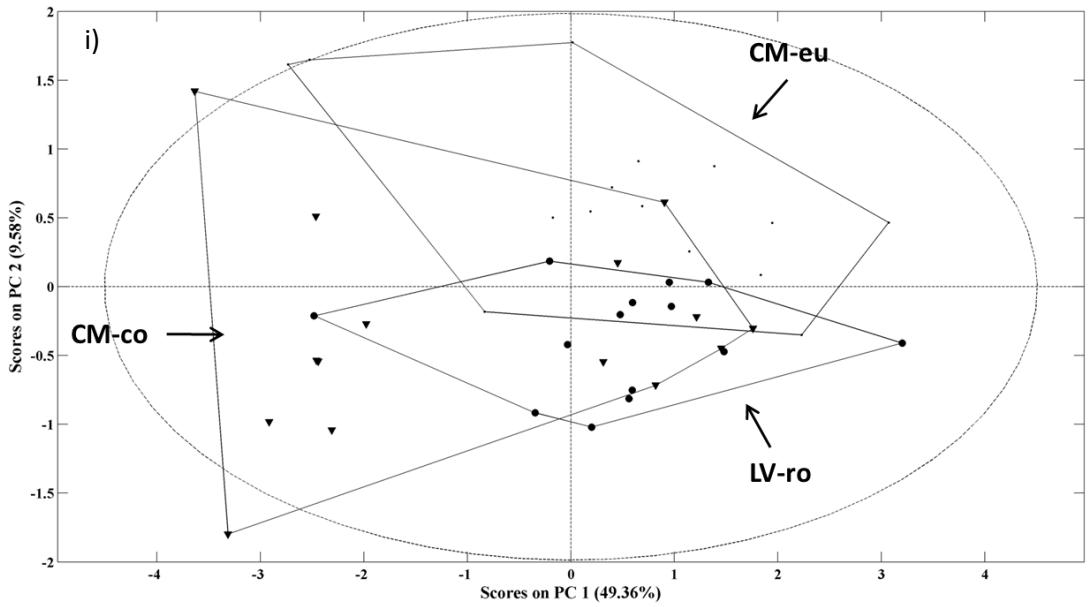


Figure 3.5.4. **(cont.).** Score plots obtained by PCA of grapevine leaves spectra in the 800 to 1375 nm and 1475 to 1850 nm regions, considering different varieties (a:*syrah*, b:*arinto*,c:*antão vaz*, d:*chardonnay*, e:*verdelho*, f:*encruzado*, g:*jaen*, h:*alvarinho*, i:*semillon*). Markers are defined accordingly to the soil type (▼ CM-ca: Calcaric cambisol; ❋ CM-je-lep: Hypereutric Cambisol Epileptic; + CM-je: Hepereutric Cambisol; ■ CM-co: Colluvic Cambisol; ◆ CM-dy: Dystric Cambisol; × CM-co: Colluvic Cambisol; ◇ CM-eu: Eutric Cambisol; □ LP-li: Lithic Leptosol; ● UM-lep: Umbrisol Epileptic; ○ UM-cm: Cambic Umbrisol; ● LV-ro: Rhodic Luvisol).

3.5.3.2. Wavelength selection and soils discrimination by multivariate regression

The promising exploratory data analysis results suggest the use of a supervised classification method aiming at discriminating soils. The supervised PLS-DA regression method has been employed thoroughly in similar situations and was used here [30]. For the application of this method, a sufficient number of samples should be used and samples for each class must be properly balanced. It was expected, based on the literature [31], that the visible range (350-800 nm) would be an appropriate region to differentiate plants grown on different soils due to chlorophyll and carotenoids pigment absorptions. However, the previous study did not intend to differentiate specific soils. It is also known that the 1375-1475 nm and 1870-1940 nm regions characteristic of O-H vibrations in water (first overtone and combination bands respectively) should not be selected. This is in agreement with the supposition that unwanted effects caused by leaves water content/humidity variability will degrade soils discrimination performance

[32]. Several PLS-DA models calibrated with different spectral windows were tested with the objective of identifying the most important spectral zones for discrimination of soils (Fig. 3.5.3). The best spectral regions are depicted in darker colors, meaning that the color gradient gives information about which regions are more relevant. From Fig. 3.5.3 it is possible to identify the spectral windows yielding the lowest error models (results obtained for a PLS-DA model considering different soils in both vineyards). Essentially, two regions were identified as producing the lowest model errors in terms of root mean square error cross-validation (RMSECV) or highest % of correct predictions: 800-1375 nm and 1475-1850 nm. Both windows were tested individually and in combination. The PLS-DA model combining the 800-1375 nm and 1475-1850 nm wavelength windows produced a lower RMSECV when compared to the models when each window was individually selected. These wavelengths were selected for all subsequent model calibrations.

Models considering each grape variety individually were calibrated and results for the test sets are summarized in Table 3.5.3. The lowest percentage of correct assignments (92.0%) was obtained for the Syrah variety at *Herdade do Peso* (model encompassing six different soil types). The highest percentage of correct assignments (98.7%) was obtained for *Antão Vaz* variety also at *Herdade do Peso.* Note that models based on few classes will tend to generate better results. These results also corroborate PCA results in the sense that very good soils discrimination was obtained (92.0-98.7% of correct predictions for the test sets depending on the modelled grapevine variety). For some situations it was difficult to see on a single PCA score plot (considering only 2 components) the discrimination of all soils (e.g., for *Syrah* variety in Fig. 3.5.4a). Additionally, these results show that models encompassing all varieties perform equally well which means that soil related information has influence over leaves' spectra independently of the grape variety. It is also interesting to note the remarkable similarity between the results obtained for *Herdade do Peso* and *Quinta dos Carvalhais* (values between 94.2 and 94.6%).

Table 3.5.3. Soils discrimination PLS-DA model results for *Herdade do Peso* and *Quinta dos Carvalhais* considering varieties individually and in combination (all varieties model).

Vineyard	Variety	Soils	%[1]	Std[2]	LV
Herdade do Peso	*Syrah*	5	92.0	4.7	7
	Arinto	3	96.4	5.4	3
	Antão Vaz	2	98.7	3.9	3
	Chardonnay	2	97.0	6.3	3
	All varieties	5	94.2	2.8	10
Quinta dos Carvalhais	*Verdelho*	4	97.0	4.7	5
	Encruzado	2	95.2	6.1	3
	Jaen	3	97.3	3.8	4
	Alvarinho	2	94.7	7.8	4
	Semillon	3	95.7	6.2	5
	All varieties	6	94.2	3.1	10

[1] Percentage of soils correct predictions for the test set; [2] Standard deviation of soils correct predictions for the test set (obtained by bootstrapping the models 500 times according to [26].

Tables 3.5.4 and 3.5.5 summarize the correct predictions (for the test set) obtained for the global soil discrimination models of *Herdade do Peso* and *Quinta dos Carvalhais*, respectively, under the form of confusion matrices. The diagonal elements sum of the confusion matrices gives the percentage of correct predictions. In both cases, circa 95% correct predictions for the soil taxonomic type were obtained. Moreover, confusion matrices convey additional information, namely some hints to interpret wrong predictions (non-diagonal confusion matrix elements). The PLS-DA models show a prediction accuracy of circa 95% (percentage of correct predictions calculated for the test sets). In *Herdade do Peso* and *Quinta dos Carvalhais*, the worst soil prediction involved the “Hypereutric Cambisol Epileptic” (CM-je-lep) and “Umbrisol Epileptic” (UM-lep) soils, respectively, with around 92% and 87% of correct predictions. Overall, these results show the ability and accuracy of the NIR spectroscopy technique to discriminate between different soils. The method produced comparable results when applied to two distinct vineyards located in different wine regions. Further studies should be performed preferably in different areas of the globe, to corroborate encouraging preliminary results.

Table 3.5.4. Confusion matrix for the soil discrimination model based on the leaves spectra of *Herdade do Peso* (94.2% of global correct predictions rate and 10 LVs). Values are in %.

% Soil type (Predicted)	Soil type (Real)					
	CM-ca	CM-je-lep	CM-je	CM-co	CM-dy	Total
CM-ca	31.0	0.4	0.0	0.2	0.0	31.7
CM-je-lep	1.1	23.6	0.5	0.3	0.0	25.5
CM-je	0.1	0.8	13.8	0.3	0.0	15.1
CM-co	0.5	0.9	0.0	18.9	0.3	20.6
CM-dy	0.1	0.0	0.0	0.2	6.8	7.2
Total	32.9	25.7	14.2	20.0	7.1	100

Table 3.5.5. Confusion matrix for the soil discrimination model based on leaves spectra of *Quinta dos Carvalhais* (94.2% of global correct predictions rate and 10 LVs). Values are in %.

% Soil type (Predicted)	Soil type (Real)						
	CM-co	CM-eu	LP-li	UM-lep	UM-cm	LV-ro	Total
CM-co	33.8	0.0	0.2	1.6	0.5	0.0	36.1
CM-eu	0.6	20.4	0.8	0.1	0.0	0.0	21.9
LP-li	0.0	0.6	6.6	0.5	0.0	0.0	7.7
UM-lep	0.4	0.2	0.0	19.0	0.0	0.2	19.8
UM-cm	0.0	0.1	0.0	0.0	7.0	0.0	7.1
LV-ro	0.0	0.0	0.0	0.0	0.0	7.4	7.4
Total	34.8	21.3	7.6	21.2	7.5	7.6	100

The loadings of the PLS-DA model encompassing all varieties for *Quinta dos Carvalhais* were analysed with the objective of better interpreting the models (Fig. 3.5.5). The goal is to provide a closer insight about specific wavenumbers that are more important for soils discrimination. Similar results were obtained for *Herdade do Peso* (data not shown). For simplicity, only the first two latent variables were compared (encompassing approximately 90% of the total variance in the Vis/NIR spectra). Five

regions appear to be more important for the soils discrimination. The region centred at 800 nm corresponds to RNH_2 and RNHR' absorption bands (third overtone region) and could be related to proteins [33] and chlorophyll [34]. A second region around 850 nm could be related with phenolic compounds, since this region is responsible for the ArCH absorptions (third overtone region) [34]. A third region between 900 and 950 nm corresponds to CH_3, CH_2, CH, ROH and ArOH absorption bands (third overtone region) and is possibly related with carbohydrates, more specifically cellulose and lignin [33, 34]. A fourth region between 1100 and 1180 nm due to CH_3, CH_2 and CH absorptions (second overtone region) is related with carbohydrates, chlorophyll and carotenoids [34]. The last region ranges from 1700 to 1800 nm which corresponds to CH_3, CH_2, CH and SH (first overtone region) absorption bands and could be related with the carbohydrates (cellulose and lignin again) and phenolic compounds [33, 34]. Shortly, the discrimination of the different types of soil by NIRS appears to be related with protein, chlorophyll, carotenoid, phenolic and carbohydrates (more specifically cellulose and lignin) content in leaves.

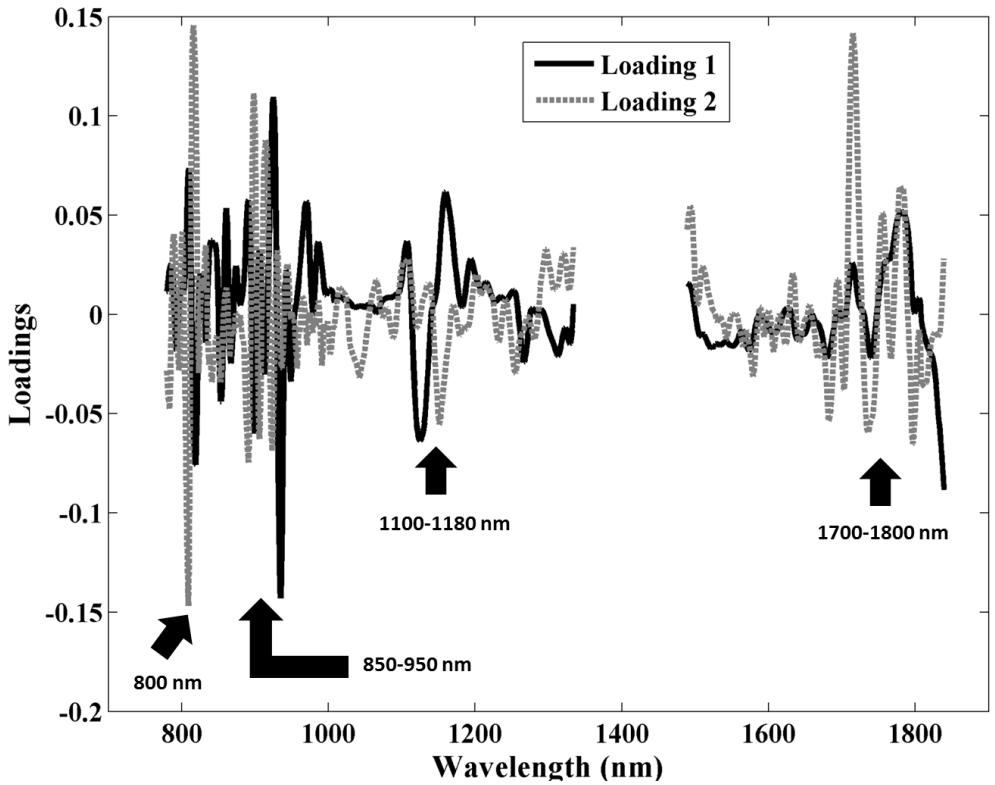


Figure 3.5.5. Representation of the loadings of the first five latent variables of the PLS-DA model developed to model all soil types of *Quinta dos Carvalhais*.

3.5.3.3. Methodology validation

The proposed methodology was further validated on a set of spectral measurements obtained from a *Quinta dos Carvalhais* block (Alvarinho grapevine) with an approximately area of 0.3ha (Fig. 3.5.1a). The existing soil map from which this work was based shows two different soil types in this block: CM-co and CM-eu. Spectroscopic measurements were made intensively in this block as described in the materials and methods section. Spectra collected in this block were projected onto the PCA and PLS-DA models constructed with all samples of *Quinta dos Carvalhais* (all soil taxonomic types). Alternatively, an option would be to calibrate PCA and PLS-DA models using only that variety samples and considering only the two soil types that are known to exist in the test block. Despite the fact that results might be better, this strategy is not viable because when scanning an unknown block, there is no prior knowledge regarding the existing soil. Therefore, and to simulate a real situation of the utilization of this method, global models considering all soil types in the vineyard were considered. The first score and the first latent variable, respectively for PCA and PLS-DA projections were used primarily to define contour maps (Fig. 3.5.6). These components are the most relevant regarding soils differentiation (especially in the case of PLS-DA). A comparison between the existing map with the maps produced with the proposed strategy revealed a very good match. As stated before, the original mapping should be considered only as a broad reference due to the rough scale used to define soil boundaries. Although very similar in shape, the maps produced with PCA and PLS-DA models, show visible transition zones which is something we would expect on the boundary of two soils with different characteristics. It is not possible to identify significant differences between the PCA and PLS-DA maps. PLS-DA predictions could be used instead of the first latent variable to define the mapping. Model predictions, in this case, consisted essentially of the two soils coexisting in this block and the resulting pattern was very similar to that shown in Fig. 3.5.6 (data not shown) although on a two color scale (each color would represent the prediction of a specific soil). PLS-DA predictions estimated 61% of the total predictions to be soil type CM-co, 35% CM-eu and 4% other soils. This is in agreement with the existing soil distribution information for this block (Fig. 3.5.6). Note that a PCA model might be directly calibrated with the spectra only, making this method not relying on any previous knowledge. However, this approach would allow only the generation of maps based on spectral variability. Whether these maps have some relation with soils would only be a hypothesis, difficult to validate without further analyses. The produced mapping could have been confirmed by scanning the soil under the plants, eventually using the same methodology hereby

used to analyse leaves [8] thus proving that the predicted detailed map is not the result of a wrong classification or artefacts. The proposed Vis/NIRS method has the potential advantage of increasing the resolution of soil maps since leaves of every plant can be measured. It has also the benefit of allowing the adjustment of desirable sampling grids and highlighting soil transition zones. An operator with a spectrometer can scan one vineyard hectare on approximately 5 hours considering one measurement in triplicate covering all plants.

Figure 3.5.6. Soil mapping for the selected block for intensive monitoring in *Quinta dos Carvalhais* (a- soil map characterized with pedology methods (black: CM-so, white: UM-lep), b- contour map generated by the PCA first principal component score, c- contour map generated by the PLS-DA first latent variable score of the spectral block).

3.5.4. Conclusions

This work proposes a methodology for indirectly scanning vineyard soil types through the in-situ VIS/NIR analysis of grapevine leaves of *Vitis vinifera* species varieties with a portable instrument in diffuse reflectance mode. The feasibility of this spectroscopic instrumental method for soils discrimination based on grapevine leaves spectra was demonstrated for two vineyards in different wine regions of Portugal, considering a total of 10 different soil types (according to the World Reference Base for Soil Resources 2014). Both unsupervised and supervised approaches, using respectively PCA and PLS-DA models generated very accurate soil maps with a very good agreement with previously existent soil maps for these vineyards. The supervised approach based on PLS-DA showed that the proposed method matched the existent soil mapping with an approximate accuracy of 95%. The proposed method was applied to intensively monitoring a single vineyard block. Results confirmed the method's potential to generate soil maps. The proposed method is therefore able to provide evidence for

soils variability with no a-priori knowledge and is able to define efficient soil mappings if previous information regarding soils (e.g. chemical and physical parameters, properties, taxonomic classification) has been revealed for the region of interest. Globally, it was demonstrated that the spectral information acquired with a Vis/NIR portable spectrometer can be used to define accurate soil maps based on plant leaves measurements with the potential advantage of highlighting soil transition zones. This work explored a new support tool to assist plantation/replantation processes with the objective of grape quality and yield. The information gathered will allow the definition of more adjusted grapevine varieties to a specific soil, instead of having one grapevine variety per entire block (as is usual through all vineyards around the world). Further investigation can be made with Vis/NIR data obtained from remote sensing devices (e.g., hyperspectral imaging) reducing dramatically the time necessary for a wide area analysis, especially interesting for large area vineyards. Results obtained in this work seem very promising and suggest that this method has the potential to be used with other crops aiming at the characterization of different soils

3.5.5 References

1. Mazzetto, F., et al., *Integration of optical and analogue sensors for monitoring canopy health and vigour in precision viticulture.* Precision Agriculture, 2010. **11**(6): p. 636-649.
2. White, R.E., *Putting It All Together* in *Understanding Vineyard Soils*. 2009, Oxford University Press, USA: New York. p. 195.
3. Jackson, D.I. and P.B. Lombard, *Environmental and management practices affecting grape composition and wine quality- a review.* American Journal of Enology and Viticulture, 1993. **44**(4): p. 409-430.
4. Bodin, F. and R. Morlat, *Characterization of viticultural terroirs using a simple field model based on soil depth I. Validation of the water supply regime, phenology and vine vigour, in the Anjou vineyard (France).* Plant and Soil, 2006. **281**(1-2): p. 37-54.
5. Peuke, A.D., *Nutrient composition of leaves and fruit juice of grapevine as affected by soil and nitrogen fertilization.* Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde, 2009. **172**(4): p. 557-564.
6. Workman, J.J. and L. Weyer, *Introduction to near-infrared spectra*, in *Pratical Guide to Interpretive Near-Infrared Spectroscopy*. 2008, CRC Press, Inc.: Boca Raton, FL, USA. p. 1-21.
7. Govender, M., et al., *Review of commonly used remote sensing and ground-based technologies to measure plant water stress.* Water Sa, 2009. **35**(5): p. 741-752.
8. Salazar, D.M., et al., *Visible-near infrared spectroscopy to assess soil contaminated with cobalt*, in *International Meeting of Electrical Engineering Research 2012*, L.A. Zuppa, et al., Editors. 2012, Elsevier Science Bv: Amsterdam. p. 245-253.
9. Cozzolino, D., et al., *In Situ Measurement of Soil Chemical Composition by Near-Infrared Spectroscopy: A Tool Toward Sustainable Vineyard Management.* Communications in Soil Science and Plant Analysis, 2013. **44**(10): p. 1610-1619.
10. Steele, M.R., et al., *Nondestructive Estimation of Anthocyanin Content in Grapevine Leaves.* American Journal of Enology and Viticulture, 2009. **60**(1): p. 87-92.

11. Calcante, A., A. Mena, and F. Mazzetto, *Evaluation of "ground sensing" optical sensors for diagnosis of Plasmopara viticola on vines.* Spanish Journal of Agricultural Research, 2012. **10**(3): p. 619-630.
12. Diago, M.P., et al., *Identification of grapevine varieties using leaf spectroscopy and partial least squares.* Computers and Electronics in Agriculture, 2013. **99**: p. 7-13.
13. Goutouly, J.P., et al., *Characterisation of vine vigour by ground based NDVI measurements.* In Proceedings of VI International Terroir Congress, ENITA, Bordeaux (France), 2006: p. 237-242.
14. Drissi, R., et al., *Nondestructive measurement of grapevine leaf area by ground normalized difference vegetation index.* Agronomy Journal, 2009. **101**(1): p. 226-231.
15. Mazzetto, F., A. Calcante, and A. Mena, *Comparing commercial optical sensors for crop monitoring tasks in precision viticulture.* Journal of Agricultural Engineering, 2009. **1**: p. 11-18.
16. Meggio, F., et al., *Grape quality assessment in vineyards affected by iron deficiency chlorosis using narrow-band physiological remote sensing indices.* Remote Sensing of Environment, 2010. **114**(9): p. 1968-1986.
17. Qin, J.L., et al., *A non-linear model of nondestructive estimation of anthocyanin content in grapevine leaves with visible/red-infrared hyperspectral*, in *Computer and Computing Technologies in Agriculture Iv, Pt 4*, D.L. Li, Y. Liu, and Y.Y. Chen, Editors. 2011, Springer-Verlag Berlin: Berlin. p. 47-62.
18. Ciraolo, G., et al., *Mapping evapotranspiration on vineyards: a comparison between Penman-Monteith and energy balance approaches for operational purposes*, in *Remote Sensing for Agriculture, Ecosystems, and Hydrology Xiv*, C.M.U. Neale and A. Maltese, Editors. 2012, Spie-Int Soc Optical Engineering: Bellingham.
19. Saiz-Rubio, V. and F. Rovira-Mas, *Dynamic segmentation to estimate vine vigor from ground images.* Spanish Journal of Agricultural Research, 2012. **10**(3): p. 596-604.
20. Kodaira, M. and S. Shibusawa, *Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping.* Geoderma, 2013. **199**: p. 64-79.
21. Stamatiadis, S., et al., *Relation of ground-sensor canopy reflectance to biomass production and grape color in two Merlot vineyards.* American Journal of Enology and Viticulture, 2006. **57**(4): p. 415-422.

22. Rinnan, A., F. van den Berg, and S.B. Engelsen, *Review of the most common pre-processing techniques for near-infrared spectra.* Trac-Trends in Analytical Chemistry, 2009. **28**(10): p. 1201-1222.
23. Naes, T., et al., *Interpreting PCR and PLS solutions*, in *A User-Friendly Guide to Multivariate Calibration and Classification*. 2004, NIR Publications: Chichester, UK. p. 39-54.
24. Barker, M. and W. Rayens, *Partial least squares for discrimination.* Journal of Chemometrics, 2003. **17**(3): p. 166-173.
25. Geladi, P. and B.R. Kowalski, *PARTIAL LEAST-SQUARES REGRESSION - A TUTORIAL.* Analytica Chimica Acta, 1986. **185**: p. 1-17.
26. Preisner, O., et al., *Application of Fourier transform infrared spectroscopy and chemometrics for differentiation of Salmonella enterica Serovar Enteritidis phage types.* Applied and Environmental Microbiology, 2010. **76**(11): p. 3538-3544.
27. Porra, R.J., W.A. Thompson, and P.E. Kriedemann, *Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophill standards by atomic absorption spectroscopy.* Biochimica Et Biophysica Acta, 1989. **975**(3): p. 384-394.
28. Zscheile, F.P., et al., *The preparation and absorption spectra of five pure carotenoid pigments.* Plant Physiology, 1942. **17**(3): p. 331-346.
29. Batovska, D.I., et al., *Seasonal variations in the chemical composition of vine-grape leaf surface.* Natural Product Research, 2008. **22**(14): p. 1237-1242.
30. Coutinho, C.P., I. Sa-Correia, and J.A. Lopes, *Use of Fourier transform infrared spectroscopy and chemometrics to discriminate clinical isolates of bacteria of the Burkholderia cepacia complex from different species and ribopatterns.* Analytical and Bioanalytical Chemistry, 2009. **394**(8): p. 2161-2171.
31. Hagg, C., F. Stober, and H.K. Lichtenthaler, *Pigment content, chlorophyll fluorescence and photosynthetic activity of spruce clones under normal and limited mineral-nutrition.* Photosynthetica, 1992. **27**(3): p. 385-400.
32. Sarraguca, M.C., et al., *Authenticity Control of Roasted Coffee Brands Using Near-Infrared Spectroscopy.* Food Analytical Methods, 2013. **6**(3): p. 892-899.
33. Kokaly, R.F. and R.N. Clark, *Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression.* Remote Sensing of Environment, 1999. **67**(3): p. 267-287.
34. Meyer, B.S., et al., *Introduction to Plant Physiology*. Second ed. 1973: D. Van Nostrand Company.

The following work is intended for future submission in a scientific journal:

Lopo, M., Páscoa, R. N. M. J., & Lopes, J. A. Grapevine ampelographic differentiation using NIR spectroscopy.

3.6. Grapevine ampelographic differentiation using NIR spectroscopy

Abstract: Grapevine leaves have a very important role in photosynthesis and plant water status (leaf water potential can be used as an indicator of overall plant water stress), among others; their health status is therefore extremely valuable to the wine industry, but their importance also extends to the ampelographic differentiation of cultivars. Before the advent of DNA fingerprinting, the identification and classification of grapevines was traditionally performed by comparison of the shape of berries and leaves. An extremely costly and long process, dependent on the availability and physical presence of few experts. Even though genetic fingerprinting is a reliable and accurate technique to identify different vine species, these analyses are usually time-consuming, expensive and require laboratory conditions. Methods based on leaf light reflectance are both faster and less expensive. Visible and near-infrared (vis/NIR) spectroscopy is a rapid, non-destructive, inexpensive and accurate analysis technique. It has been widely used in various fields of the wine industry, namely in wine and grape analysis; it has been used to obtain certain information about plant composition, such as chlorophyll, water content, nitrogen (N) concentrations, but has hardly been explored for the ampelographic differentiation of vines. In this study, 15 different varieties of grapevines leaves from four different vineyards, in four different Wine Regions of Portugal (Alentejo, Dão, Douro, Vinhos Verdes) were scanned *in-situ* using a FieldSpec 4 Wide-Res (ASD Inc, Boulder, CO) in diffuse reflectance mode over a spectral range of 28571 – 4000 cm^{-1}. The spectra were further processed with chemometric tools such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). The objective of this work is to realise if vis/NIR spectroscopy technology is an adequate technology for plant phenotyping and is able to correlate, complement and become a rapid *in-situ* alternative to costly, time-consuming genetic techniques. Results have shown an excellent correlation between grapevine leaves and vine species through near infrared spectroscopy (NIRS) with percentages of correct predictions above 90%.

3.6.1. Introduction

Plant phenotyping refers to the quantitative description of the plant's physiological, biochemical and morphological properties [1]. In the context of the current worldwide industrial demand of quality and efficiency in crop and food production, plant phenotyping is undoubtedly a very important topic in agriculture and its importance rises every day. Plant phenotyping consists on the identification of effects on the phenotype as a result of genotype differences and the environmental conditions to which a plant has been exposed [2]. *Vitis vinifera* L. leaves are regarded as extremely valuable to the wine industry due to their role in photosynthesis, plant water status (leaf water potential can be used as an indicator of overall plant water stress), but also for their importance in ampelographic differentiation of the cultivars. Before the advent of DNA fingerprinting, the identification and classification of grapevines was traditionally achieved by classic ampelometry [3] where a comparison between the shape of berries, leaves and trunks was performed and the differences analysed by trained experts making its utilization limited both in space and time. Even though genetic fingerprinting is a reliable and accurate technique to identify vine species, these analyses are time-consuming, expensive and must be made under laboratory conditions.

Modern viticultural methods based on precision agriculture have recently began to show a growing interest in remote sensing methods for the temporal and spatial monitoring of grapevine species due to their potential for estimating vine biophysical variables such as shape, size and vigour, as well as vine yield and grape quality [4]. These techniques are proving useful not only for short-term monitoring, but also in long-term decision-making processes regarding vineyard management such as grapevine growth and harvesting [5].

Visible and near-infrared (vis/NIR) spectroscopy is a rapid, inexpensive, non-invasive and accurate analytical technique suited for several agricultural applications due to its rapid data acquisition time, the capability of determining more than one parameter using the same measurement, and its easy and fast usage. Near infrared (NIR) radiation is influenced by combinations and overtones of fundamental vibrational transitions, essentially of C–H, N–H, O–H and S–H bonds present in the sample [6]. Moreover, this technology is non-destructive and allows the acquisition of spectral fingerprints with little or no sample preparation. NIRS has previously been applied for plant varietal discrimination in crops as diverse as wheat [7], pear [8], bayberry [9], strawberry [10] or even tomato [11]. This technology has also been widely used in various fields of the wine industry; it has been used for grape composition analysis [12,

13] and to obtain specific information about plant composition, such as chlorophyll [14], nitrogen concentrations (N) [15] or water content [16]. The possibility of using vis/NIR technology for grapevine phenotyping arises thus as an attractive and promising tool for precision viticulture, especially because this technique is able to characterize more than one parameter using just one spectral measurement [17]. Another important feature of this technique is that it can be incorporated in remote sensing devices or used to obtain multispectral images [18]. Spectral measurements can be made in plant leaves, either at-line in field laboratory or *in-situ* at the vineyard, or by remote devices (remote sensing) [19].

There are numerous studies addressing the discrimination of vegetation species [20, 21], but similar investigations of *V. vinifera* L. varieties remain limited [22]. Nevertheless, datasets using hyperspectral sensors are becoming more and more common, constituting a robust platform of non-destructive measurements that can be used for the estimation of biochemical constituents as well as health condition assessment. A more thorough understanding of within species spectral variation could also prove useful for the refinement of these processes [23].

Vis/NIR spectrometers are able to acquire large amounts of spectral data, making it necessary to manage them in efficient and automatic ways. Chemometrics has become one of the most valuable research fields in the latest few years due to its knowledge discovery power, direct applicability in several areas and, especially, its proven effectiveness in those problems where it is applied. Methods such as principal components analysis (PCA) [24], partial least squares (PLS) [25] and partial least squares discriminant analysis (PLS-DA) [26] have provided procedures for both descriptive (characterizations of the properties of the data) and predictive (learning and induction of the data for forecasting) tasks.

The main objective of this work was to understand if spectral grapevine leaves measurements acquired through vis/NIR spectroscopy. Samples from four different vineyards from specifically different wine regions in Portugal were used with chemometric analysis, for grapevine ampelographic differentiation. Furthermore, an intensive monitoring was performed on single leaves to assess which part of the leaf would provide more meaningful information from a spectral point of view.

3.6.2. Materials and Methods

3.6.2.1. Infield vineyard monitoring

Four vineyards, in four different delimited wine regions in Portugal were selected: *Quinta de Azevedo* (Barcelos, 41º34' 12.45''N 8º32'25.05''O) in the Vinho Verde Wine Region, (north of Portugal); *Quinta do Mourão*, (Lamego, 41°07′23.5″N 7°48′09.5″W) in the Douro Wine Region (north of Portugal); *Quinta dos Carvalhais* (Mangualde, 40.556721°N–7.787247°W) in the Dão Wine Region (center of Portugal) and *Herdade do Peso* (Vidigueira, 38.141579°N–7.677813°W) in the Alentejo Wine Region (south of Portugal) (Fig. 3.6.1).

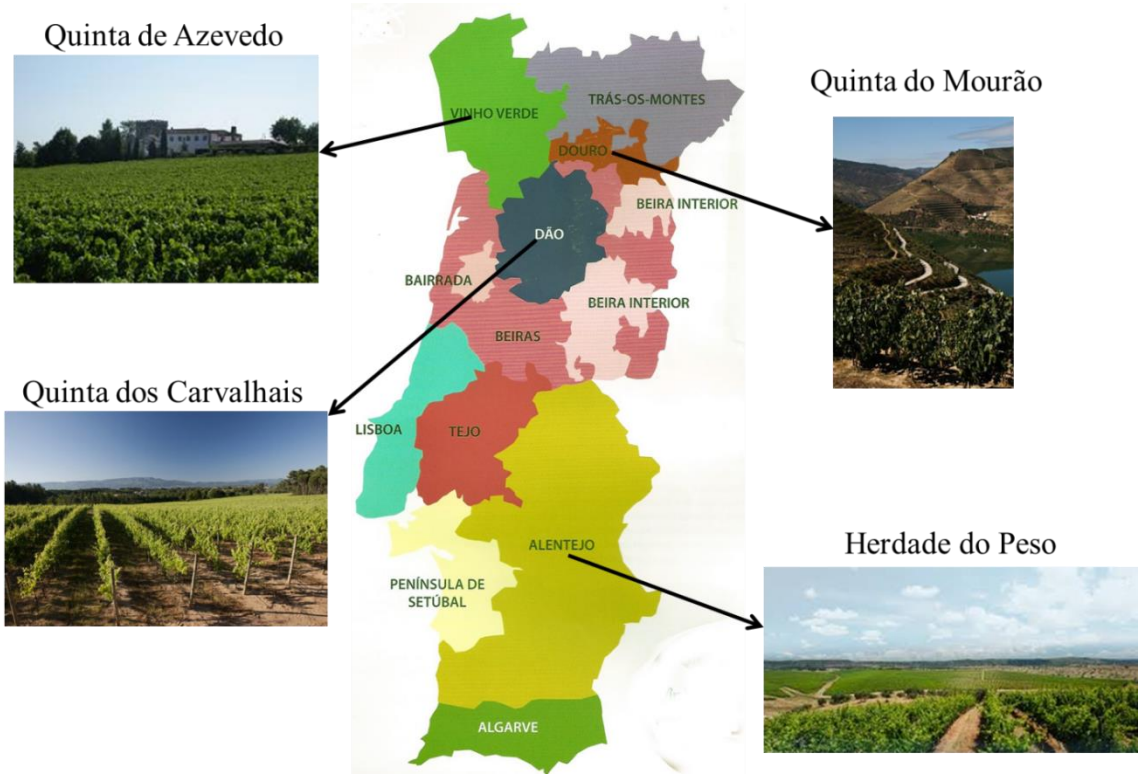


Figure 3.6.1. Vineyards locations in the different wine regions of Portugal.

Vineyards are divided in numbered blocks, each containing a single *V. vinifera* L. cultivar. A total of 15 different varieties were analysed: Alvarinho, Loureiro, Pedernã, Trajadura, Touriga Nacional, Touriga Franca, Tinta Barroca, Verdelho, Encruzado, Jaen, Sémillon, Syrah, Arinto, Antão Vaz, and Chardonnay (Table 3.6.1).

Table 3.6.1. Vineyards grapevine species in the four different estates used in this study.

Vineyard	Grapevine varieties	Sampling spots
Quinta de Azevedo	Alvarinho	1
	Loureiro	7
	Pedernã	2
	Trajadura	1
Quinta do Mourão	Touriga Nacional	Full block
	Touriga Franca	Full block
	Tinta Barroca	Full block
Quinta dos Carvalhais	Verdelho	4
	Encruzado	2
	Jaen	3
	Alvarinho	2
	Sémillon	3
Herdade do Peso	Syrah	8
	Arinto	3
	Antão Vaz	2
	Chardonnay	2

In *Quinta de Azevedo* and *Herdade do Peso* four different blocks were analysed whereas in *Quinta dos Carvalhais*, five different blocks were monitored – one variety per block. In *Quinta do Mourão* three different blocks were monitored, each block containing a single variety (Table 3.6.1). Sampling was performed considering, 11 spots in *Quinta de Azevedo*, 14 spots in *Quinta dos Carvalhais* and 15 spots in *Herdade do Peso*. At each defined sampling spot, a total of twenty leaves in five different plants were monitored (four monitored leaves per plant). A total of 300 spectra were collected in *Herdade do Peso*, 280 in *Quinta dos Carvalhais* and 264 in *Quinta de Azevedo*. In *Quinta do Mourão* a different strategy was implemented. An intensive monitoring was performed on each block (again, four leaves per plant) on three different blocks, each block containing five different rows of 20 plants yielding a total of 1200 spectra. This analysis was performed at the ripening stage shortly after the veraison period.

3.6.2.2. NIR spectra acquisition and pre-processing

Near infrared leaves spectra were acquired using a Field-Spec 4Wide-Res (ASD Inc, Boulder, CO) in diffuse reflectance mode spanning the 28571 – 4000 cm^{-1} range (Fig. 2). The system incorporated a diffuse reflectance contact probe (Hi-Brite, ASD Inc., Boulder, CO) with a measurement surface area equivalent to a 10 mm diameter circle, enclosing a halogenous light source. Leaves were measured in diffuse reflectance mode directly in the plant (in-situ) with no cleaning process involved. Average size leaves located at one shoot of the first arm were selected. Spectral measurements were performed approximately at the leaf centre between 7 am and 11 am. All measurements, within each vineyard, were made in the same day, performed in triplicate and the average considered for further processing. A background was taken every hour with a Spectralon disk (ASD Inc., Boulder, CO). Furthermore, a single leaf was subjected to intensive spatial vis/NIR monitoring to assert which area of the leaf, if any, provided better spectroscopic information regarding grapevine variety differentiation. All vis/NIR spectra were pre-processed with Savitzky–Golay filter (15-points filter size, second order polynomial, and second-order derivative) [27].

3.6.2.3. Spectra modelling

Chemometric analysis of the spectra was performed using principal components analysis (PCA) [24] which was primarily used to perform exploratory data analysis as well as detect outliers. Partial least squares discriminant analysis (PLS-DA) [28] was used to develop calibration models for grapevine varieties discrimination. The PLS-DA models were developed using the PLS-2 algorithm, which handles multiple dependent variables codifying PLS-DA outputs (classes) in multiple variables [28], which is the case here. Model predictions are converted into class assignments using the distribution of calibration predictions obtained from a PLS model built on two or more classes to determine the threshold level yielding the lowest level of false classifications. Basically, in the construction of the PLS-DA models, the data were divided into calibration (70% of the available samples) and test (remaining 30%) sets [29]. This division was made considering blocks of 15 contiguous spectra in order to avoid overfitting problems (e.g., preventing the existence of spectra collected from nearby leaves both in the calibration and test sets). Samples were divided randomly while ensuring that the same proportion between grapevine varieties was present in the calibration and test sets to avoid unbalanced classes across sets [30]. The optimal number of latent variables (LVs) was estimated by leave-one-block-out cross-validation

(contiguous blocks of 15 samples) using only the calibration set [24]. For all developed models, the test set was used to test the accuracy of the PLS-DA models and the corresponding results expressed as confusion matrices [31]. Confusion matrices compare each known grapevine variety with the corresponding NIRS prediction, and entries are expressed as percentages. The PLS-DA loadings were also analyzed to understand which specific wavenumbers are more important for grapevine variety discrimination. Sets of spectra subjected to chemometric analysis were scaled using the mean centring method.

All chemometric methods and spectra processing were performed using Matlab version 7.9 (Mathworks, Natick, MA) and the PLS Toolbox version 5.5.1 (Eigenvector Research, Inc., Wenatchee, WA).

3.6.3. Results and Discussion

The vis/NIR spectra collected from the different vineyards presented the same typical pattern found in previous works [32]. Major bands in the NIR region are related to C–H bonds (combinations and overtones) and O–H (water). The peaks found in the visible region are commonly attributed to pigments, i.e. peaks around 23800 and 15000 cm^{-1} relate to chlorophyll a, while those around 22000 and 15500 cm^{-1} to chlorophyll b. The peak around 22200 cm^{-1} is normally attributed to carotenoids [19] (Fig. 3.6.2).

Figure 3.6.2. Leaves spectral acquisition strategy and vis/NIR spectrum of a grapevine leaf with indication of the major bands.

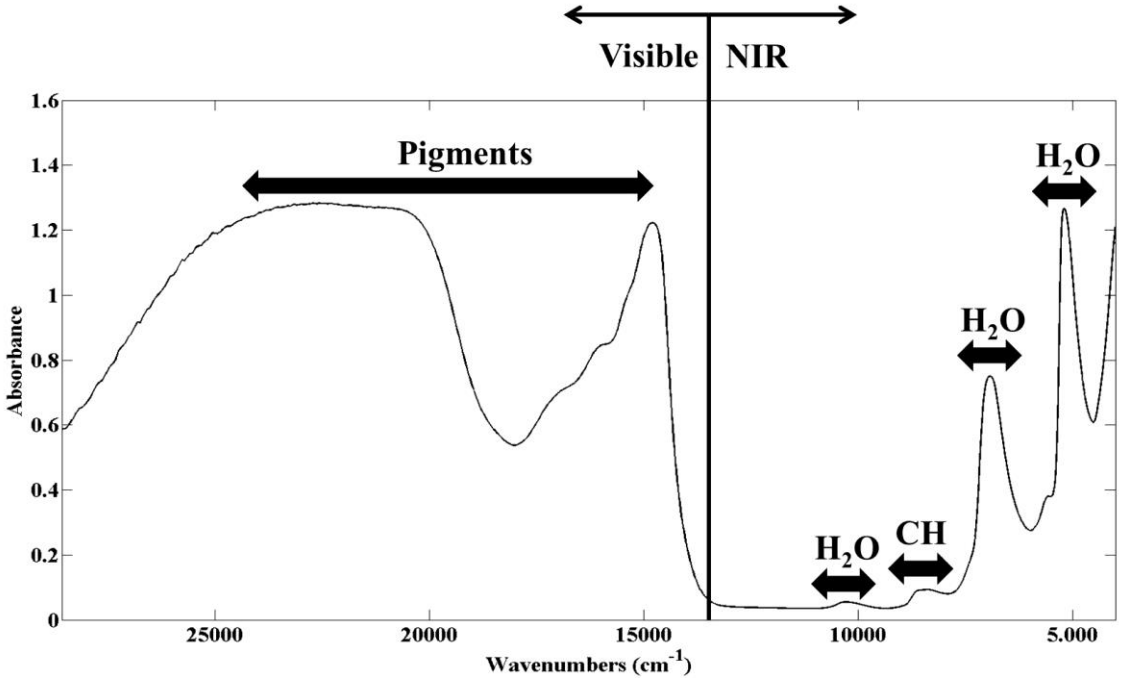


Figure 3.6.2. **(cont.).** Leaves spectral acquisition strategy and vis/NIR spectrum of a grapevine leaf with indication of the major bands.

3.6.3.1. Vine leaf intensive spectral monitoring

A spectroscopic spatial mapping of a single leaf was performed to see if it was possible to discover and understand which part of the leaf, if any, proved more important and contained more information for grapevine species differentiation (Fig. 3.6.3). Equidistant points were established in the subject leaf and the spectra collected. A PCA model was constructed and the scores of the first principal component (91.4% of variance captured) plotted and perfectly transposed onto the photographical representation of the analysed leaf. The first PCA first score revealed indeed that different areas of the leaf yield a specific spectral signature.

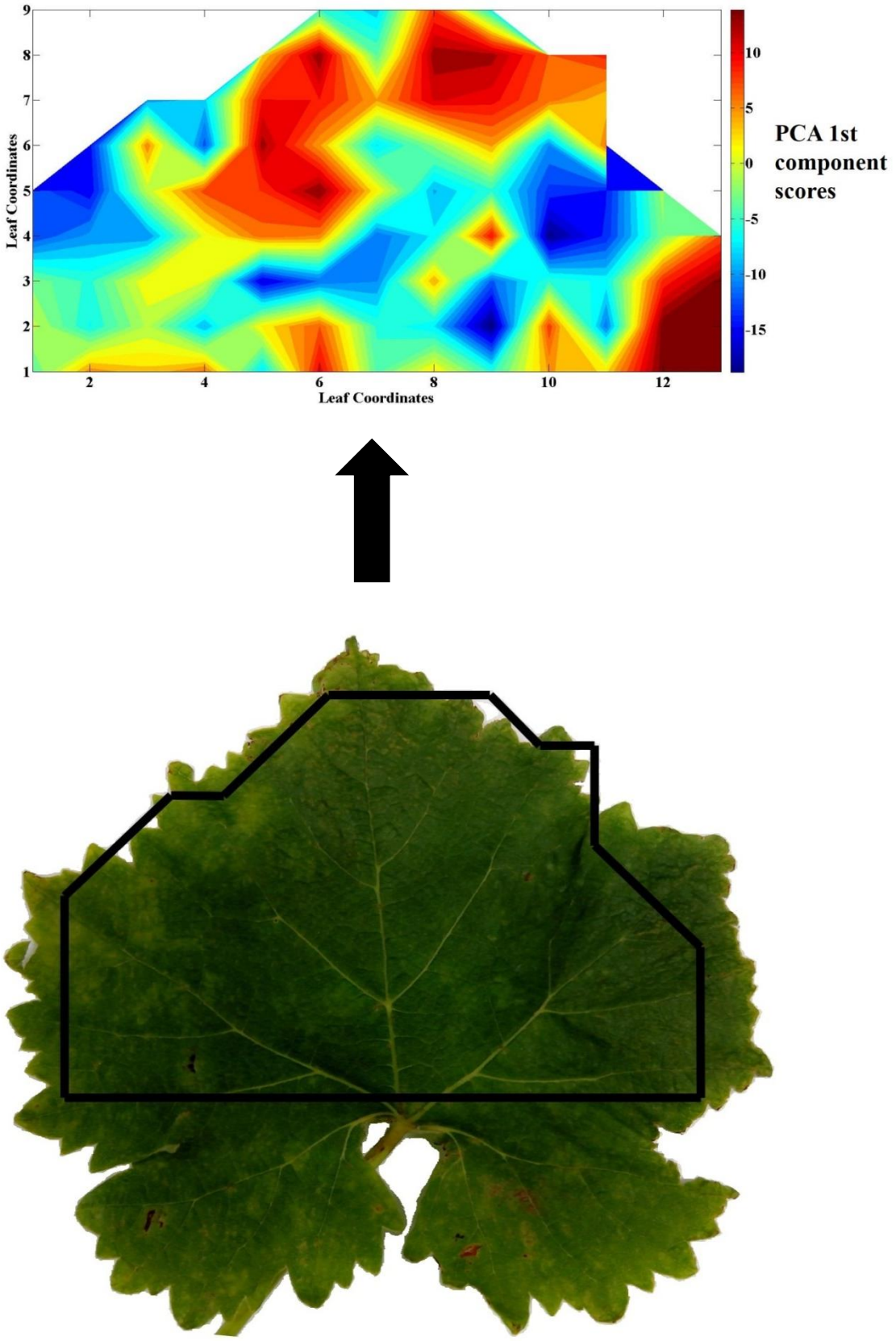


Figure 3.6.3. Grapevine leaf intensive monitoring with leaf spectral map using PCA first component scores and juxtaposition onto the leaf photograph.

3.6.3.2. Grapevine leaves spectra exploratory analysis

The vis/NIR spectral variability obtained within the leaves was analysed. PCA models comprising the entire spectral region (28571 – 4000 cm^{-1}), except noisy areas at both ends of the spectra, were developed for each vineyard, considering the different grapevine varieties. The models revealed evident clustering separation within the grapevine species analysed in all the vineyards and no detection of outliers. A simple scatter plot between the first and second principal components was enough for most varieties to verify that samples cluster according to the specific grapevine species. Score plots of the first two components of the PCA models are shown as an example in Figure 3.6.4. For visual purposes, only two different varieties are plotted at the same time. In cases where satisfactory differentiation was not achieved considering only two components, such as Figure 3.6.4e, a detailed case-by-case analysis was performed using combinations of other components but no further clustering was achieved and therefore this analysis was kept to two major components. The models were able to differentiate equally well red grapevine varieties and white ones as well as endogenous (Trajadura, Pedernã) varieties and very well-known international ones (Sémillon, Jaen).

One of the reasons for unsatisfactory differentiation of grapevine species could be due to the time of year and day of spectral collection. It is known that leaves pigments, metabolites, protein content and photosynthesis vary significantly over the vegetative cycle, especially during the ripening period [33]. Ideally, measurements should be made on the same day or over a relatively short period of time and if possible in similar weather conditions. Further studies addressing the issue of which time of year is best for leaves spectra collection, with the objective of grapevine species differentiation or any other specific characteristic analysis, would certainly be most welcomed. Furthermore, the lack of differentiation of grapevine varieties through PCA could also be due to external factors such as soil humidity, solar exposure or soil treatment. It should also be mentioned that PCA maximized the variance criterion and provides no information whatsoever about grapevine species. However, results appear promising, indicating that there are spectral features of leaves that can be correlated with its phenotype.

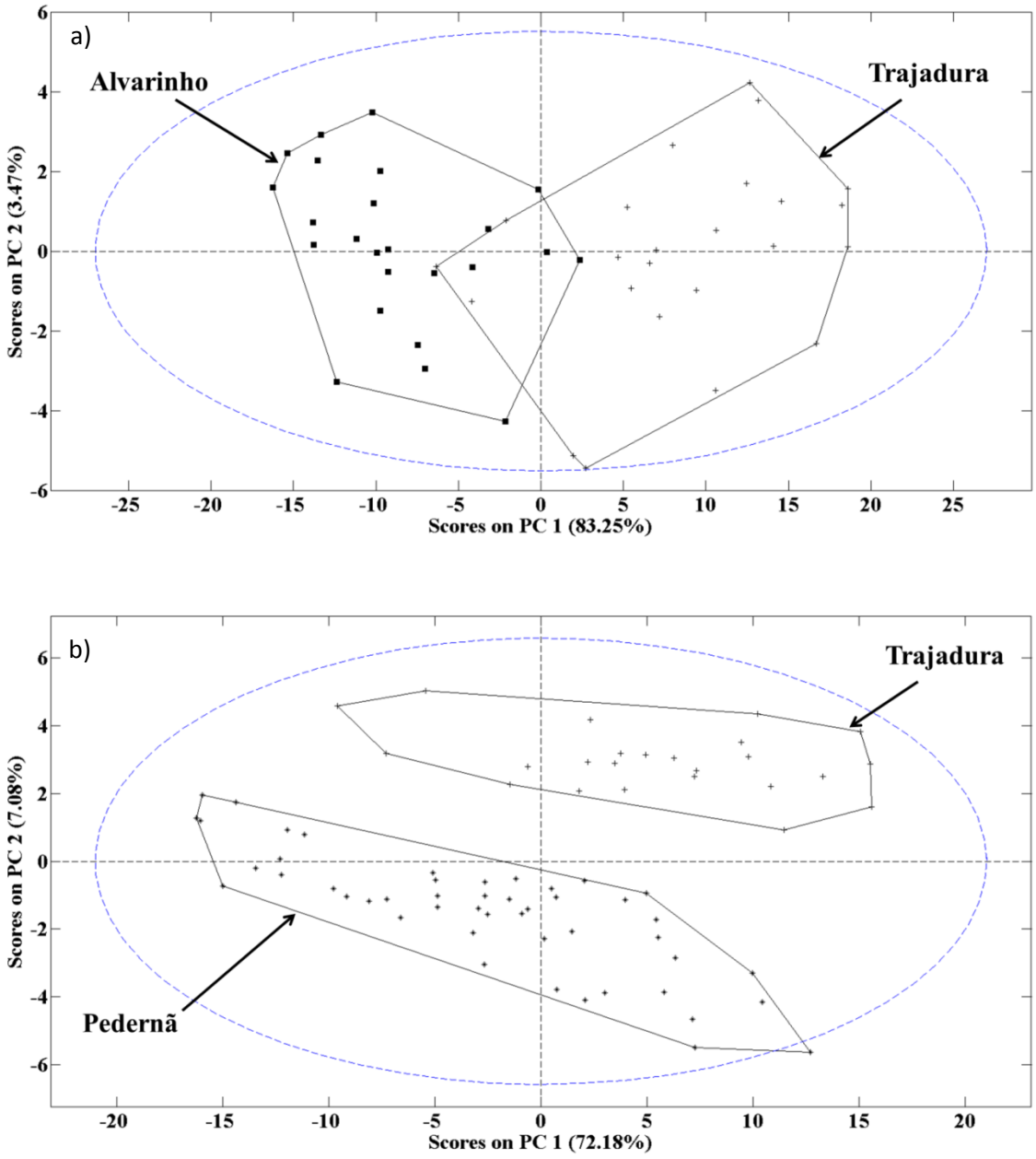
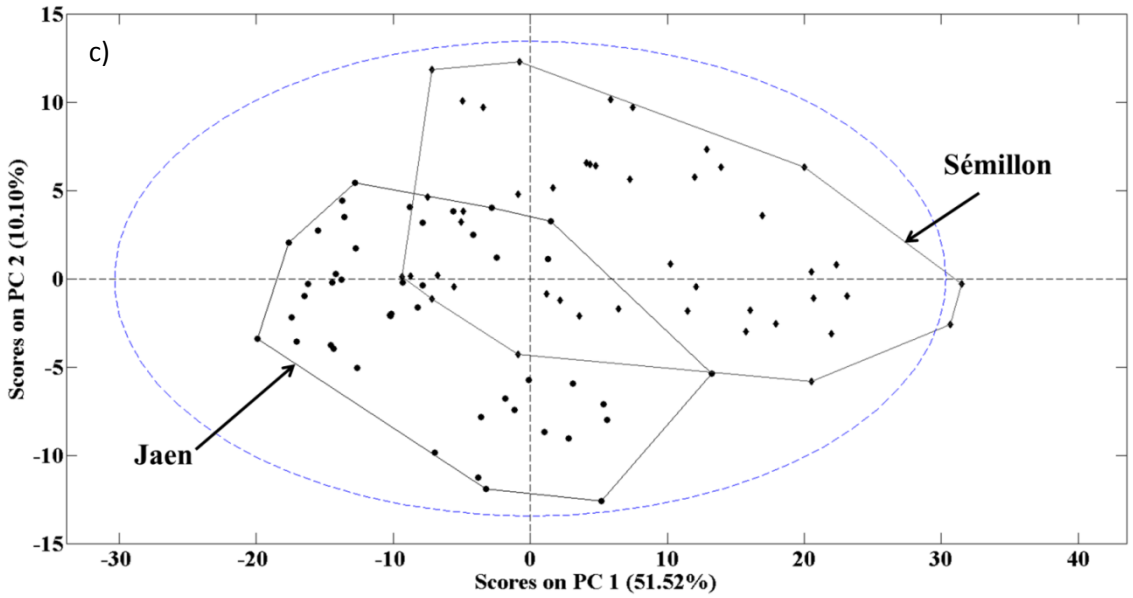


Figure 3.6.4. Score plots obtained from PCA models built on vis/NIR data using the entire spectral region (28571 – 4000 cm^{-1}) from leaves of the different vineyards.
a) and b) *Quinta de Azevedo*; c) and d) *Quinta dos Carvalhais*; e) *Herdade do Peso*.
■ – Alvarinho; **+** – Trajadura; * – Pedernã; ⧫ – Sémillon; ● – Jaen; ▼ – Verdelho; ▲ – Antão Vaz; ★ – Arinto.



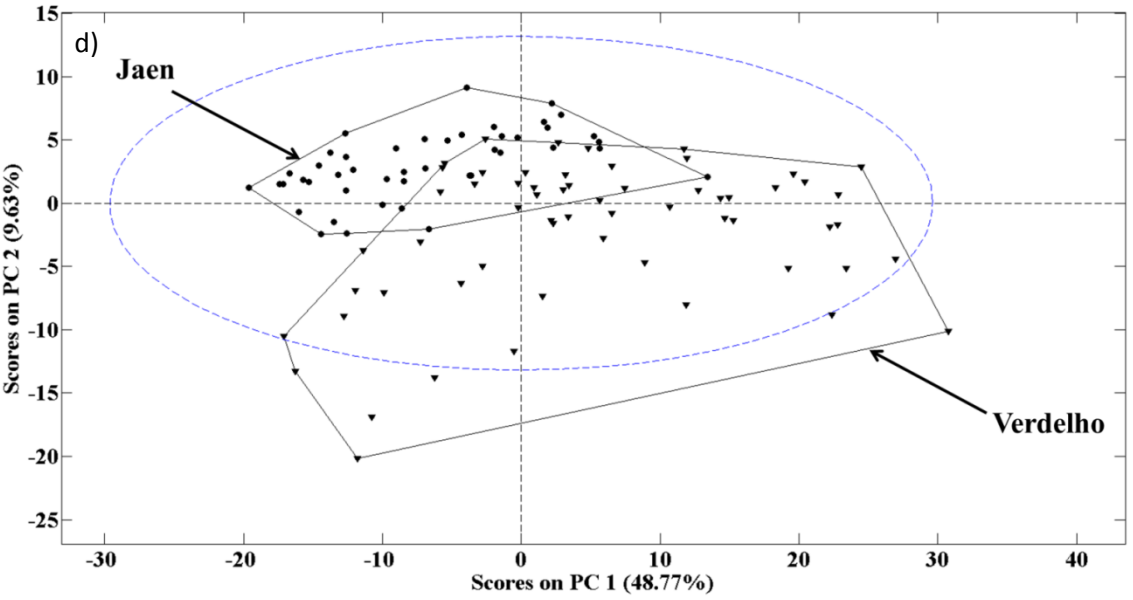


Figure 3.6.4. **(cont.).** Score plots obtained from PCA models built on vis/NIR data using the entire spectral region (28571 – 4000 cm^{-1}) from leaves of the different vineyards.

a) and b) *Quinta de Azevedo*; c) and d) *Quinta dos Carvalhais*; e) *Herdade do Peso*.

■ – Alvarinho; **+** – Trajadura; * – Pedernã; ⧫ – Sémillon; ● – Jaen; ▼ – Verdelho; ▲– Antão Vaz; ★ – Arinto.

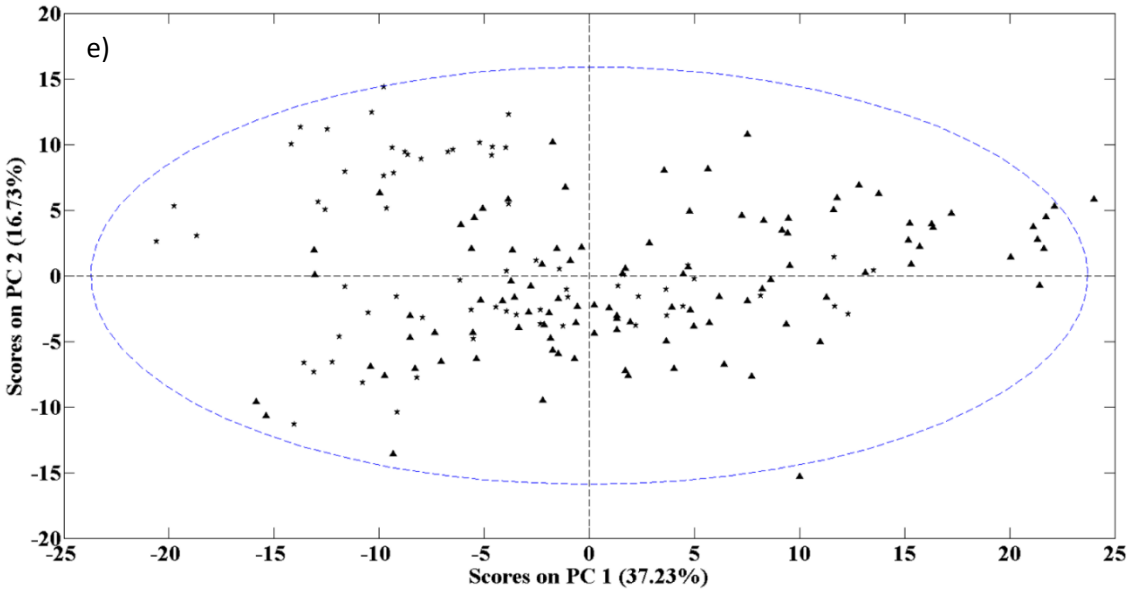


Figure 3.6.4. **(cont.).** Score plots obtained from PCA models built on vis/NIR data using the entire spectral region (28571 – 4000 cm^{-1}) from leaves of the different vineyards.

a) and b) *Quinta de Azevedo*; c) and d) *Quinta dos Carvalhais*; e) *Herdade do Peso*.

■ – Alvarinho; **+** – Trajadura; * – Pedernã; ⧫ – Sémillon; ● – Jaen; ▼ – Verdelho; ▲– Antão Vaz; ★ – Arinto.

3.6.3.3. Varieties discrimination

To further extend the promising results obtained with the spectral exploratory data analysis, the use of a supervised classification method for the discrimination of grapevine varieties was conducted. The supervised PLS-DA regression method has been thoroughly employed for classification problems with cultivars such as strawberry varieties [10] and olive varietal identification [34], but also for vineyards soils discrimination purposes [19, 35]. For the application of this method, a sufficient number of samples should be used and samples for each class must be properly balanced. PLS-DA expects to find a proper correlation of spectral variations and a set of defined classes, this is accomplished by maximizing the covariance value between different class variables and rejecting variance within a class. A selection of different spectral windows was performed (Region 1: 28571 – 12953 cm^{-1}; Region 2: 12952 – 7819 cm^{-1}; Region 3: 7818 – 5974 cm^{-1}; Region 4: 5973 – 5438 cm^{-1}; Region 5: 5437 – 4496 cm^{-1}; Region6: 4495 – 4000 cm^{-1}) with the objective of ascertain which region or combination of regions proves to be more effective for ampelographic discrimination of grapevine varieties. PLS-DA models with all individual regions and possible combinations were

calibrated, thus revealing, by way of extensive testing, which zone or combinations of spectral zones were the most important for grapevine species identification. The 7272 – 6779 cm^{-1} and 5347 – 5154 cm^{-1} regions, characteristic of O–H vibrations in water (first overtone and combination bands respectively) are to be avoided due the possibility of unwanted effects caused by leaves water content/humidity which may degrade the performance of the model for grapevine variety discrimination [36]. However, when performing *in-situ* monitoring this is something difficult to control. Reports from the literature suggested that, due to chlorophyll and carotenoids pigment absorptions, the visible range (28571–12500 cm^{-1}) should be the appropriate region to differentiate plants phenotypes [37], but that was not the case. The two main regions identified as producing the highest percentage of correct predictions or the lowest model errors in terms of root mean square error cross-validation (RMSECV) were: 12952 – 7819 cm^{-1} and 5973 – 5438 cm^{-1}. PLS-DA models combining both regions produced a lower RMSECV than models constructed with each region separately and were therefore selected for all subsequent model calibrations.

Models considering each grapevine variety analysed were calibrated for each vineyard and results for the test sets are summarized in Table 3.6.2 along with the spectral regions used. The models exhibited a very high percentage of correct assignments, with *Quinta do Mourão* exhibiting the best result (98.0% correct prediction). These high results on *Quinta do Mourão* were not totally unexpected due to the different monitoring strategy and fewer grapevine varieties (only three). It is known that models based on few classes will tend to generate better results. Predictions for the models are further presented under the form of confusion matrices (Tables 3.6.3, 3.6.4, 3.6.5 and 3.6.6). The sum of the diagonal values of each confusion matrix gives the percentage of correct predictions for each model. From the analysis of the confusion matrices, it is also possible to calculate the percentage of “misclassifications” for each variety and know the percentage of correct predictions, within the each vineyard model, for each grapevine variety (Table 3.6.7). By calculating the percentage of correct predictions for each variety, one reaches the percentage value with which the specific variety was accurately predicted.

Table 3.6.2. Grapevine varieties overall differentiation PLS-DA models results for all the vineyards.

Vineyard	Grapevine Species	Total number of spectra	Spectral regions (cm^{-1})	Correct predictions (%)	Number of LV's
Quinta dos Carvalhais	5	280	12952–7819 5973–5438	97.2	9
Herdade do Peso	4	300	12952–7819 5973–5438	97.4	8
Quinta de Azevedo	4	264	12952–7819 5973–5438	96.5	7
Quinta do Mourão	4	1600	12952–7819 5973–5438	98.0	9

Table 3.6.3. Confusion matrix for grapevine varieties differentiation using leaves spectra of *Herdade do Peso* (97.4% of global correct predictions rate and 8 LVs). Values are in %.

Predicted Grapevine Varieties	Real Grapevine Varieties				
	Syrah	Arinto	Antão Vaz	Chardonnay	Sum
Syrah	**52.9**	0.0	0.3	0.2	53.5
Arinto	0.0	**20.1**	0.0	0.0	20.1
Antão Vaz	0.0	0.4	**12.4**	0.4	13.2
Chardonnay	0.0	0.4	0.8	**12.0**	13.2
Sum	52.9	20.9	13.5	12.7	100

Table 3.6.4. Confusion matrix for grapevine varieties differentiation using leaves spectra of *Quinta dos Carvalhais* (97.2% of global correct predictions rate and 9 LVs). Values are in %.

Predicted Grapevine Varieties	Real Grapevine Varieties				
	Verdelho	Encruzado	Jaen	Alvarinho	Sum
Verdelho	**27.8**	0.0	0.0	0.0	0.0
Encruzado	0.0	**13.9**	0.0	0.1	0.1
Jaen	0.7	0.0	**20.9**	0.2	0.0
Alvarinho	0.0	0.0	0.7	**12.8**	0.5
Sum	28.5	13.9	21.9	13.2	100

Table 3.6.5. Confusion matrix for grapevine varieties differentiation using leaves spectra of *Quinta do Mourão* (98.0% of global correct predictions rate and 9 LVs). Values are in %.

Predicted Grapevine Varieties	Real Grapevine Varieties			
	Touriga Nacional	Tinta Barroca	Touriga Franca	Sum
Touriga Nacional	**15.2**	0.0	0.4	15.6
Tinta Barroca	0.1	**40.5**	0.4	41.0
Touriga Franca	0.6	0.5	**42.3**	43.4
Sum	15.9	41.0	43.1	100

Table 3.6.6. Confusion matrix for grapevine varieties differentiation using leaves spectra of *Quinta de Azevedo* (97.2% of global correct predictions rate and 9 LVs). Values are in %.

Predicted Grapevine Varieties	Real Grapevine Varieties				
	Loureiro	Pedernã	Alvarinho	Trajadura	Sum
Loureiro	**60.0**	1.8	0.0	0.4	62.2
Pedernã	1.3	**17.0**	0.0	0.0	18.3
Alvarinho	0.1	0.0	**9.7**	0.0	9.8
Trajadura	0.0	0.0	0.0	**9.8**	9.8
Sum	61.3	18.8	9.7	10.2	100

Table 3.6.7. PLS-DA correct predictions for grapevine species differentiation, for each variety, based on the global model. Values are in %.

Vineyard	Variety	Correct predictions (%)
Quinta deAzevedo	Loureiro	97,8
	Pedernã	90,5
	Alvarinho	100
	Trajadura	95,9
Quinta do Mourão	Touriga Nacional	95,6
	Tinta Barroca	98,8
	Touriga Franca	98,1
Quinta dos Carvalhais	Verdelho	97,4
	Encruzado	99,7
	Jaen	95,2
	Alvarinho	97,5
	Sémillon	97,4
Herdade do Peso	Syrah	100
	Arinto	96,1
	Antão Vaz	91,7
	Chardonnay	94,6

There are not many previous studies on the use of NIRS for the ampelographic differentiation of grapevine species, but a few have revealed similar results to the ones obtained in this study. For instance, Gutiérrez and co-workers [38] analysed 400 different leaves of 20 different grapevine varieties and reported 100% of correct predictions for three varieties (Cabernet Franc, Cabernet Sauvignon and Touriga Nacional), correct predictions greater than 90% for six varieties: Albariño, Treixadura, Viognier, Grenache, Carmenere and Caladoc, while most of the remaining varieties exhibited correct predictions greater or equal than 75% and lower or equal than 90%. The same authors, in a subsequent study [17], were able to discriminate one hundred forty-one samples out of 159 (88.7% correct predictions) with Cabernet Sauvignon obtaining 100% of correct predictions. All other analysed varieties were above 80% mark, excluding the Viognier variety, which obtained a score of 75% correct predictions.

To outline some explanations for the model separation between grapevine species and understand which part of the NIR signal used in the PLS-DA discrimination models is more important, the regression coefficient vectors were plotted (Fig. 3.6.5). Results for all vineyards were quite similar. It is clear that there are two distinct wavenumber regions fundamental for the discrimination of grapevine species. These zones are comprised between 12936 – 7819 cm^{-1} and 5970 – 5438 cm^{-1}. In the first zone, the peaks around 12500 $cm^{-1}$are characterized by R-NH (amides) bonds (third overtone region) and could be related to proteins [39] and chlorophyll [40]. The peaks around 11000 cm^{-1} are characterized by Ar-OH (phenolic antioxidant), R-OH (alcohol), C-H bonds as well as water bonds and are possibly related with carbohydrates, more specifically cellulose and lignin [39, 40]. The region between 9000 and 8000 cm^{-1} approximately, is related with carbohydrates, chlorophyll and carotenoids [40], mainly due to C-H bonds (second overtone region). The second zone (5970 – 5438 cm^{-1}) is characterized by C-H and S-H (thiol) bonds (first overtone region) and could be related to phenolic compounds and carbohydrates (cellulose and lignin) [39, 40]; the peaks found in that zone are most probably related to water bonds. This regression coefficient vectors analysis reveals that carotenoids, carbohydrates and chlorophyll seem to be the most important compounds for grapevine species ampelographic characterisation.

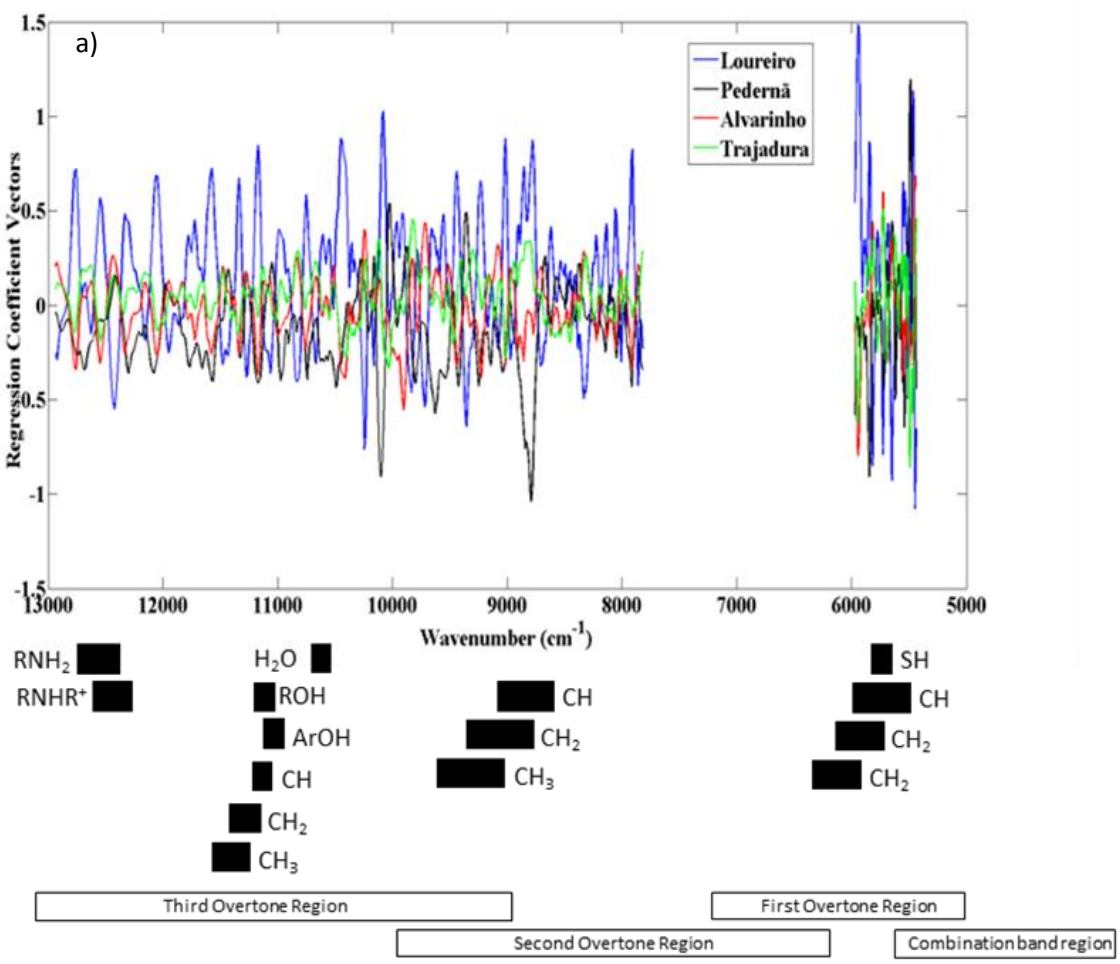


Figure 3.6.5. Representation of the regression coefficient vectors of the PLS-DA models developed for grapevine variety identification in each vineyard. a) *Quinta de Azevedo*; b) *Quinta do Mourão*; c) *Quinta dos Carvalhais*; d) *Herdade do Peso*.

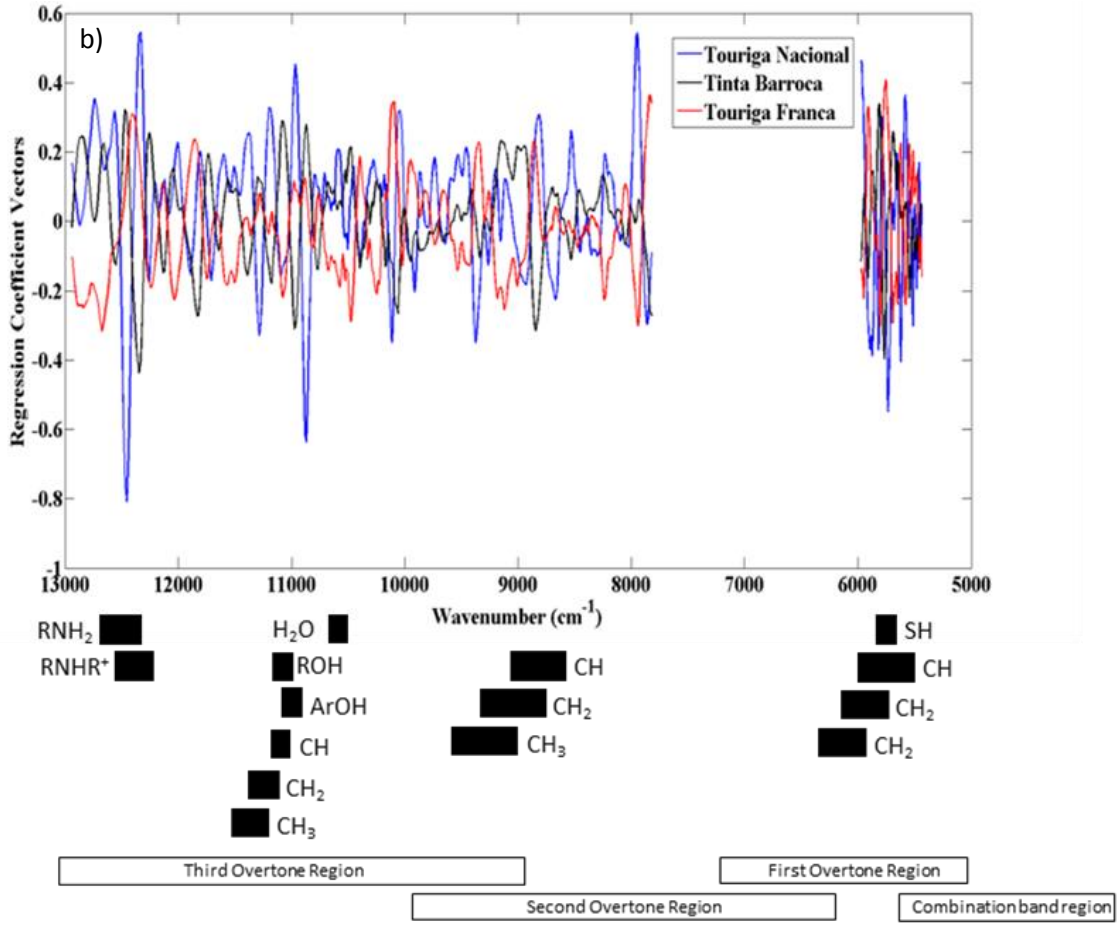


Figure 3.6.5. **(cont.).** Representation of the regression coefficient vectors of the PLS-DA models developed for grapevine variety identification in each vineyard. a) *Quinta de Azevedo*; b) *Quinta do Mourão*; c) *Quinta dos Carvalhais*; d) *Herdade do Peso*.

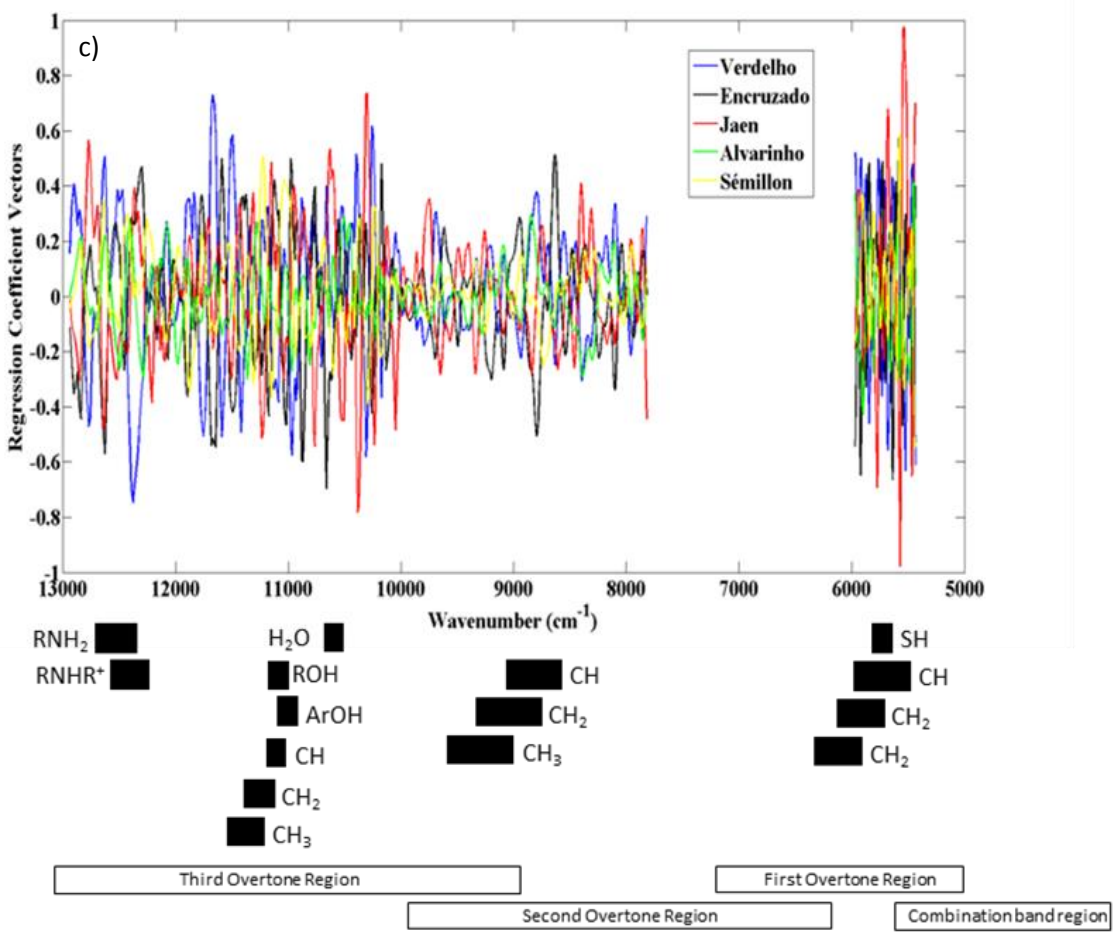


Figure 3.6.5. **(cont.).** Representation of the regression coefficient vectors of the PLS-DA models developed for grapevine variety identification in each vineyard. a) *Quinta de Azevedo*; b) *Quinta do Mourão*; c) *Quinta dos Carvalhais*; d) *Herdade do Peso*.

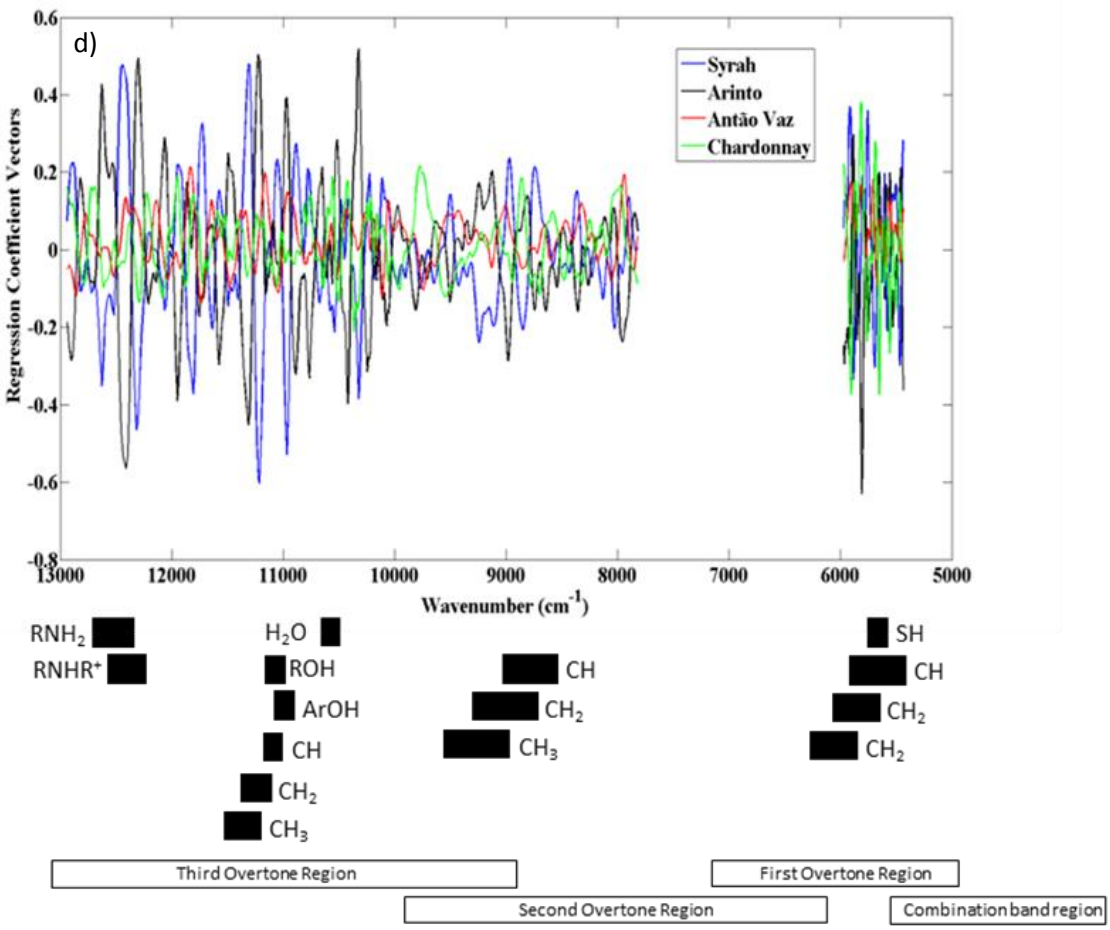


Figure 3.6.5. **(cont.).** Representation of the regression coefficient vectors of the PLS-DA models developed for grapevine variety identification in each vineyard. a) *Quinta de Azevedo*; b) *Quinta do Mourão*; c) *Quinta dos Carvalhais*; d) *Herdade do Peso*.

3.6.4. Conclusions

The present study proposes the use of a fast, reliable, environmental friendly and cost-effective methodology to assert plant phenotypes in vineyards of four different wine regions of Portugal. A total of 16 different grapevine species were analysed using both unsupervised (PCA) and supervised (PLS-DA) methods. The applied methodology yielded spectral models with accurate grapevine variety differentiation above 90%. Furthermore, an intensive monitoring of a single grapevine leaf was performed to understand if the spectral data of the leaf's surface was different according to the positioning of the spectrometer probe on the leaf. The resulting image revealed that there are indeed differences. However, further studies are needed to understand if these differences are significant for the building of the model or relevant from a biological or vinicultural point of view. Overall results for all the specific vineyards

analysed were extremely positive (all above 95% of correct predictions), proving that NIRS can be a strong alternative in the ampelographic differentiation of grapevine species, but possibly also for other plants.

The successful emergence and development of new technologies such as NIRS into different science fields have made their way to industries as diverse as wine, coffee, soil and is also starting to intrude onto plant phenotyping. Using such technologies it is possible to perform phenotyping tasks with reduced time and monetary costs, making it extremely attractive to the industrial sector. Viticulture can, no doubt, benefit from the latest research results that have been developed for a number of wine and vineyard related problems using NIRS.

3.6.5. References

1. Walter, A., F. Liebisch, and A. Hund, *Plant phenotyping: from bean weighing to image analysis.* Plant methods, 2015. **11**(1): p. 14.
2. Minervini, M., H. Scharr, and S.A. Tsaftaris, *Image analysis: the new bottleneck in plant phenotyping [applications corner].* IEEE Signal Processing Magazine, 2015. **32**(4): p. 126-131.
3. Galet, P., *A practical ampelography*. 1979: Cornell University Press.
4. Zarco-Tejada, P.J., et al., *Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy.* Remote Sensing of Environment, 2005. **99**(3): p. 271-287.
5. Steele, M., A.A. Gitelson, and D. Rundquist, *Nondestructive estimation of leaf chlorophyll content in grapes.* American Journal of Enology and Viticulture, 2008. **59**(3): p. 299-305.
6. Workman, J. and L. Weyer, *Practical Guide to Interpretive Near.* Infrared Spectroscopy, 2008. **1**: p. 1-21.
7. Wang, W. and J. Paliwal, *Spectral data compression and analyses techniques to discriminate wheat classes.* Trans. ASABE, 2006. **49**(5): p. 1607-1612.
8. Fu, X., et al., *Discrimination of pear varieties using three classification methods based on near-infrared spectroscopy.* Trans. Asabe, 2007. **50**(4): p. 1355-1361.
9. Li, X., Y. He, and H. Fang, *Non-destructive discrimination of Chinese bayberry varieties using Vis/NIR spectroscopy.* Journal of Food Engineering, 2007. **81**(2): p. 357-363.
10. Sánchez, M.-T., et al., *Non-destructive characterization and quality control of intact strawberries based on NIR spectral data.* Journal of Food engineering, 2012. **110**(1): p. 102-108.
11. Xu, H.-r., et al., *On-site variety discrimination of tomato plant using visible-near infrared reflectance spectroscopy.* Journal of Zhejiang University Science B, 2009. **10**(2): p. 126-132.
12. Fernandez-Novales, J., et al., *Assessment of quality parameters in grapes during ripening using a miniature fiber-optic near-infrared spectrometer.* International journal of food sciences and nutrition, 2009. **60**(sup7): p. 265-277.
13. Pérez-Marín, D., et al., *Miniature handheld NIR sensor for the on-site non-destructive assessment of post-harvest quality and refrigerated storage behavior in plums.* Journal of Food Engineering, 2010. **99**(3): p. 294-302.

14. Carter, G.A. and B.A. Spiering, *Optical properties of intact leaves for estimating chlorophyll concentration.* Journal of environmental quality, 2002. **31**(5): p. 1424-1432.
15. Abdel-Rahman, E.M., F.B. Ahmed, and M. Van den Berg, *Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy.* International Journal of Applied Earth Observation and Geoinformation, 2010. **12**: p. S52-S57.
16. Santos, A.O. and O. Kaye, *Grapevine leaf water potential based upon near infrared spectroscopy.* Scientia Agricola, 2009. **66**(3): p. 287-292.
17. Gutiérrez, S., et al., *Data mining and NIR spectroscopy in Viticulture: Applications for plant phenotyping under field conditions.* Sensors, 2016. **16**(2): p. 236.
18. Govender, M., et al., *Review of commonly used remote sensing and ground-based technologies to measure plant water stress.* Water SA, 2009. **35**(5): p. 741-752.
19. Páscoa, R., et al., *Exploratory study on vineyards soil mapping by visible/near-infrared spectroscopy of grapevine leaves.* Computers and Electronics in Agriculture, 2016. **127**: p. 15-25.
20. Cochrane, M., *Using vegetation reflectance variability for species level classification of hyperspectral data.* International Journal of Remote Sensing, 2000. **21**(10): p. 2075-2087.
21. Schmidt, K. and A. Skidmore, *Spectral discrimination of vegetation types in a coastal wetland.* Remote sensing of Environment, 2003. **85**(1): p. 92-108.
22. Cemin, G. and J.R. Ducati, *Spectral discrimination of grape varieties and a search for terroir effects using remote sensing.* Journal of wine research, 2011. **22**(1): p. 57-78.
23. Parton, D., et al. *A temporal investigation of leaf spectroscopy for grape variety discrimination*. in *Hyperspectral Image and Signal Processing (WHISPERS), 2012 4th Workshop on*. 2012. IEEE.
24. Naes, T., et al., *Interpreting PCR and PLS solutions*, in *A User-Friendly Guide to Multivariate Calibration and Classification*. 2004, NIR Publications: Chichester, UK. p. 39-54.
25. Perez, D., et al., *Authentication of green asparagus varieties by near-infrared reflectance spectroscopy.* JOURNAL OF FOOD SCIENCE-CHICAGO-, 2001. **66**(2): p. 323-327.
26. Naes, T., et al., *A user friendly guide to multivariate calibration and classification*. 2002: NIR publications.

27. Rinnan, Å., F. van den Berg, and S.B. Engelsen, *Review of the most common pre-processing techniques for near-infrared spectra.* TrAC Trends in Analytical Chemistry, 2009. **28**(10): p. 1201-1222.
28. Barker, M. and W. Rayens, *Partial least squares for discrimination.* Journal of Chemometrics, 2003. **17**(3): p. 166-173.
29. Geladi, P. and B.R. Kowalski, *PARTIAL LEAST-SQUARES REGRESSION - A TUTORIAL.* Analytica Chimica Acta, 1986. **185**: p. 1-17.
30. Preisner, O., et al., *Application of Fourier Transform Infrared Spectroscopy and Chemometrics for Differentiation of Salmonella enterica Serovar Enteritidis Phage Types.* Applied and Environmental Microbiology, 2010. **76**(11): p. 3538-3544.
31. Borrelli, P., et al., *Assessment of the cover changes and the soil loss potential in European forestland: First approach to derive indicators to capture the ecological impacts on soil-related forest ecosystems.* Ecological Indicators, 2016. **60**: p. 1208-1220.
32. Porra, R., W. Thompson, and P. Kriedemann, *Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy.* Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1989. **975**(3): p. 384-394.
33. Batovska, D., et al., *Seasonal variations in the chemical composition of vine-grape leaf surface.* Natural product research, 2008. **22**(14): p. 1231-1236.
34. Vanloot, P., et al., *Artificial vision and chemometrics analyses of olive stones for varietal identification of five French cultivars.* Computers and Electronics in Agriculture, 2014. **102**: p. 98-105.
35. Lopo, M., et al., *Classification of Vineyard Soils Using Portable and Benchtop Near-Infrared Spectrometers: A Comparative Study.* Soil Science Society of America Journal, 2016. **80**(3): p. 652-661.
36. Sarraguça, M.C., et al., *Authenticity control of roasted coffee brands using near-infrared spectroscopy.* Food Analytical Methods, 2013. **6**(3): p. 892-899.
37. Hagg, C., F. Stober, and H. Lichtenthaler, *Pigment content, chlorophyll fluorescene and photosynthetic activity of spruce clones under normal and limited mineral nutrition.* Photosynthetica, 1992. **27**(3): p. 385-400.
38. Gutiérrez, S., et al., *Support vector machine and artificial neural network models for the classification of grapevine varieties using a portable NIR spectrophotometer.* PloS one, 2015. **10**(11): p. e0143197.

39. Kokaly, R.F. and R.N. Clark, *Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression.* Remote sensing of environment, 1999. **67**(3): p. 267-287.
40. Meyer, B.S., D.B. Anderson, and R.H. Böhning, *Introduction to plant physiology.* Introduction to plant physiology., 1960.

Chapter IV: Concluding Remarks and Future Perspectives

The main objective of this thesis was to demonstrate the possibilities and advantages of using infrared spectroscopy, mainly near infrared and mid infrared, combined with chemometric techniques, during certain important steps in the wine industry, mainly related to vineyard management. The development and validation of infrared spectroscopy methodologies in such processes could provide solutions with environmental and financial advantages, when compared to reference methods.

4.1. Conclusions

The work presented in this thesis demonstrated the application of IR spectroscopy as a monitoring tool to assess multiple quality indicators of vineyards (soil and plants). Soil identification and classification with the intent of vineyard soil mapping was accomplished both directly and indirectly (using grapevine leaves spectra) and *in-situ* and in the laboratory using portable and benchtop spectrometers. Important chemical, biological and physical soil parameters for the growth and health of the grapevine were also determined using both NIR and MIR, with MIR exhibiting better predictive results. A thorough comparison of the performance of several portable and handheld instruments was carried out to understand which spectrometer would exhibit better results in a portability/performance relationship for the aforementioned processes. Grapevine ampelographic differentiation was also attempted and successfully accomplished, meaning that IR spectroscopy can also play an important role in the identification and plantation/replantation strategies of grapevine varieties. All mentioned processes are undoubtedly paramount for the sound management of a vineyard and hence the prosperous thriving of wine production.

The NIR potential for soil discrimination was addressed by soil mapping two vineyards from different wine regions in Portugal. A strategy using chemometric methods such as PCA and PLS-DA revealed that dried-ground soil samples presented better results, but not significantly different when compared with wet or dried samples unground samples. Discriminant models showed that NIRS is able to discriminate the different vineyard soil types, reproducing very accurately the mapping generated by pedology methods (79.3% of correct predictions). Furthermore, variations within the same soil type (present at different locations in the vineyard) were also detected and it was found that, for soil classification, there is no significant difference in using samples from different depths, at least for the depths studied.

A comparative study was performed to assess the difference in accuracy between a benchtop and a portable NIR spectrometer for vineyard soil classification. The same soil from the same vineyard was analysed with both instruments with the objective of understanding if a portable spectrometer would yield satisfactory results for *in-situ* measurements. Results showed, quite surprisingly that the performance of the portable instrument was equivalent to that of the benchtop spectrometer. Modelled spectra revealed, through PLS-DA, that 75 to 100% successful soil identification rates are achieved when samples are collected within the same vineyard block and that lower prediction percentages (around 70-75%) are obtained when soils from the entire vineyard are analysed simultaneously.

The promising results from the previous study led to another, more thorough and complex comparative study, using a wider range of portable and handheld spectrometers and different vineyards. Three Fourier-transform infrared (FT-IR) and two visible and near infrared (vis-NIR) were used to analyse the same soil samples from vineyards in the McLaren Vale wine region in South Australia. The objective was to understand if the spectrometers were suitable for the analysis of important soil properties such as TN, TOC, pH, moisture and eCEC. Models were first developed using the entire spectral range followed by models using specific wavelengths. Results were quite varied for each spectrometer with PLS models exhibiting that, at least for the spectrometers used, there is no best equipment for all the different soil constituents analysed. Nevertheless, the FT-IR instruments had an overall better performance than the vis-NIR spectrometers. Furthermore, models developed using specific regions of the spectra yielded better results than models constructed with the whole spectral range.

The suitability of two handheld IR spectrometers, one Fourier-transform infrared (FT-IR) and a vis-NIR one, for the prediction of important soil properties for grapevine's growth was assessed. Parameters such as phosphorous (P), potassium (K), sulphur (S), conductivity, pH ($CaCl_2$), calcium carbonate, chloride, exchangeable cations: calcium (exch. Ca), potassium (exch. K), magnesium (exch. Mg), sodium (exch. Na) and exchangeable sodium percentage (ESP) were estimated. It was also investigated if the use of these hand held spectrometers was adequate to accurately classify different soil types and thus enable the possible development of a robust soil mapping. The PLS models revealed R^2 values for parameters ranging from 0.27 to 0.99 using both MIR and NIR instruments. Soil discriminant analysis was also successfully achieved through PLS-DA with the FT-IR spectrometer yielding 74.9% of correct predictions and the vis-NIR one 67.2% of correct predictions when classifying the same soils. These results corroborate the results obtained in the first works.

Indirect soil classification was also attempted through the analysis of grapevine leaves. *In-situ* spectra of grapevine leaves were collected in two different vineyards with the intent of understanding if it was possible to discriminate vineyard soils just by analysing leaves spectra. Models build using PLS-DA returned around 95% of correct soil taxonomic predictions. This methodology was then applied to monitor all plants within a 0.3ha vineyard block in one of the vineyards, resulting in a highly detailed soil taxonomic map built exclusively from leaves Vis/NIR spectra. A comparison with the existing soil map proved that the NIR spectroscopy based estimation was not only extremely reliable and accurate but also a more detailed than traditional pedological soil maps.

Grapevine leaves are no doubt of the utmost importance for vineyard overall status and hence, management, but their importance also extends to the ampelographic differentiation of cultivars. Grapevines leaves of 15 different varieties from four different vineyards, in four different Wine Regions of Portugal (Alentejo, Dão, Douro, Vinhos Verdes) were scanned *in-situ* using a portable vis/NIR spectrometer. The spectra were processed using chemometric tools such as PCA and PLS-DA with the objective of understanding if NIRS could successfully identify different grapevine varieties and become a rapid, *in-situ* alternative, to costly, time-consuming genetic techniques. Results have shown an excellent correlation between grapevine leaves and vine species with percentages of correct predictions above 90%.

The work developed during this thesis had the objective of understanding and demonstrate the potential of infrared spectroscopy for vineyard management, mainly through the use of portable instruments that can be used for *in-situ* routine monitoring of several processes. The methodologies presented, could indeed become benchmark tools for vineyard control and management with both financial and environmental advantages when compared to reference methods.

4.2. Future Perspectives

The work developed during this thesis opened perspectives of future work for the development of infrared spectroscopy in vineyard management but also in the wine industry. The following items summarise some of those future perspectives:

- to assist in the plantation or replantation of varieties in specific soil types with the objective of producing grapes/wines with specific characteristics;

- to understand which correlations between soil and plant provide trademark characteristics that can be used as tool to direct wine production;
- to establish correlation between soil mineral properties and wine characteristics;
- to establish better, more effective correlations between spectra and soil/plant composition;
- to use IR spectroscopy to assess important specific parameters for the health of the grapevine such as water stress;
- to study the possible correlations between the taxonomic soil type and the characteristics of the grape within a single grapevine variety;
- to understand if there are correlations between the multielemental compositon of grape components (seeds, skin, and berry) and the soil in which the grapevine is planted;
- to use IR spectroscopy in intensive monitoring of the vineyard for thorough characterisation that may help in vineyard management, planning but also harvest preparation and
- to resource on IR spectroscopy to assess composition and nutritional variability of a given grapevine variety and thus evaluate its yield.