The Rapid Characterisation of Natural Products and Herbal Medicines by Near-Infrared Spectroscopy

MARIKO KUDO

A thesis submitted in partial fulfilment of the requirements of

The University of London

for the degree of Doctor of Philosophy

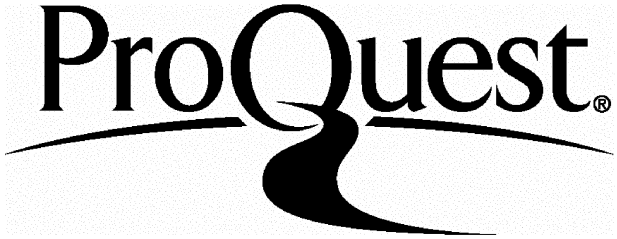
in the Faculty of Medicine

The School of Pharmacy, University of London,

29/39 Brunswick Square,

London WC1N 1AX.

2004



To Charlotte

Acknowledgements

I would like to thank my supervisors at the School of Pharmacy, Professor Tony Moffat and Dr. Robert Watt for making this project possible. As well as providing me with the opportunity to attend numerous conferences, which I found useful, they have been helpful in giving me invaluable advice and guidance.

I would also like to thank FOSS NIRSystems and Mrs. Sheelagh Halsey for providing the instrument that I used for the majority of the project. In addition to the instrument, they also provided me with the Vision software and numerous useful pieces of advice. My thanks also go to NDC Infrared Engineering and Mr. Clive Rowe for the loan of a second NIR instrument.

My colleagues at the School of Pharmacy also are greatly appreciated for making the lab an interesting and vibrant environment to work in. Special thanks go to Dr. Andrew O'Neil for his special support and encouragement over the years.

Finally, all my love and gratitude go to my parents, Professor Norikazu Kudo and Mrs. Sachiko Kudo for their support over the past 28 years, and to my sister Machiko for her countless words of wisdom and encouragement.

Abstract

The Rapid Characterisation of Natural Products and Herbal Medicines by Near-Infrared Spectroscopy

Although the use of herbal medicines has increased in popularity, there is very little quality control carried out on them. Thus, in the industry, potentially lethal problems such as misidentification and adulteration can occur. Traditional methods of analysis, including chromatography and microscopy, can be destructive and time-consuming. Near-infrared spectroscopy (NIRS), however, would be advantageous in that it is rapid, easy, and requires little or no sample preparation.

The potential for NIRS to analyse natural products was explored. Results showed that the technique could be used to successfully identify different samples within a family (e.g. *Umbelliferae*) as well as within a genus (e.g. *Digitalis*). It was also possible to identify various parts of the same plant, as well as detect adulteration of a sample with a different plant part. NIRS could also be used to successfully discriminate between samples of the same species of different geographical origins, such as *Cannabis sativa*. The use of NIRS in the food industry was also briefly examined, and it was found that it could be used to discriminate between various tea blends.

NIRS was also effective in measuring the moisture content of samples. This is advantageous over other techniques in that being rapid and non-destructive, it could easily be used on-line in the production process.

Various statistical and chemometric methods were investigated to allow for optimum power of identification and sample discrimination. The Maximum Distance in Wavelength Space method was the most successful one, although others, such as Correlation in Wavelength Space and correlation coefficients also had limited success. Manual techniques, such as the two-wavelength method and the Polar Qualification method were also investigated.

Overall, it could be said that NIRS could be a powerful tool for the analysis of natural products, with its speed and non-destructiveness being a major advantage.

TABLE OF CONTENTS

CHAPTER 2: Materials and Methods

CHAPTER 3: The Use of Near-Infrared Spectroscopy for the Analysis of Herbal Materials- Preliminary Studies

CHAPTER 7: An Investigation into Some Leaf Components Using Near-Infrared Spectroscopy

CHAPTER 8: The Discrimination of Commercial Teas Using Near-Infrared Reflectance Spectroscopy

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial neural network
AOTF	Acousto-optically tuneable filter
ASTM	American Society for Testing and Materials
BP	British Pharmacopoeia
CBN	Cannabinol
CBD	Cannabidiol
DT	De-trend
FT-NIR	Fourier-Transform Near-Infrared
GC	Gas chromatography
HPLC	High performance liquid chromatography
KF	Karl Fischer
***k*NN**	*k*-Nearest Neighbour
MLR	Multiple linear regression
NIR	Near-infrared
NIRS	Near-infrared spectroscopy
PCA	Principal components analysis
PLS	Partial least squares
PQS	Polar Qualification System
RCA	Rapid Content Analyser
SNV	Standard normal variate
THC	Tetrahydrocannabinol
TLC	Thin layer chromatography

Chapter 1: Introduction

1.1. Background

The use of herbal medicines is becoming increasingly popular in recent times. However, there are as yet very few quality control procedures carried out on them. Thus, in industry, potentially lethal problems such as misidentification and adulteration can occur. Also, traditional methods of analysis, including chromatography and microscopy, can be destructive and time-consuming. Near-infrared spectroscopy (NIRS), on the other hand, would be advantageous in that it is rapid, easy, and requires little or no sample preparation. The use of NIRS for the analysis of natural products has so far been little documented or attempted. This may be partly due to the fact that vegetable materials can vary unpredictably in response to a number of factors, including temperature, rainfall and age, making the setting up of references difficult. Some early studies included the discrimination of ginseng and grape seeds (Corti et al 1990), the assignation of herbal medicines such as *Cassia*, *Ganoderma*, *Smilacis Rhizoma* and *Astragali Radix* according to geographical origins (Woo et al 1998, Woo et al 1999) and the determination of the essential oil content of certain herbal materials (Fehrmann et al 1996).

Light absorption in the NIR is primarily due to overtones and combinations of fundamental vibration bands occurring in the mid-infrared region (Blanco et al 1998). As a result, peaks appear neither well defined nor sharp, and absorbances in the NIR region are much smaller than those in the mid-infrared by as much as a factor of 1000 (Khan et al 1997). As a result, unlike mid-infrared spectroscopy, it is virtually impossible to characterise samples by mere visual inspection of spectra

(Blanco et al 1998). As a result, complex, computer based statistical analysis (chemometrics) is necessary in order to analyse the resulting spectra. With the development of modern NIR instrumentation and sophisticated software packages, however, NIRS has been more important in later years. The low absorptivities of absorption bands in the NIR are compatible with moderately concentrated samples and longer path lengths compared to the mid-infrared region (US Pharmacopeoia 1998). These longer path lengths enable spectra to be measured even through intact materials, and thus NIRS is advantageous in that it is rapid, non-destructive, and requires little or no sample preparation (Blanco et al 1998). Thus, most samples, including herbal materials, can theoretically be analysed as they are. In addition, NIRS can provide simultaneous information about the chemical composition of materials as well as their physical state including moisture and particle size (Moffat et al 1997).

This thesis explores the potential of NIRS for the identification and characterisation of plant materials, mainly of pharmaceutical interest. The main identification methods investigated were Maximum Distance in Wavelength Space, Correlation in Wavelength Space, a two-wavelength analysis, and the Polar Qualification System. These were used due to their wide use and also to their mathematical simplicity. It is hoped that any techniques established could be applied to the analysis of herbal medicines.

1.1.1 Theory of NIRS

The NIR region lies between the visible and mid-infrared regions of the electromagnetic spectrum and is defined by the American Society for Testing and

Materials (ASTM) as the spectral region spanning the wavelength range 780-2526nm (Blanco et al 1998). William Herschel recorded the first NIR spectrum in 1800 when he continued his measurements of the heat energy of solar emission beyond the red portion of the visible spectrum. In recognition of this discovery, it has been proposed that the NIR region between 780 and 1100nm should be named the Herschel infrared (Davies 1990, Osborne et al 1993a).

At the basic level, spectroscopy is a technique based on the vibrations of the atoms of a molecule. A spectrum is obtained by passing radiation through a sample and determining what fraction of the incident radiation is absorbed at a particular energy. The energy at which any peak in an absorption spectrum appears corresponds with the frequency of a vibration of a part of a sample molecule (Stuart et al 1996).

Electromagnetic radiation can be thought of as a stream of particles or quanta for which the energy, E, is given by the Bohr equation, below:

$$E = h\nu \tag{1.1}$$

where h is Planck's constant (6.626×10^{-34} Js^{-1}) and ν is the frequency of the absorbed radiation.

Processes of electronic change can be shown in terms of quantised discrete energy levels E_0, E_1, E_2.....etc (Figure 1.1)

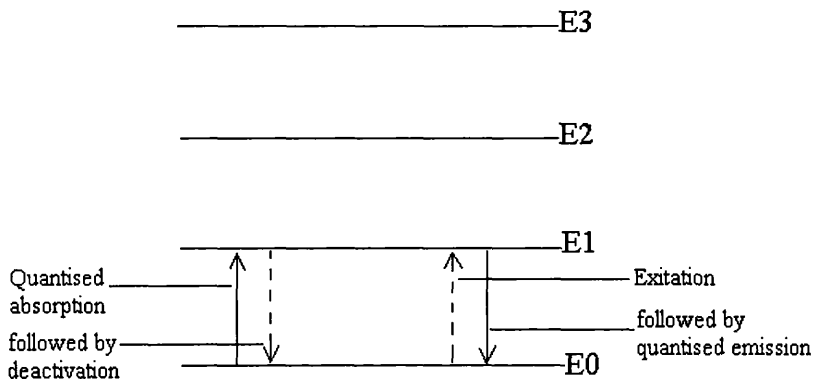


Fig. 1.1. Energy levels

In a large assembly of molecules, all atoms or molecules will be distributed among these various energy levels. Whenever a molecule interacts with radiation, a quantum of energy (photon) is either absorbed or emitted. In each case, the energy of the radiation quantum equals the energy gap, i.e. $E_1 - E_0$, $E_2 - E_1$, etc. That is,

$$\Delta E = h\nu = E_1 - E_0 \tag{1.2}$$

Therefore, the frequency of emission or absorption of radiation for a transition between the energy levels E_1 and E_0 can be given by the following:

$$\nu = \frac{E_1 - E_0}{h} \tag{1.3}$$

Associated with the uptake of energy (absorption) is a deactivation mechanism in which the atom or molecule returns to its original state. Associated with loss of

energy is some prior exitation mechanism. These mechanisms are shown by the dotted lines in Figure 1.1(Stuart et al 1996).

For the absorption of radiation to occur the lower energy level must be populated, with the greater population leading to a more intense absorption. The absorption of NIR radiation is associated with changes in the vibrational and rotational energy of molecules. Thus in order for the molecule to interact with the radiation, there must be a change in the electric dipole moment when the atoms are displaced. Vibrations in which there is no dipole moment change will be NIR inactive (Banwell, 1996).

1.1.2 Vibrations of Molecules

A molecule can be thought of as a system of masses joined by bonds with spring-like properties. Diatomic molecules have three degrees of translational freedom and two degrees of rotational freedom. In addition, atoms can move relative to each other. That is, bond lengths can vary or one atom can move out of its present plane. These stretching and bending movements are collectively known as vibrations. There is only one vibration possible for a diatomic molecule, which corresponds to the stretching and compression of the bond. This means that there is one degree of vibrational freedom. In polyatomic molecules containing N atoms, several degrees of vibrational freedom exist as bonds may stretch and angles may bend. These are $3N$-5 for any linear molecule and $3N$-6 for any non-linear molecule (Stuart et al 1996).

The stiffness of the bond and the masses of the atoms at each end of the bond are also used to explain the frequency of vibrational modes. The bond stiffness can be characterised by a proportionality constant called the force constant, k. k is derived from Hooke's law that assumes that for the vibration of a diatomic molecule, the potential energy curve may be approximated to a parabola. The reduced mass, μ, provides a useful way of simplifying calculations by combining the individual atomic masses, and may be represented as follows:

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \tag{1.3}$$

The equation relating force constant, reduced mass and the frequency of absorption is as follows:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{1.4}$$

This can be modified so that direct use of the wavenumber values for bond vibrational frequencies can be made:

$$\nu = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}} \tag{1.5}$$

where c is the speed of light in a vacuum ($2.997925 \times 10^8\ ms^{-1}$).

The vibrational energies are given by the following:

$$E_{vibration} = (v + \tfrac{1}{2})h\nu_0 \tag{1.6}$$

The vibrational quantum number, v, can take the values 1,2,3.....This represents a series of evenly spaced energy levels in the molecule, with the selection rule being $\Delta v = \pm 1$. Thus, the absorption transition for $E_{1\leftarrow 0}$, for example, would be:

$$E_{1\leftarrow 0} = (1 + \tfrac{1}{2})hv - \tfrac{1}{2}hv = hv_0 \tag{1.7}$$

and so on (Banwell 1996).

It is only possible for a molecule to absorb radiation when the incoming infrared radiation is of the same frequency as one of the fundamental modes of vibration of the molecule. This means that the vibrational motion of a small part of the molecule is increased while the rest is left unaffected The complexity of an infrared spectrum arises from the coupling of vibrations over a large part, or over, the whole molecule. Bands associated with these vibrations are likely to conform to a pattern or fingerprint of the molecule as a whole, rather than a specific group within the molecule (Stuart et al 1996).

In practice, the vibrations of polyatomic molecules tend to be non-harmonic. That is, vibrations about the equilibrium position are not symmetric, with the potential energy curve for actual bonds being only roughly parabolic (Blanco et al 1998). This is because although real bonds are elastic, repulsion between atoms causes the potential energy to rise more rapidly than predicted by the harmonic approximation, resulting in bond breakage (Osborne et al 1993a). To rectify this, one can include additional terms of higher order than those used in Hooke's law. Therefore, the energy E_v for each level can be given by:

$$E_v = f_e(v + \tfrac{1}{2}) - f_e x_e(v + \tfrac{1}{2}) + h$$

(1.8)

Here, v is the vibrational quantum number, x_e is the anharmonicity constant (i.e. the deviation of the potential function from the parabola), f_e is the uniform spacing between levels corresponding to a parabola with its centre at the equilibrium distance and the same curvature as the real potential energy function, and h is a higher order term (Blanco et al 1998).

One consequence of this is that the selection rule becomes $\Delta v = \pm 1, \pm 2, \ldots$etc. As a result, transitions such as $v_{2\leftarrow 0}$ and $v_{3\leftarrow 0}$ are now possible, and in addition to the fundamental band (+1), other, higher frequencies known as overtones appear as multiples of the fundamental frequency. For many molecules, the energies of these overtones lie in the NIR region (Khan et al 1997). The energy required for the first overtone is twice the fundamental, assuming that energy levels are equally spaced. Since the energy is proportional to the frequency absorbed and this is proportional to the wavenumber, the first overtone will appear in the spectrum at twice the wavenumber of the fundamental (Stuart 1996). As the first overtone is approximately $1/10^{th}$ and the second overtone $1/100^{th}$ the intensity of the fundamental absorption, NIR spectra can be recorded directly on undiluted samples.

Polyatomic molecules may display simultaneous changes in the energies of two or more vibrational modes, leading to combination and subtraction bands. Combination bands arise when two fundamental bands absorbing at $v_1 + v_2$ absorb

energy simultaneously. The resulting weak band will appear at (v_1 +v_2) wavenumbers (Stuart et al 1996). Combination bands are of very low probability unless they arise from no more than two vibrations involving bonds that are either connected through a common atom or multiple bonds. Subtraction bands, which are caused by absorption by molecules residing in excited vibrational states, are of very low probability at room temperature (Osborne et al 1993a).

1.2 NIR diffuse reflectance spectroscopy

Random reflections, refractions, and scatter at various interfaces inside the sample diffuse NIR light that penetrates a powder's surface before it emerges back through the surface (Osborne 1993a). This is known as diffuse reflectance (Figure 1.2). The low molar absorptivity of solids in the NIR region significantly restricts sensitivity. However, it allows operation in diffuse reflectance mode and therefore the recording of spectra for solid samples (Blanco et al 1998).

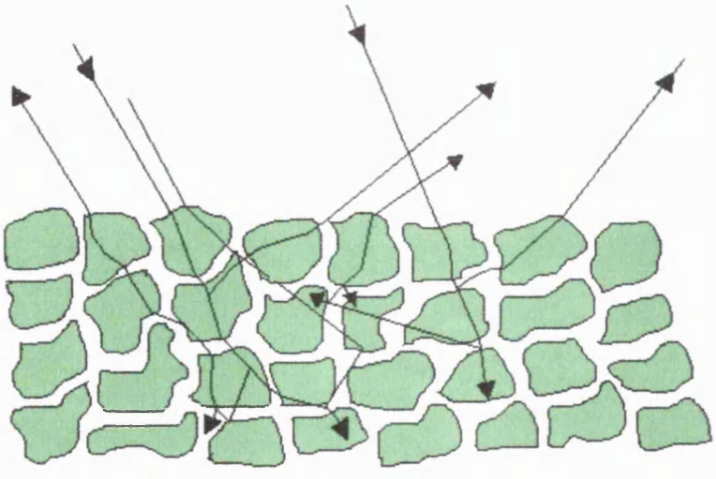


Figure 1.2 Diffuse reflectance

Reflectance spectroscopy measures the light reflected by the sample surface, which contains a specular component (Figure 1.3) and a diffuse component. Little information about composition is contained in specular reflectance, and thus its contribution to measurements is minimised by adjusting the detector's position relative to the sample (Blanco et al 1998). Specular reflectance is a mirror reflection, and is described by the Fresnel equation (Kortum, 1969):

$$R_{spec} \equiv \frac{I_{refl}}{I_O} = \frac{(n-1)^2 + n^2k^2}{(n+1)^2 + n^2k^2} \tag{1.9}$$

where I_0 is the intensity of the incident radiation, I_{refl} is the intensity of the reflected radiation, k is the absorption coefficient and n is the refractive index.

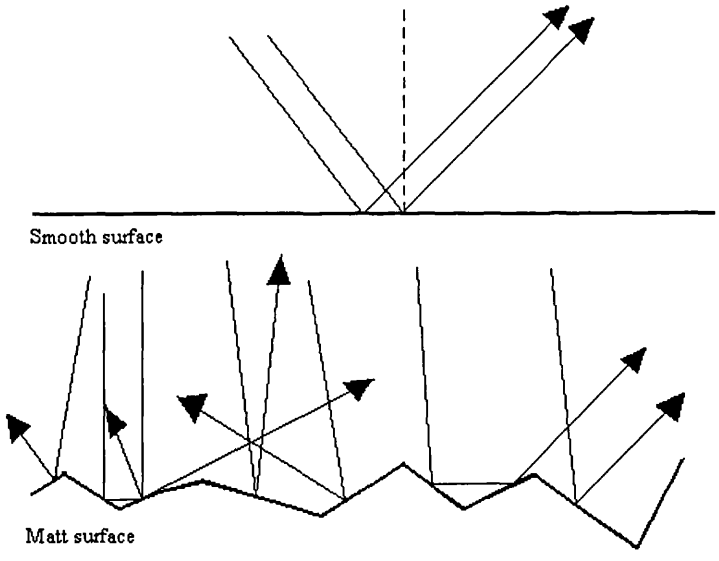


Figure 1.3 Specular reflectance

If the surface of a sample is matt, the boundary between the sample and the surrounding medium may be said to comprise of a series of small interfaces

orientated at all possible angles to the normal. Thus, although reflection at each interface follows the Fresnel equation, the effect is to diffuse the specular component of reflectance. For most materials excluding metals, about 4 percent of the incident radiation undergoes specular reflectance. As a result, there is a level of superimposition on diffuse reflectance arising from internal effects (Osborne et al 1993a). The interference caused by specular reflection can be eliminated partly by using powders with small particle size and by diluting the absorbing species with suitable diluents (Kortum 1969).

However, diffuse reflectance and tranmission are the basis of measurements by NIRS. There are no rigorous rules for diffuse reflectance, but a number of theories have been proposed. The most popular one is the Kubelka-Munk theory, which proposes that the power of reflected radiation could be described by two constants s and k, or the scattering and absorption constants, respectively. A layer of infinite thickness which is completely opaque may be described thus (Osborne et al 1993a):

$$\frac{(1-R_\infty)^2}{2R_\infty} = \frac{k}{s} \tag{1.10}$$

where R_∞ is the reflectance of the infinitely thick layer. In practice, relative reflectance (R), or the ratio of the intensity of the light reflected by the material to that by a standard, is preferred to absolute reflectance. Typically, a standard is a stable material with a high and constant absolute reflectance. These include Teflon, barium sulphate, magnesium oxide, and high-purity alumina ceramics. In these standards, k is assumed to be zero and absolute reflectance to be one. That is,

the perfect standard is a material that absorbs no light at any wavelength and reflects light at an identical angle with the angle of incidence. As no single material meets these requirements, the standards used in this context are stable, homogeneous, non-transparent, non-fluorescent substances of high, fairly constant relative reflectance (Blanco et al 1998).

The Kubelka-Munk equation can be rewritten in terms of the relative reflectance and the absorbing analyte concentration (c):

$$f(R) = \frac{(1-R)^2}{2R} = \frac{k}{s} = \frac{\varepsilon c \ln_{10}}{s} = \frac{c}{a} \tag{1.11}$$

where ε is the molar absorptivity and $a = s/2.303e$. As with Beer's law, the Kubelka-Munk theory is only applicable to weak absorption bands, or when the product of absorptivity times concentration is small, as is so in the NIR region. However, deviations from the above equation result (Blanco et al 1998). If the absorption of the matrix is high, the effective depth of penetration is too small for the beam to be able to interrogate enough particles for the Kubelka-Munk assumptions to be valid. This is often the case in many cereal products, which are highly absorbing materials (Ollinger and Griffiths 1993).

A practical approach for this is to use NIR data either as raw data (relative reflectance, R; relative transmission T) or as apparent absorbance, A (Osborne et al 1993):

$$A = \log_{10} \frac{1}{R} = a'c \tag{1.12}$$

where c equals concentration and a' is a proportionality constant. With transmission measurements, transformation to A is given by:

$$A = \log_{10} \frac{1}{T} = a'c \tag{1.13}$$

While this transformation has no theoretical basis on the Kubelka-Munk equation, it provides very adequate results when used under the particular conditions used in many diffuse reflectance spectroscopic applications (Blanco et al 1998).

1.3 NIR instrumentation

The basic instrument configurations for transmittance and reflectance measurements are shown in Figure 1.4.

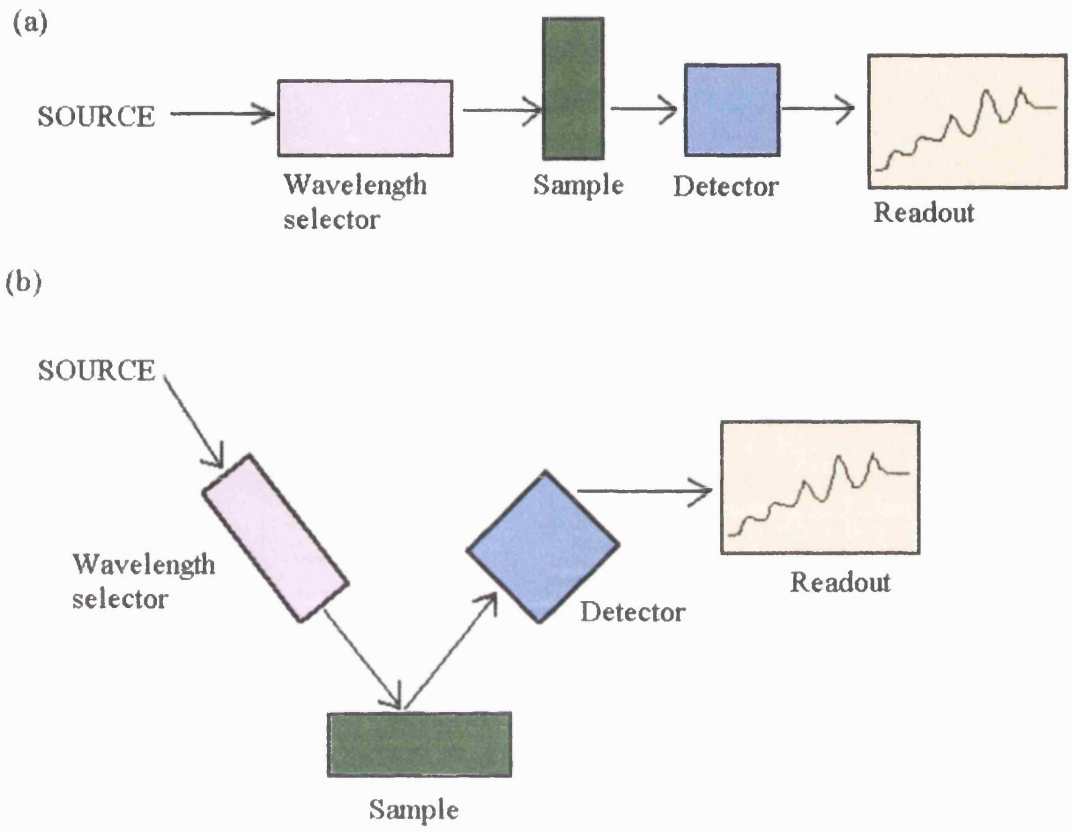


Figure 1.4 Basic dispersive instrument configurations for (a) transmittance and (b) reflectance NIR measurements

All NIR instruments are performed by passing light radiation through a sample and measuring the emerging (either transmitted or reflected) light intensity. There are numerous spectrometers that are based in different operating principles. Instrument design has been greatly influenced by modern technology, such as microprocessors, lasers, solid state modulators and optical fibres have had a massive impact on instrument design. Long-established techniques, such as Michelson interferometry have been utilised practically, while new dispersing

elements, such as acousto-optically tuneable filters (AOTFs) are being increasingly used as alternatives to diffraction gratings and interference filters. Data analysis has also constantly been made easier by powerful software and chemometric techniques (Osborne et al 1993).

1.3.1 Instrument requirements

While NIR instruments may vary according to manufacturer, they are all alike in that they aim to adhere to the following basic requirements (Osborne et al 1993):

1) The instrument must work over the whole or part of, the wavelength range 780-2500 nm.
2) It must be able to resolve a small wavelength interval compared with an absorption feature. A typical interval is about 0.1-2 percent of the measuring wavelength.
3) Diffusely scattered radiation is desirable for solid samples, while transmission is ideal for liquids.
4) If small or weak concentrations or absorption changes are under investigation, a wide dynamic range of 10^4-10^5 will be needed.
5) Measurements must be made in a reasonably rapid time. In order to achieve this, the instrument must be able to sufficiently illuminate the sample.
6) The instrument must be compact, robust, and stable, and be useable for process control both in situ and in the laboratory.

In addition, the European Pharmacopoeia states that an NIR instrument must have a means of collecting and measuring the intensity of the transmitted or reflected radiation, such as an integration sphere or a fibre optic probe coupled to an appropriate detector, and a means of mathematical treatment of the spectral data obtained (European Pharmacopoeia 1997).

1.3.2 Sources

NIR instruments fall into one of three groups. These are dispersive, interferometric, and non-thermal. Dispersive and interferometric instruments typically use broad band, thermal radiation produced by an incandescent filament. Often, quartz halogen lamps are used.

Non-thermal or 'cold' sources emit radiation from a much narrower range of wavelengths down to individual emission lines. The main advantage of the non-thermal source is its efficiency, as most of the energy used appears as emitted radiation over a narrow range of wavelengths. As power consumption is much reduced, compact, even battery-powered portable instruments are possible. Non-thermal sources include laser diodes, light-emitting diodes, and lasers (Osborne et al 1993).

1.3.3 Detectors

The most widely utilised detectors in the NIR regions are compound lead-salt semiconductors. These include lead sulphide (PbS) and lead selenide (PbSe). Lead sulphide is used over the wavelength range 1000-2500nm and lead selenide can be

used from 2500-3500nm, and is much less sensitive than lead sulphide. These detectors can be used at room temperature or cooled. Cooling increases sensitivity of the cell to longer wavelengths and increases the signal to noise ratio (Osborne et al 1993).

A relatively new detector is the indium gallium arsenide, or InGAs detector. This operates over 1-1.8μm, is generally smaller and is a few times more sensitive than lead sulphide, size for size.

1.3.4 Wavelength selection

Most laboratory NIR instruments depend on the dispersion-type monochromator for wavelength selection. The use of a grating monochromator provides for a high resolving power instrument. However, the efficiency of the radiation is not always high, as scanning is achieved by rotating the grating and therefore traversing an image past the exit slit. The measurement time is comparatively long for a monochromator system, and they have moving parts. To overcome this, multi-channel detectors have been used recently with dispersive systems (Chalmers1999). The old mechanically grooved and replicated gratings are now replaced by holograpic gratings, which are produced by a photoetching technique. As well as being easier to manufacture, the newer gratings cost less than their predecessors (Morisseau and Rhodes, 1995).

Interference filter-based instruments are advantageous in that they are light, compact and robust. In some cases where analysers work on-line, they may need as few as four or five dedicated wavelengths. In these cases, a filter system may be

a relatively inexpensive option. However, a disadvantage may be that it is restricted for measurements at pre-determined wavelengths. For this reason, filter instruments are used in relatively simple on-line applications (Chalmers 1999).

Acousto-optically tuneable filters (AOTFs) are an alternative technology and are becoming increasingly popular. The required wavelengths are produced by the interaction of ultrasonic waves and white light in a birefringent crystal such as tellurium dioxide crystal (Chalmers 1999). They are highly efficient and small and have the ability to change quickly between wavelengths or to scan the spectrum with no moving parts and all under digital control. They also have good luminous efficiency and are compact and robust (Osborne et al 1993).

In recent years, Fourier-transform NIR (FT-NIR) spectroscopy has also been finding increasing favour. This method is based on the interference of radiation between two beams to yield an interferogram. An interferogram is a signal produced as a function of the change of pathlength between the two beams. The two domains of distance and frequency are interconvertible by the mathematical method of Fourier Transformation. In a FT-IR spectrometer, the radiation emerging from the source is passed to the sample through an interferometer before reaching a detector. Upon amplification of the signal, in which high-frequency contributions have been eliminated by a filter, the data are converted to a digital form and transferred to the computer for Fourier transform to be carried out. The most common type of interferometer is the Michelson interferometer, although quartz wedge types are also used (Stuart et al 1996, Chalmers 1999). The latter type is more compact and has stronger vibration resistance, whereas the former has

higher spectral resolution (Chalmers 1999). The main advantage of FT-NIR spectroscopy is its speed- it is possible to obtain spectra at a millisecond time-scale (Stuart et al 1996).

1.3.5 Operational procedures

The methods employed in NIRS are extremely similar to those used for liquids in the UV and visible regions, and much less labour-intensive than those for the mid-infrared region (Blanco et al 1998). Both quartz and glass are transparent in the NIR, and hence the optical components of NIR instrumentation do not have to be made of fragile materials. This lack of absorption by glass and quartz also enables these materials to be used as transparent containers (Morisseau and Rhodes 1995). For a liquid or solution, the absorbance can be easily measured by using quartz cuvettes or fibre optic probes. As the absorption of NIR radiation follows the Beer Lambert principle, no special precautions need to be taken. Solvents containing N-H and C-H groups are the most ideal ones, as they show little or no absorption in the NIR region (Blanco et al 1998).

For solids such as powders and grains, NIR spectra can be measured by using cuvettes with a transparent window material such as quartz, which are rugged and inexpensive (Stark and Luchter, 1986). However, as glass vials are cheaper and disposable, they are often used instead. The other advantage of using glass vials is that they eliminate the possibility of sample contamination (Simard and Buijs 1996). Another way of obtaining spectra of powders and grains is by using a fibre

optic probe. Although this makes recording spectra significantly easier, light losses resulting from transport along the fibre can cause increased signal noise.

1.4 Data pre-treatment

In NIRS, after acquiring a spectrum, quantitative or qualitative analysis can be carried out using specific algorithms. This is a simple task if an accurate classification method is already established, but to develop such a model takes effort, time and experience. It is necessary to define the classification goal of the application, and to select an appropriate classification method. When using NIRS, it is important to decide whether a data pre-treatment method is necessary. Thus, the selection of a suitable pre-treatment method is another important step in method development (Candolfi et al 1999). NIR spectra are subject to large baseline shifts due to variations such as particle size, shape and compaction. Mathematical treatments are usually carried out on sample spectra to minimise the contribution of physical effects in the NIR. In addition, data pre-treatment can help to provide clear separation between peaks that overlapped substantially in the original spectra.

Adams (1995) states that there are three main aims of data pre-treatment in spectral analysis:

1) to reduce the amount of data and eliminate data that are not relevant to the study being undertaken,
2) to preserve or enhance sufficient information within the data in order to achieve the desired goal,

3) to extract the information in, or transform the data to, a form suitable for further analysis

The most commonly used data pre-treatment methods are the calculating of derivatives, standard normal variate (SNV), and de-trending (DT) (Figure 1.5).

One way to sharpen the resolution of broad or overlapping bands is to take derivative spectra. Derivative spectra are usually generated by differentiating the recorded signal with respect to wavelength. Whereas early applications relied on hard-wired units for electronic differentiation, modern derivative spectroscopy is normally carried out digitally using mathematical functions. The simplest way to produce the first derivative spectrum is by difference (Adams 1995):

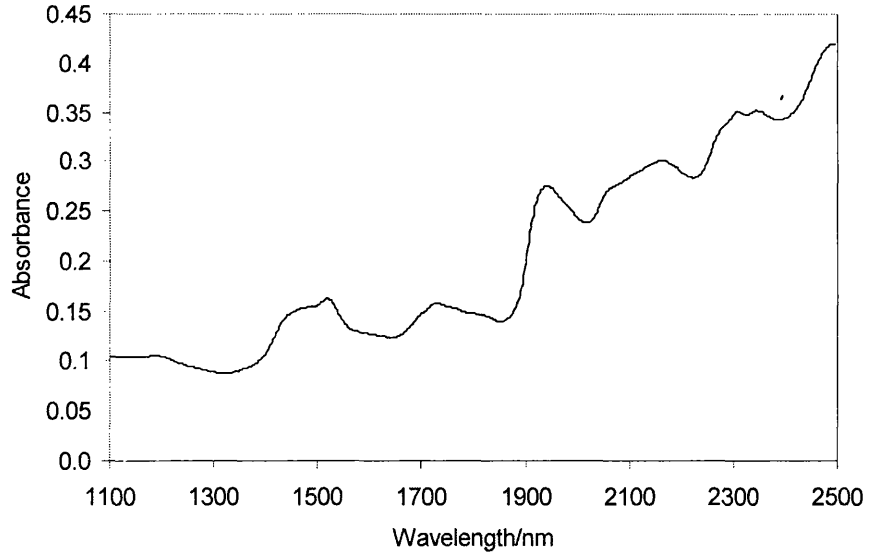
$$\frac{dy}{d\lambda} = \frac{y_{i+1} - y_i}{\Delta\lambda} \tag{1.14}$$

or,

$$\frac{dy}{d\lambda} = \frac{y_{i+1} - y_{i-1}}{2\Delta\lambda} \tag{1.15}$$

where y represents the original signal at wavelength (λ) i.

A



B

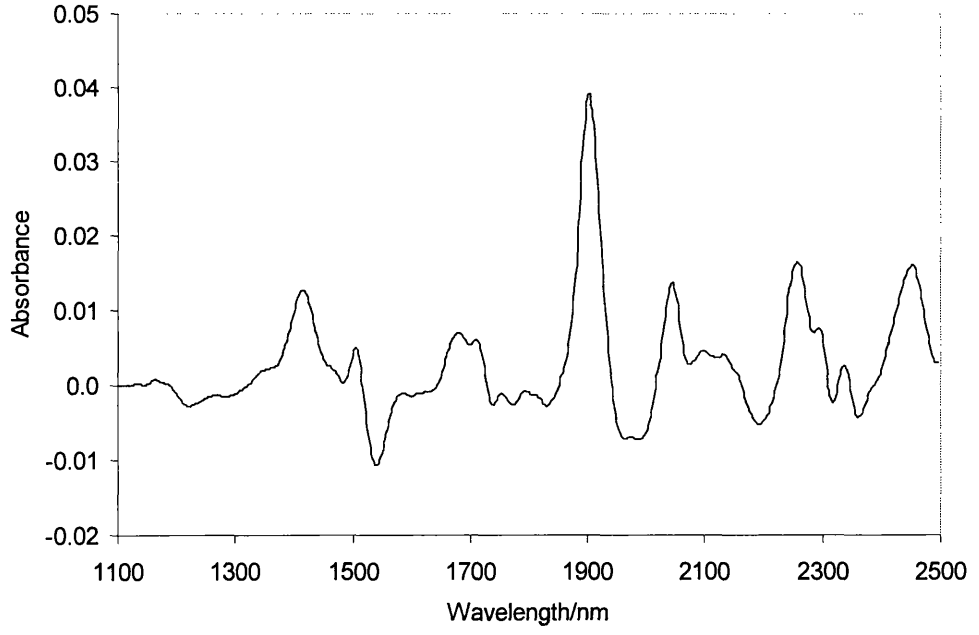
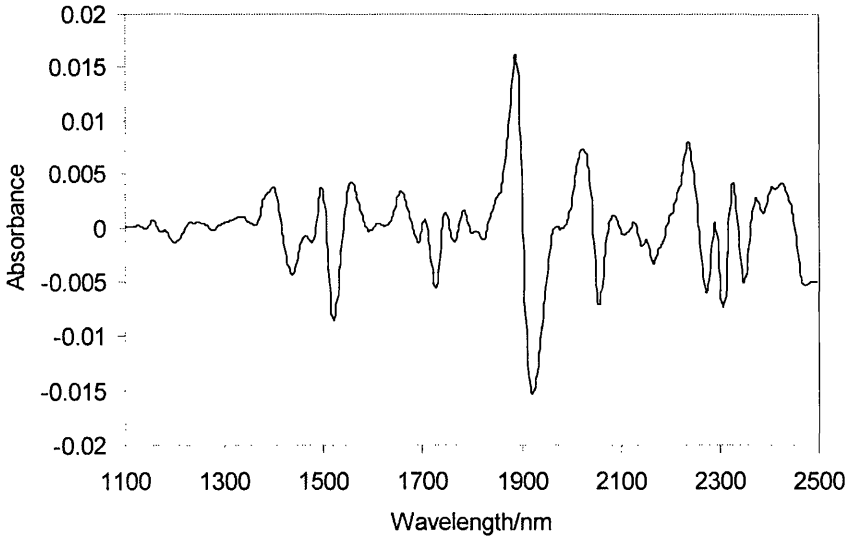


Figure 1.5 A spectrum of belladonna leaf before and after various data pre-treatments. A = raw spectrum (no pre-treatment), B =1st derivative

C



D

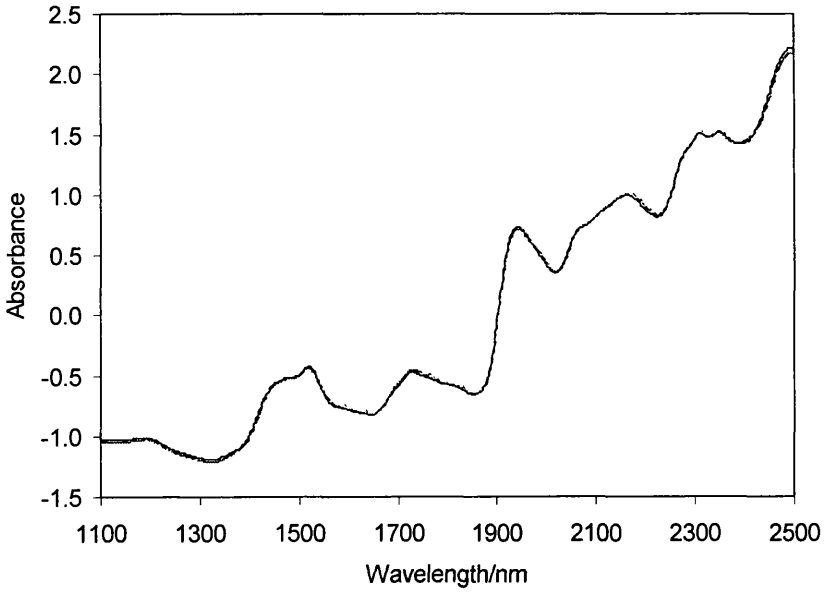
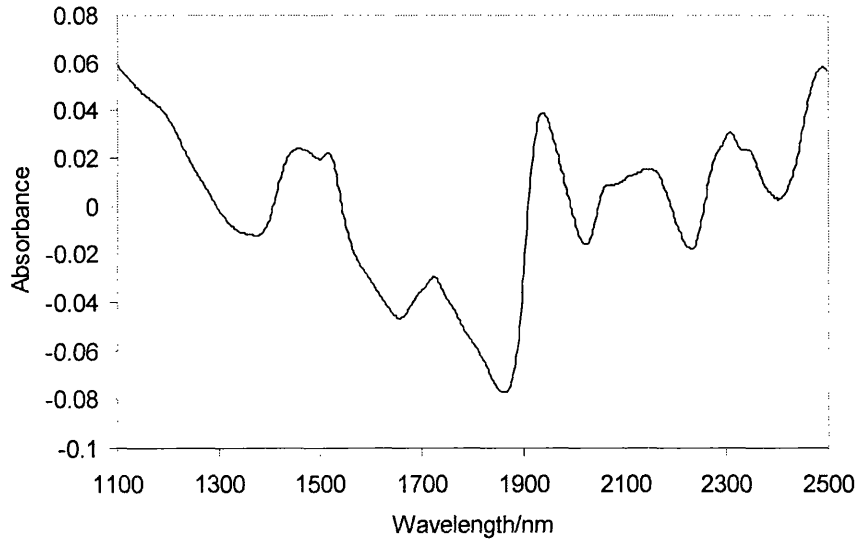


Figure 1.5 A spectrum of belladonna leaf before and after various data pre-treatments. C = 2^{nd} derivative, D = SNV

E



F

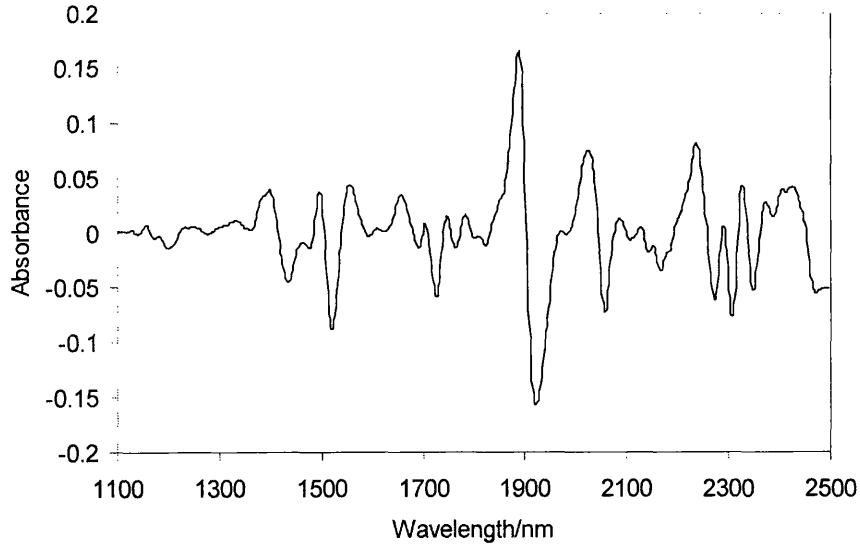


Figure 1.5 A spectrum of belladonna leaf before and after various data pre-treatments. E = de-trend, F = SNV, 2nd derivative

A first derivative spectrum has a peak where the upward slope in the original spectrum reaches a maximum, a trough where the downward slope reaches a minimum, and zero at any peak maximum or trough minimum in the original.

A second derivative spectrum is obtained by differentiating a second time with respect to wavelength, and therefore has peaks and valleys corresponding to each peak in the original. In particular, it has valleys corresponding to each peak in the zero order spectrum. As a result, it provides clear separation between peaks that overlap substantially in the original. A second derivative spectrum can be calculated by the following (Adams 1995):

$$\frac{d^2y}{d\lambda^2} = \frac{y_{i+1} - 2y_0 - y_{i-1}}{\Delta\lambda^2} \qquad (1.16)$$

Second derivatives eliminate baseline curvature and spectral offset which are the result of multiplicative scatter. Third and fourth derivative spectra may also be obtained by differentiating a third and fourth time. Derivatives can also be calculated using a Savitzky-Golay smoothing filter. Here, convolution filters based on polynomial functions are used, in which coefficients are derived from a least squares fit. This method is useful in preserving peak heights more efficiently than the original method (Bromba and Ziegler 1981, 1983, Savitzky and Golay 1964).

One common correction performed on NIR spectra prior to other pretreatments is the use of SNV. This treatment effectively removes the multiplicative interferences of scatter and particle size. (Barnes et al 1989).

This can be performed using the following equation:

$$y_i = \frac{(x_i - x_{mean})}{s} \tag{1.17}$$

In this equation, x_i represents the ordinate values at wavelength i, s is the standard deviation of the ordinate values over the wavelength range being standardised, and x_{mean} is the mean ordinate value of the wavelength range being standardised.

DT is a technique which accounts for the variation in baseline shift, which is normally the case in the reflectance spectra of powdered or densely packed samples. This may use a second degree polynomial regression (Barnes et al 1989):

$$y_{\lambda_j} = a + bj_{\lambda_j}^2 + cj_{\lambda_j} + e_\lambda \tag{1.18}$$

Here, y is the spectral response in the original spectrum at wavelength λ_j, j=1…j wavelengths, b and c are the coefficients of the quadratic least squares equation, e is the residual signal at wavelength λ_j, and a is the spectral offset. The first three terms on the right hand side of equation 1.19 give an estimate of the spectral baseline:

$$\hat{y}_{\lambda_j} = a + bj_{\lambda_j}^2 + cj_{\lambda_j} \tag{1.19}$$

The difference between the original spectrum and the estimated baseline is calculated to produce the residual or de-trend spectrum. Both curvature and spectral offset are removed from the original spectrum using this correction.

The influence in data pre-treatment depends on the data and the pattern recognition method employed. Transforming NIR spectra mostly decreases within-class variance, so that possible errors may be eliminated. However, in the case of using

derivatives this transformation can sometimes emphasise small spectral differences within a class. Thus, this method should be applied with care and if feasible, in conjunction with another method such as SNV (Candolfi et al 1999). Both SNV and DT spectra achieve correlation statistics superior to those obtained from untreated spectra, and are equal to or better than those for the first or second-derivative transformations. Derivatisation of SNV and DT produces statistics which are better than those of untreated spectra.

1.5 Chemometric Techniques used in NIR qualitative analysis

There is a wide range of chemometric techniques available for the qualitative analysis of NIR data. These include Correlation in Wavelength Space, correlation coefficients, Maximum Distance in Wavelength Space, Residual Variance in Principal Components Space, Mahalanobis Distance in Principal Components Space, two wavelength plot, Polar Qualification System, and k-Nearest Neighbour analysis.

1.5.1 Correlation in Wavelength Space

This method calculates the product moment correlation coefficient for the sample spectrum and the average spectrum for each product included in the library (Blanco 1998):

$$\rho_{jk} = \frac{\sum_{i=1}^{p}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{p}(x_{ij} - \bar{x}_j)^2}\sqrt{\sum_{i=1}^{p}(x_{ik} - \bar{x}_k)^2}} \qquad (1.20)$$

where p represents the number of wavelengths, k and j denote the sample and reference product respectively, and x_i is the measured value at wavelength i. $\bar{x}_j$ is the average spectrum for the reference product j and $\bar{x}_k$ equals the average spectrum for the sample. The sample is qualified as the product when the resulting value is higher than a pre-set threshold. A frequently used threshold is 0.85 (FOSS 1998). In theory, if two spectra are taken from the same product (i.e. a perfect match), the correlation coefficient should be exactly 1. However, random noise associated with any type of spectral measurement usually precludes this (Blanco et al 1998).

The calculation of Correlation in Wavelength Space is relatively easy and can be useful for discriminating between closely related compounds. For instance, it has been used successfully for the identification of different types of cellulose (Van der Vlies et al 1993) and for the identification of a pharmaceutical preparation by using a library of 163 substances including active compounds, excipients, amino acids and vitamins (Blanco et al 1994). It has also been successful in non-invasively assessing the contamination of injections by yeast, mould and bacteria (Galante et al 1990).

This method, particularly when using second derivative spectra, is less sensitive to shifts or the slight differences and changes in spectral values that reflect differences in the physical properties and the presence of impurities. On this ground, it can often be less suitable for qualification purposes (Mark 1992).

1.5.2 Correlation Coefficients

This method correlates the absorbance value at each wavelength. The correlation coefficient, r, is denoted by the following (Graham 1999):

$$r = \frac{S_{xy}}{S_x S_y} \tag{1.21}$$

where x represents values for one species, and y represents those for another species. S_{xy} is the covariance, given by equation 1.22:

$$S_{xy} = \frac{\sum xy}{N} - x_{mean} y_{mean} \tag{1.22}$$

where N equals number of variables and x_{mean} and y_{mean} are mean values for x and y.

S_x is the standard deviation of x, given in equation 1.23:

$$S_x = \sqrt{\frac{\sum x^2}{N} - \bar{x}^2} \tag{1.23}$$

and S_y is the standard deviation of y, given in equation 1.24:

$$S_y = \sqrt{\frac{\sum y^2}{N} - \bar{y}^2} \tag{1.24}$$

This method can be used to identify unknowns by calculating the correlation coefficient between unknowns and reference products, and classifying the

unknowns as the species which gave rise to the highest r values. This method produces results that are highly comparable to the Correlation in Wavelength Space method, and thus, it is often sufficient to use either one or the other.

1.5.3 Maximum Distance in Wavelength Space

This method assumes that at each wavelength, measurements are distributed according to the normal law. Data from a library of spectra that define the accepted variability for a product is used to calculate a mean product spectrum and a standard deviation spectrum. During analysis, the distance between the unknown sample and the average spectrum for the reference product is calculated at each wavelength. The most 'unfavourable' situation, i.e., the maximum distance, is determined using the following equation:

$$d_{kj} = \max \frac{\left| x_{kp} - \bar{x}_{jp} \right|}{s_{jp}} \tag{1.25}$$

Here, subscripts k and j denote sample and reference product, respectively; x_{kp} is the measured sample value at wavelength p; $\bar{x}_{jp}$ is the average spectrum of reference product j at wavelength p; and s_{jp} is the standard deviation of the measured values for reference product j at wavelength p (Blanco et al 2000).

If the sample under investigation belongs to the same population (e.g. species) as the reference product, then there will be a probability of 99.7% that the distance will be less than three times the standard deviation. If the maximum distance is larger than this, then the sample must belong to a different population, and

therefore will fail the identification procedure. Although the qualification criterion $D_{max} \leq 3$ is often thought to be a typical match value, it is usually more practical for users to identify and decide on the most suitable limit based on their own problems and methods. In general, however, this method has been found to be more successful than the use of correlations (Blanco et al 1998).

Correct usage of this method requires extensive control of the instrument in order to ensure that noise remains fairly constant, as this method can be sensitive to noise, water peaks and wavelength shifts (Gonzalez and Pous 1995).

One problem which is often encountered when using this method, is the risk of obtaining so-called false negatives at those wavelengths coinciding with x-intercepts in derivative spectra. If there is a very small standard deviation for the average spectrum at a given wavelength, then the resulting distance will be very large and an identification failure will result. This can be the case when the values observed in second derivative spectra are very close to zero. One way to circumvent this problem is to use a wavelength library stabilisation method (Blanco et al 2000).

This method has been successful in various applications. These include the discrimination of various strengths of tablets still in their blister packaging (Dempster et al 1993) and the identification of ten pharmaceutical excipients (Candolfi et al 1999). Another study involved the identification of cellulose, methyl cellulose, cellulose acetate phthalate and corn starch (Gemperline and Boyer 1995) in training sets of 8-10 spectra per library, using a probability threshold of 95% instead of a distance threshold.

1.5.4 Residual Variance in Principal Components Space

The American Society for Testing and Materials defined Principal Components Analysis (PCA) as:

"A mathematical procedure for resolving sets of data into orthogonal components whose linear combinations approximate the original data to any desired level of accuracy. As successive components are calculated, each component accounts for the maximum possible amount of residual variance in the set of data. In spectroscopy, the data are usually spectra, and the number of components is smaller than or equal to the number of variables or the number of spectra, whichever is less." (ASTM)

PCA effectively reduces the data to a small number of variables, or principal components. These are linear combinations of the original variables. More than 90% of the variation is modelled in the first principal component. In NIRS, this usually represents information relating to either water content or particle size. These factors can often overwhelm the spectrum and cause co-linearity, where the same information is revealed in measurements at different wavelengths. PCA helps to reveal the inherent structure of the data that correlates to the analyte in question.

Residual Variance in Principal Components Space is a method that calculates the local PC model for each product in the library. Each product PC model is used to reconstruct the unknown spectrum. The difference between the original and reconstructed spectrum is then used to calculate the residual variance. The residual variance for a spectrum k to be identified (S^2_k) that is assumed to belong to class j

(defined by the spectra for reference product *j*) is divided into the total variance for the samples belonging to class *j* (S^2_0) in order to obtain the following variance relation (Equation 1.26) (Blanco et al 2000):

$$F = \frac{S_k^2}{S_0^2} \frac{n}{n-a-1} \tag{1.26}$$

where n is the number of spectra for the reference product and a the number of Principal Components used to construct the class model (Blanco et al 1998). From the resulting F function, one can calculate the probability that a given sample belongs to the distribution represented by the training set (Blanco et al 2000). The unknown is identified as a product when the residual variance for this product's Principal Component model is within the threshold value. The default threshold in FOSS Vision© software is 0.95 for match value and 0.84 for probability level (FOSS 1998).

1.5.5 Mahalanobis Distance in Principal Components Space

In this method, the local PC model is calculated for each product in the library. During the analysis, the unknown spectrum PC scores are calculated for each product in the library. During the analysis, the unknown spectrum PC scores are calculated for each product model and Mahalanobis distance is calculated (FOSS 1998). The Mahalanobis distance is a useful tool for cluster analysis, and can be calculated for multi-dimensional spaces. The distance (D) between the sample and the centre of the cluster formed by the spectra of the reference product is defined in equation 1.27:

$$D^2 = (x_j - \bar{x}_k)^T C (x_j - \bar{x}_k) \tag{1.27}$$

x_j is the vector describing the spectrum of sample j, x_k is the vector for the average spectrum of reference k, C is the matrix that describes the distance measurements in the multidimensional space studied and superscript T denotes a transpose matrix (Blanco 1998). The unknown is identified as a product when the Mahalanobis distance for this product is within the threshold value. Usually, the threshold is set at three times the Mahalanobis distance (Mark 1992), but as in Maximum Distance in Wavelength, thresholds can vary depending on the users own problems and working methods. In FOSS Vision® software, the default threshold is 4 for match value and 0.84 for probability level (FOSS 1998).

1.5.6 Two-wavelength analysis

In optical spectroscopy, the most common method of classifying an unknown is to visually examine the spectrum under investigation. Samples can often be classified or identified by matching the location and strength of absorbance peaks to those of known substances. It is possible to generalise this procedure by noting that, if the absorbance is measured with sufficient accuracy, then any wavelength where there are absorbance differences between the substances to be distinguished can serve to classify them (Mark 1992).

This concept can be used to generate a visual representation of differences observed. For instance, if the spectra of two or more materials are seen to vary at

two wavelengths, it should be possible to dispense with all other spectral data except the absorbances at these particular wavelengths, and still be able to classify the samples.

Thus, data corresponding to absorbances at two wavelengths can be plotted against each other to provide groups of points. If the spectra of the materials studied are sufficiently different at two wavelengths, these wavelengths alone can be used to characterise the product. Analysis of unknowns can thus be performed by plotting their data corresponding to the absorbances of the same wavelengths and carrying out a comparison. Various data pre-treatments and wavelengths can be tried in order to find the most successful combination(s). The most ideal pre-treatment and wavelength combinations will vary according to the type materials under investigation.

In general, when only a few materials are to be distinguished, the differences between the spectra can be detected visually and the wavelengths to use for discrimination selected manually. However, if some of the spectra are extremely similar, or there are so many different materials involved that visual inspection becomes impossible, then manual selection would not be possible. If there is no prior knowledge as to which are the suitable wavelengths, then a method for detecting the optimum set of wavelengths is necessary. One common way in which this is done is to compute the distances D_{ij} between all pairs of groups i and j. Using these results, the sum of the inverse squared distance, i.e., $(1/D_{ij})^2$ is formed. The groups that are most similar (closest together) will contribute most heavily to this sum. As a result, selecting those wavelengths that cause this sum to

be the smallest results in the selection of the wavelengths that best separate the closest groups. This approach is advantageous in that it will optimise among all groups that are comparably closely spaced, rather than concentrating on only the single closest pair (Mark 1992) However, in many cases, as will be revealed in later chapters, simple visual inspection and trial and error can serve as sufficient means in selecting suitable wavelengths.

1.5.7 Polar Qualification System

The consistency of starting materials for the quality of end products in the pharmaceutical industry is of utmost importance. Traditional testing of these raw materials usually look for compliance with the pharmacopoeial criteria for identity, purity, and strength (Van der Vlies et al, 1995). These are mainly chemical parameters, and unfortunately, monographs seldom include tests to measure physical parameters, which may be just as important for the production of drugs of consistent quality (Van der Vlies et al, 1995). Deficiencies in raw material testing can lead to batch-to batch variation in the end products (Brittain et al 1991).

Common methods employed for the qualitative analysis of materials such as Maximum Distance in Wavelength and the use of Mahalanobis Distance often assume the availability of a reference and learning set necessary to determine the criteria for compliance with the confidence limits of these methods (Van der Vlies et al, 1995). Frequently, reference and learning sets may be unavailable, and in

addition, certain methods may need the use of sophisticated chemometric techniques which are not easily nor quickly mastered (Van der Vlies et al, 1995).

In comparison, the Polar Qualification System (PQS) has its advantages in that by using relatively simple mathematics, the information present in an NIR spectrum can be reduced to one single "quality point" in a two-dimensional plane (Kaffka and Gyarmati, 1990). This is achieved by plotting the second derivatives of the NIR reflectance spectra of materials in a polar coordinate system, and calculating the "centre of gravity" of the plot obtained (Kaffka and Gyarmati, 1990). This method can be used to test the acceptability of certain products, as the position of the quality point is sensitive to changes in the composition of the product being analysed. Thus, a deviation in the quality point can serve to indicate that the sample does not meet the required quality level (Van der Vlies et al 1995). Van der Vlies reported that PQS could effectively differentiate between two different batches of amoxycillin trihydrate originating from two different sources, which only differed in terms of their particle size.

The calculation of polar coordinates is reasonably easy. Firstly, the selected part of the spectrum is transformed into polar coordinates X_i and Y_i by using Equations 1.28 and 1.29, where n+1 is the number of datapoints (wavelengths) and $A_0 \ldots A_n$ the second derivative absorbances at datapoints i=0…n (van der Vlies et al, 1995).

$$X_i = |A_i| \cos(2\pi i/n) \tag{1.28}$$

$$Y_i = |A_i| \sin(2\pi i/n) \tag{1.29}$$

It should be noted that in the above equations, the absolute values of the absorbances were used. This is because in the second derivative, a negative peak has, in most cases, two adjacent positive peaks. As a result, by taking the absolute values, three adjacent positive peaks are obtained. This is important, as in the next step, the centre of gravity of the polar plot is calculated. If the absolute values are not used, the peaks would be situated in opposite quadrants, reducing the sensitivity of the method (Van der Vlies et al, 1995).

The next step is to calculate the coordinates of the centre of gravity ("quality point") of each plot using Equations 1.30 and 1.31 (Van der Vlies et al, 1995):

$$Z_x = \frac{1}{n+1} \sum_{i=0}^{n} X_i \tag{1.30}$$

$$Z_y = \frac{1}{n+1} \sum_{i=0}^{n} Y_i \tag{1.31}$$

This method can be carried out along the whole spectrum from 1100-2500n, or on sections of the spectrum, for example at 100nm increments. With its simplicity and effective discriminating power, PQS has great potential for qualification purposes.

Figure 1.6 shows a section of the second derivative spectra of five plant resins from 1700-1800nm. Figure 1.7 shows the same spectra after they have been transformed into polar coordinates X_i and Y_i. All spectra show a "flower"-shaped pattern which is characteristic of PQS analysis.

Figure 1.8 shows the final plot of the "quality points" after calculation using Equations 1.30 and 1.31. It is clear from this resulting plot that the PQS method allows a good separation and clustering of all the materials according to their identity. Even though each sample within each group was different, all clustered at specific regions depending on what they were.

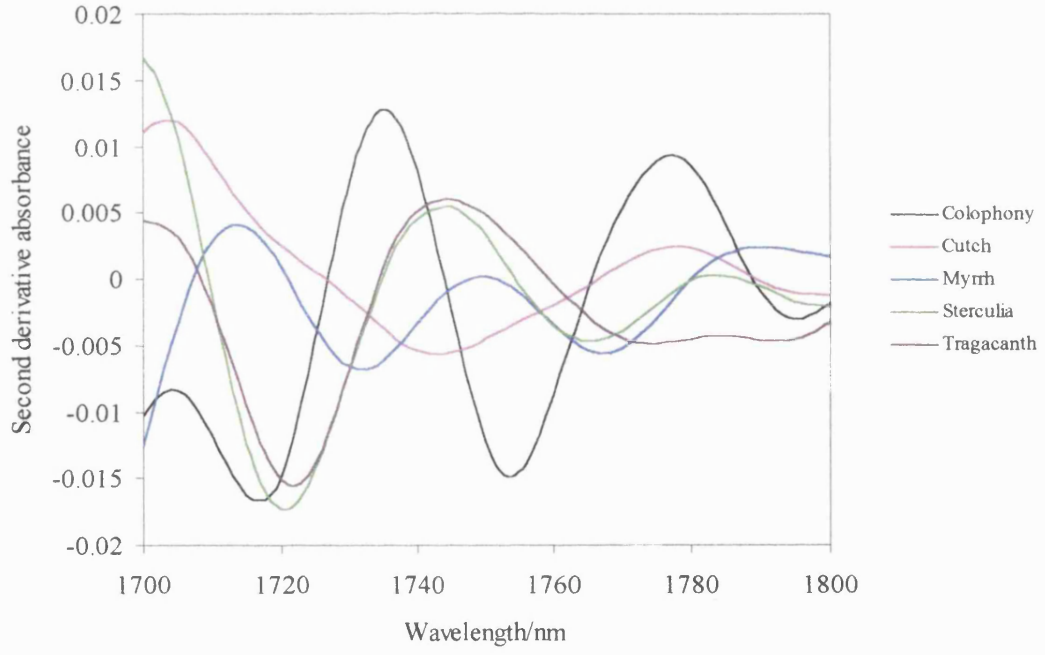


Figure 1.6 Representative second derivative spectra of five resin samples in the wavelength range 1700-1800nm

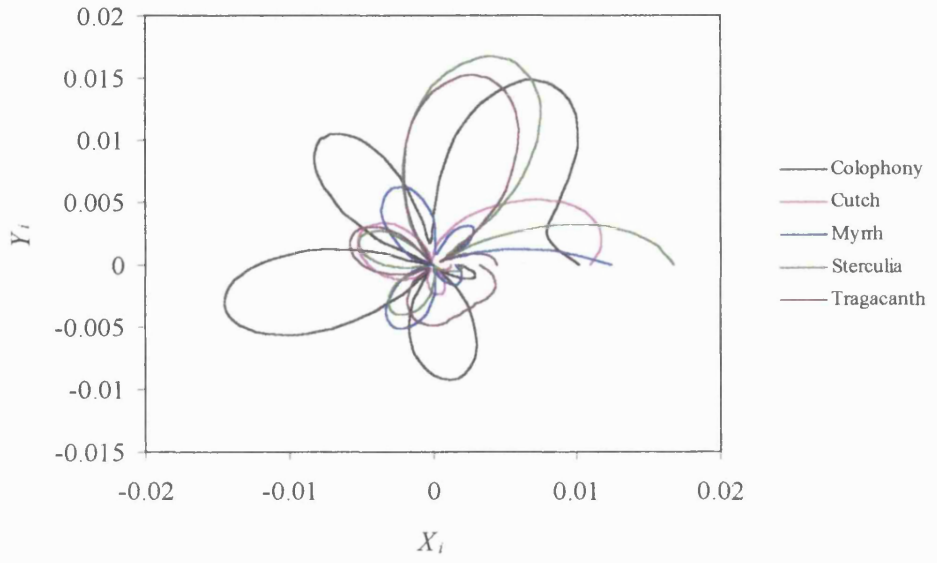


Figure 1.7. Representative second derivative spectra of five resin samples plotted as polar coordinates X_i and Y_i using the wavelength range 1700-1800nm

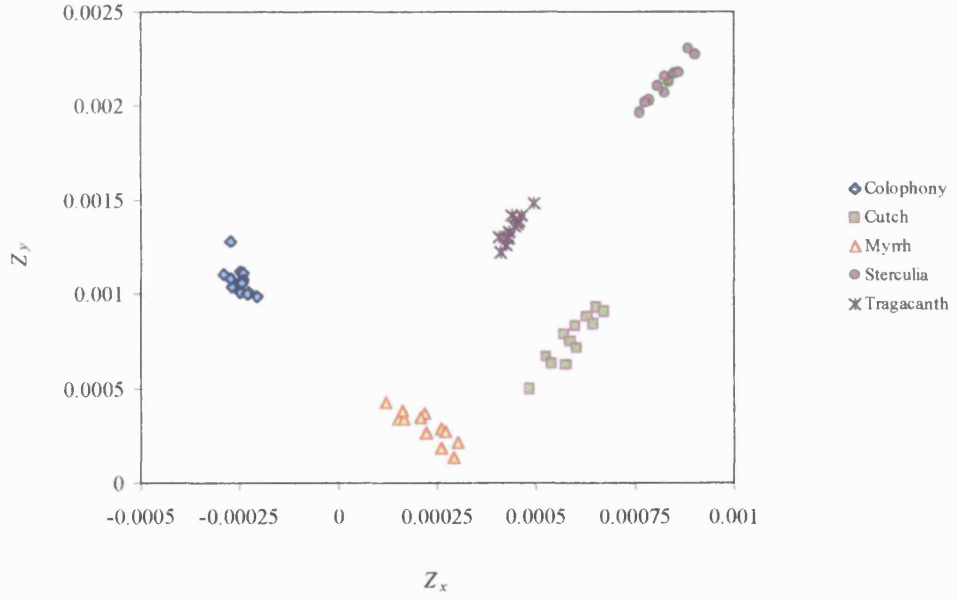


Figure 1.8. Centres of gravity ("quality points") of five types of resins in the wavelength range 1700-1800nm

1.5.8 k-Nearest Neighbour

Both the two-wavelength and PQS method can be used to identify unknown substances. One way in which this can be done is by the *k*-Nearest Neighbour (*k*-NN) method. This is a very simple mathematical classification procedure where

the distance between the datapoint of an unknown and those of a training set are calculated. Usually, the Euclidean distance is used. If the training set is made up of a total of *n* samples, then *n* distances are calculated. The *k* nearest samples to the unknown are then selected which classifies the unknown to the group that produced the group to which the majority of the *k* samples belong. The choice of k can be determined by optimisation, although it is usually found that small values of k (3 or 5) are to be preferred. This method works best where there is little overlap between the groups (Massart et al 1988).

The Euclidean distance, *D*, between the datapoint of the 'unknown' and every remaining point, which can be used for *k*-NN, is calculated by the following (*U* = 'Unknown', *K* = 'Known', *x* and *y* are x and y coordinates, respectively):

$$D = \sqrt{(U_x - K_x)^2 + (U_y - K_y)^2} \tag{1.32}$$

If the training set is of a sufficient size, the mathematical simplicity of this method does not prevent it from producing classification results as good as or better than other more complex methods. It also has the advantage of being a multi-category method (Massart et al 1988). Figure 1.9 shows a PQS plot of plant resins with datapoints for samples treated as unknowns. From this, it can be seen that the datapoints of the "unknown" samples lie closest to the group they theoretically belong to.

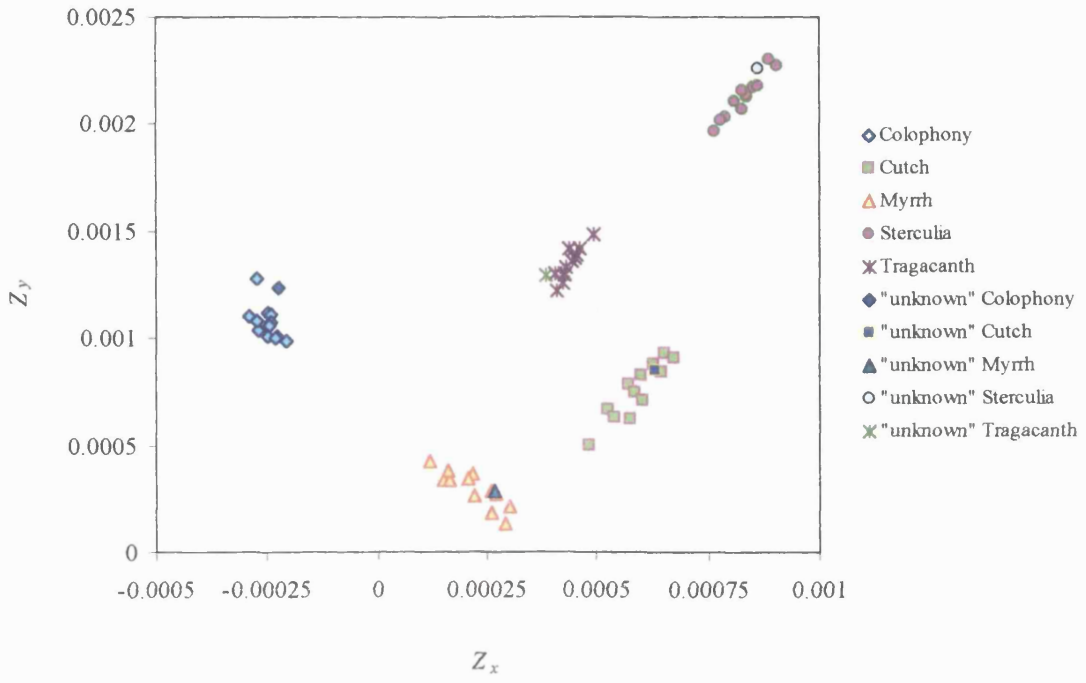


Figure 1.9. PQS plot of plant resins using the wavelength range 1700-1800nm incorporating "unknown" samples.

1.6 Applications of NIRS in food and beverage analysis

Although NIRS is used extensively as an important analytical tool in the pharmaceutical industry, it first found its use in the food and agricultural industries for the determination of moisture, protein, and starch in grains (Morisseau and Rhodes 1995). Its success in these industries, which often involve plant-based substances (e.g. food grains), suggests that NIRS could also have potential in the herbal medicine industry, where similar methods could be applied.

1.6.1 Moisture

The NIR absorption spectrum of pure water includes 5 bands with maxima at 1940, 1450, 1190, 970 and 760nm at room temperature. These are subject to shifts, as results of variations in temperature and hydrogen bonding when water is in a solvent or food matrix. The determination of moisture in liquids (such as fuming

nitric acid, glycerol and methanol) was one of the earliest analytical applications of NIRS (Osborne 1993a). It was realised that the same principle could be applied to foodstuffs, and now the determination of water is still probably the most widely used application of NIRS in food analysis. It is based on measurements of well-defined absorption bands, but background correction is complex and empirical as a result of the limitations of diffuse reflectance when the background is absorbing, and due to varying absorbance from sample to sample (Osborne et al 1993). Moisture analysis by NIRS is commonly used to measure the moisture content and other characteristics of wheat and other commercially important small grains (Delwiche et al 1992). In fact, it can be said that it is not often in the history of science that an instrumental technique and a particular industry become so closely interlinked as is the case for NIRS and the grain trade (Downey 1994). High accuracy, minimal sample preparation and rapidity of analysis are the main reasons, giving it the potential to be used for on-line analysis and control. With the development of whole grain instrumentation and single-kernel sample holders, non-destructive analysis of many grains on an individual kernel basis is feasible (Downey 1994). Water content is directly related to the occurrence of biological reactions which is the cause of degradation of carbohydrates. Therefore, investigating water activity and moisture content have great implications in the storability of grains (Delwiche et al 1992). In addition, moisture content can affect the suitability of certain raw materials, for example soybeans, for their use in food processing (Takahashi et al 1996). As well as grains, NIRS is used to measure

moisture content in other foodstuffs, one example being cheese (Wehling and Pierce 1988).

1.6.2 Other applications

NIRS is also used in the analysis of fats, such as shortenings in foods, carbohydrates (e.g. the cellulose content of flour) and sugars, such as the determination of glucose, fructose, and sucrose in fruit juices, wine, chocolate and baked products (Osborne et al 1993a). The application of NIRS in food analysis is important for various commercial reasons. For instance, it is vital in determining the baking quality of wheat. In yeast leavened bread, for instance, flour of a minimum of 11 percent protein is preferred. Therefore, NIRS is regularly used to verify that the wheat used has a protein content of approximately 12 percent. NIRS can also be used to check for starch damage and assess the dough visco-elastic behaviour and to indicate when the milling process has been out of specification (Downey 1994). It can also be used to detect the presence of inorganic additives in flour (Sirieix and Downey 1993). In addition, NIRS can be used for such novel purposes as determining the geographical origins of chocolate (Osborne et al 1993a) Other instances where NIRS has been used in the analysis of foodstuffs is briefly described below.

1.6.2.1 Coffee

Commercial coffee is either made up of Arabica or Robusta varieties or blends of the two types. As Arabica is of better quality, it is more expensive, and as a result, there exists the high potential for fraud (Downey et al 1994). In the coffee trade, most of the business is between the producers and processors. The processors use

trained personnel to identify and confirm bean variety using sensory methods that are hard to replicate. When coffee is sold, it is in the form of roasted beans or as instant coffee, and these are hard to assess, even by trained sensory panelists (Downey et al 1994).

Studies were performed by Downey et al (1994) to investigate the potential of NIR reflectance spectroscopy for discriminating between pure Arabica and Robusta coffees and blends of the two. Investigations were carried out on whole and ground beans, and it was found that success rates for identification using principal components analysis ranged from 82.9% to 96.2%, with pure samples yielding more successful results than blended ones. A high degree of success was also achieved when using lyophilised (freeze-dried) coffee (Downey and Boussion 1996).

1.6.2.2 Orange juice

With the increase in interest in healthy eating, fruit juices have become commercially important commodities (Evans et al 1993). In the EU, most orange juice is prepared from imported frozen concentrate which is reconstituted by adding water. In the UK market, most of the frozen orange concentrate originates from Brazil and Israel. Adulteration of fruit juices has long been practised, including the over-dilution of the concentrate with water, further extraction with water of the residual products of juice extraction (pulp), addition of sugars, acids and water to swell the volume while maintaining sweetness, and the addition of metal ions and amino acids (Evans et al 1993). As the methods for detecting adulteration have become more advanced, the methods of adulteration have

progressed as well. Detection of adulteration is made harder due to the fact that authentic fruit juice can vary naturally, thanks to species, maturity, growing location, climate, season, planting system and processing techniques (Evans et al 1993). Testing the potential of NIRS to discriminate between three sources of orange juice (Brazilian, Israeli unadulterated, Israeli adulterated) showed that a 100 percent success rate could be achieved when using 25 principal components on raw spectra (Evans et al 1993). The determination of various components in orange juice which are of importance in quality control procedures (e.g. glucose, fructose, sucrose, citric acid, malic acid) using NIRS and Partial Least Squares has also been investigated, showing that NIRS is a promising technique in this aspect (Li et al 1996).

1.6.2.3 Rice

Long grain rice accounts for over half of the total UK rice market, with a retail value of £200 million per year. Basmati, a class of rice grown in the Punjab region of India and Pakistan, is recognised as a market sector in its own right, with about 20 percent of total rice sales (Osborne et al 1993b, Osborne et al 1994). As Basmati can only be grown once a year with a yield about half that of other rices, it commands a high price, especially due to its popularity in the UK, and thus there is a high incentive for adulteration. The potential of NIRS to discriminate Basmati from cheaper, inferior rice types was investigated as a quick and non-destructive method that would protect both the consumer and the honest trader (Osborne et al 1993b). Results were encouraging, and were further supported by Krzanowski (1995).

1.6.2.4 Other food ingredients

In addition to the examples shown above, the potential of NIRS for the authentication of food and food ingredients has been tested for a wide range of other products. These include the geographic origin of so-called Italian olive oil, the vegetable origin of vegetable oil, and vanilla (Downey 1996). Clearly, as food authenticity is an issue of concern to food processors, retailers, regulatory authorities and consumers, a rapid technique for testing this is desirable, and thus NIRS has many advantages in that it is fast, non-destructive and does not require highly skilled staff or chemical reagents. Ready transfer of NIR models to the factory as either on- or near-line methods is already possible (Downey 1996).

NIRS has also shown its potential in the brewing industry. For instance, it has been used in conjunction with a fibre optic probe to measure the alcohol content in beer, with an accuracy of ± 0.04% alcohol (Benson, 1996).

In addition to food for human consumption, NIRS has also been used for the analysis of animal feed, such as feed barley (Edney et al 1994). This is of importance, as feed manufacturers need to get maximum value out of their ingredients. In order to achieve this, they need the exact information on the nutritive value of their feed ingredients. Traditionally, feed formulators relied on book values for nutrient composition, but this was not always economical, and thus NIRS possesses advantages in that it is quick, accurate and relatively inexpensive (Edney et al 1994).

1.6.2.5 Parasites

NIRS has also been investigated as a method for detecting parasites in crops. For instance, it has been used to detect grain weevil larvae and pupae inside single wheat kernels. Although modern control measures mostly prevent infestations reaching levels where significant loss of commodity occurs, consumers are increasingly demanding food to be of the highest quality, and as a result, cereal buyers now operate a zero-tolerance to insects (Ridgway and Chambers 1999). As many producers apply pesticides as an insurance against insects, this can lead to unnecessary levels of residues in the final product. Thus, the efficient detection of insects would allow better food quality, as well as the application of pesticides to be more efficiently targeted. Although there are established, rapid methods to detect insects external to the grain, no method exists which can detect insects hidden inside the kernel. Using NIRS on weevil-infested and uninfested grains, Ridgway and Chambers showed that it could be used to correctly classify over 96% of samples. Spectral differences appeared to arise from the actions of the developing insect, rather than from any feature specific to kernels selected by adult females for egg-laying (Ridgway and Chambers 1999). It was possible to detect infestation by measurements at just two wavelengths, 1194nm and 1304nm. Detection of larvae inside grains was also possible using NIR imaging, where physical images were recorded at carefully chosen wavelengths in the NIR region (Ridgway and Chambers 1998).

1.6.2.6 Animal proteins

Although NIRS in food analysis is usually confined to agricultural crops, it can also be used to analyse animal protein (Rannou and Downey 1997, McElhinney et al 1999). This is of particular importance as there has been a tendency of later years to the increased consumption of pre-processed foods. As a result, there have been concerns voiced over the authenticity of meat products. Often, there are cases where meat from high-value species is substituted by meat from a lower value species. Often, there are cases where meat products are extended with fat, rind, or non-meat proteins or water (Rannou and Downey 1997). In addition, consumers may desire to avoid certain meat species for religious (e.g. pork) or health (e.g. beef) reasons. Species identification is not a problem when the meat is in whole cuts, but when they have been chopped or minced, it becomes virtually impossible (McElhinney et al 1999).

Traditional methods for analysis of animal proteins include capillary gas chromatography, electrophoresis, immunoassays and DNA-based procedures. However, these methods can be time-consuming, and all require sophisticated facilities and highly trained staff (Rannou and Downey 1997).

Investigations carried out on NIR spectra of raw pork, turkey, chicken, beef and lamb showed that there is a great potential for NIRS to address the issue of raw meat speciation (Rannou and Downey 1997, McElhinney et al 1999). Overall, NIRS appeared to be more accurate than mid-infrared spectroscopy (Rannou and Downey 1997), although success rates appeared to be higher in smaller group

classifications (McElhinney et al 1999) and the molecular phenomenon for this is as yet unclear (Rannou and Downey 1997).

1.7 NIRS of natural products

In the qualitative analysis of materials using NIRS, it is always necessary to build a qualitative library incorporating the training set data for each material grouping, any subsequent transformation and discriminatory analysis. These will depend on the intended use of the particular library. A typical qualitative library development will most likely involve the following stages (PASG NIR Sub Group 1998):

1. Define the purpose of the library
2. Selection of samples/spectra for the training set
3. Data pre-treatment
4. Library construction
5. Determination of thresholds (e.g. for Maximum Distance in Wavelength Space

In defining the purpose of the library, it is important to define its scope in terms of its intended use prior to starting development, that is, identification and/or qualification. Consideration should be given to the chemical similarity and numbers of material groups to be discriminated in both their chemical similarity and number. In constructing the library, a full or reduced wavelength range may be used. Wavelength segments can be useful to remove unwanted effects or to enhance small differences (PASG NIR Sub Group 1998).

In terms of selecting spectra for the training set, sample variability due to factors such as moisture, particle size and other chemical/physical properties may be built

into the library. This is especially important for qualification libraries. It may also be desirable to select representative samples from a large population. Library thresholds can be determined following internal validation of the library, and assessment of external sample performance and a consideration of the next best match (PASG NIR Sub Group 1998).

Thus, in the qualitative analysis of all materials, including natural products and drugs derived from plants by NIRS, the process involves working logistically with libraries of samples to bring about statistical identification as well as discrimination.

Previously, very little work has been carried out on herbal materials, due to the fact that most interest has been on the analysis of materials of agricultural and commercial importance. In addition, plant materials are extremely difficult to characterise by NIRS, as it is often the case that there is no clear-cut reason as to why spectral differences may occur. That is, even in samples of the same species, the presence or efficacy of active components may vary for numerous reasons, such as season, climate, temperature, and geographical origin. Perhaps the closest application of NIRS of natural products, which is currently used in a commercially important setting, is its use in the tobacco industry. Commercial tobacco blends are a complex mixture of different tobaccos and additives that are typically formulated to meet a specific composition. These can vary substantially and therefore the characterisation and quantitative analysis of their properties is difficult. Thus, in the tobacco industry, there is the challenge of assessing the intake quality of the raw materials and the analysis of tobacco blends for process verification. It is in

the interest of the manufacturers that the tobacco blends that have been mixed and their quantities can be estimated as rapidly as possible with minimal sample preparation (Lo, 1996). Dominguez and Seymour (1992) proposed the use of NIRS and discriminant analysis for the differentiation of various tobacco types and for the prediction and verification of blend identity. Another study by Lo (1996) discussed the use of PCA, artificial neural networks (ANN) and *k*-NN for the differentiation of individual types of tobacco blends on 127 samples of pulverised tobacco leaves containing six different leaf types. ANN is a robust and flexible method of modelling data that has primarily been focused on the classification of single categories rather than on the identification of separate components from spectra of mixtures. Results showed that ANN provided good discriminant power and generalisation in the identification of individual components from mixtures, and that the tobacco blending ratio could be estimated within 2.5% (w/w) error. Although useful, the success of the *k*-NN method in this scenario was limited, as it was not possible to use the technique to detect whether a sample was a pure component or a mixture (Lo, 1996). Thus, to some extent, NIRS has been shown to be successful for the characterisation of herbal materials, although the transition from the tobacco to phyto-pharmaceutical industry has been approached with caution.

In one early study, however, NIRS was applied to the identification of plant drugs, namely ginseng and grape seeds (Corti et al 1990). Qualitative analysis was carried out using a type of discriminant analysis to identify the various substances by their absorption spectra. Results showed that the distinction of grape seeds

according to their geographical origin was successful. In the case of ginseng, two important aspects were tested. Firstly, the possibility of identifying adulterated or false specimens was considered, and secondly the distinction of young and old roots was considered. Results showed that when using samples of ginseng with and without added starch, the normalised distances between different samples increased with increased percentage starch. In the case of age having an effect, it was successfully shown that old and young roots were distinguishable. In addition, it was also possible to semi-quantify ginsenosides. Analysis in all cases was possible with minimal sample preparation. Thus, in this study, four important aspects in the herbal industry were approached: the identification of country of origin, the assessment of adulteration, the age of the sample and the quantification of the active component(s).

In a more recent study, herbal medicines such as *Cassia*, *Ganoderma*, and *Smilacis Rhizoma* were shown to be discriminated accurately according to geographical origins by NIRS and by using the PLS regression method (Woo et al 1998).

The concept of discriminating herbal materials according to geographical origin was followed up in a further investigation on samples of *Astragali Radix*, *Ganoderma*, and *Smilacis Rhizoma*. It was shown in this study that second derivative NIR reflectance spectra of representative samples did show slight differences according to the places of origin. It was then shown that samples from the two different places of origin (in this case, China and Korea) could be differentiated using PCA. In addition, the techniques of using Mahalanobis distance and discriminant analysis with PLS were used to successfully

discriminate the various samples (Woo et al 1999). Another study, by Kwon and Cho (1996) on the identification of the geographical origins of ginseng and sesame seeds using PLS showed that NIRS could be used to allocate the country of origin with a 99% accuracy rate for sesame seeds and about 95% for ginseng. The identification of the geographical origin of natural products is of importance not only for quality control measures, but also for the prevention of unlawful trading and illegal smuggling. For instance, in Korea, the Ministry of Finance strictly controls the monopoly of red ginseng trading, as red ginseng produced elsewhere is considered to be cheaper and inferior (Kwon and Cho 1996). In addition, it has been known that although samples of a certain herbal medicine may come from the same species, the quality and content of active can vary according to where they were cultivated and also according to the growing conditions of the different regions (Woo et al 1999). Thus, the development of an ultra-fast, simple and non-destructive inspection technique for identifying the country of origin and for determining the quality of herbal commodities would be of great importance.

NIRS has also been used to determine the essential oil content of certain herbal materials. For instance, in the case of fennel and caraway, the content of the monoterpene substances carvone, trans-anethole, fenchone, and estragole is influenced by breeding (Fehrmann et al 1997). It was found that the levels of these could be determined using NIRS. Ideally, rapid, non-destructive methods are desired for the evaluation of fruits during the selection process, and as traditional methods such as gas chromatography and steam distillation are time-consuming, NIRS has been found to be useful.

NIRS has also been found to be a successful method for the quality control of various commercial essential oils. As essential oils used in complementary medicine are subject to very few quality control procedures, it is difficult to tell whether an oil which is labelled as pure is free from adulterants and contaminants. For instance, rosemary oil is often diluted with the cheaper eucalyptus oil (Wilson et al 2000). A study by Wilson et al (2000) on rosemary essential oil, using Correlation in Wavelength Space and Maximum Distance in Wavelength Space showed that it was possible to discriminate between rosemary and eucalyptus oils. A forward search multiple linear regression method (MLR) was also used which chose the wavelength at which the absorbance values correlated the most with the reference data. Results showed that in the prediction of eucalyptus oil content in rosemary oil, less than 5% could be detected over the 0 to 100% range and less than 0.2% in the 0 to 10% range (Wilson et al 2000). The determination of Cineole (Eucalyptol) content in Eucalyptus oils, which, according to the British Pharmacopoeia, is present in amounts not less than 70% w/w, has also been investigated (Wilson et al 2001a), as was the determination of constituents present at much lower levels. For example, lemon oil generally contains between 2-4% citral. Using the BP assay method as a reference, it was found that the use of NIRS and MLR produced a correlation of 0.962 and a Standard Error of 0.19% m/m citral, thus suggesting that NIRS may be a suitable method for assessing the quantities of small amounts of constituents in essential oils (Wilson et al 2001b), as well as determining their identity.

If successful, the use of NIRS for the characterisation of natural products is of potential importance, as although the use of herbal medicines in the treatment of various conditions is becoming increasingly popular, there are as yet very few quality control procedures carried out on them. This can be a cause for concern, as herbal materials can vary in the presence and levels of their active constituents in response to a variety of factors. Thus, stringent quality control procedures would be useful in verifying that all herbal materials adhere to certain safety regulations. A rapid and non-destructive method that can also be used on site in the production environment would be advantageous in making quality control of these materials quicker, easier and possibly cheaper. The ease of operation of NIR instruments would mean that staff need not be extensively trained to carry out the procedures, thus making analysis of natural products by NIRS even more attractive.

In addition to the pharmaceutical industry, due to its rapidity and non-destructive nature, NIRS could also provide an attractive method for the identification and characterisation of forensic samples, such as unidentified and suspect plant materials seized at customs or crime scenes. Traditional chemical methods of plant drug analysis such as High Performance Liquid Chromatography (HPLC), Thin Layer Chromatography (TLC) and Gas Chromatography (GC) are destructive and require the use of harmful reagents, while physical ones like microscopy require extensive knowledge of plant morphology and also contain a margin for human error. In addition, some methods are unsuitable for certain types of plant drugs. For instance, cannabis deteriorates at high temperatures, and thus GC is not a suitable method for its analysis (White 1998). The determination of geographic

origin as well as identity of such samples would potentially be a crucial step in deducing important factors such as drug-trafficking routes, main exporters of drugs, and also to determine whether certain street seizures lead to a common supplier. Thus, if successful, the use of NIRS for the rapid characterisation of natural products has great potential, both in the pharmaceutical industry and in forensic science.

1.8 Aims and objectives

In this investigation, an attempt will be made to use NIRS to characterise natural products, in the hope that any techniques established could be applied to the analysis of herbal medicines. This would be of particular importance in that normally employed traditional techniques such as HPLC and TLC are time consuming, destructive and require extensive sample preparation. NIRS on the other hand, is rapid and requires little or no sample preparation. The aim therefore of this project is to first establish a database of various natural products and consequently, through analysing the spectra obtained, attempt to establish a method for the rapid characterisation and identification of herbal medicines. Sample presentation, and the effect of various external factors, such as temperature and moisture, as well as geographical origin, will also be taken into account. It is hoped that any techniques developed could also be used for quality control as well as forensic purposes. The aims and objectives of this project therefore were:

1. To establish a database of NIR spectra of natural products of pharmaceutical interest.
2. To establish a routine analysis identification method for natural products using NIRS, taking into account factors such as sample presentation, geographical origin, and age.
3. To develop further chemometric techniques for the analysis of herbal products using NIRS.
4. To explore the potential of NIRS for forensic and quality control purposes.

Chapter 2: Materials and Methods

2.1 Introduction

Qualitative analysis in spectroscopy depends in one way or another on comparing spectra of products under investigation with spectra of reference materials (i.e. the 'knowns'). Many methods of making such comparisons has been devised, many of which depend on powerful computer hardware. These apply powerful algorithms that allow accurate identifications to be made by distinguishing what are often very small absorbance differences. The high signal-to-noise ratios that modern NIR instruments provide are of importance, but even then, a computer is needed to separate information that cannot be always detected visually by the analyst (Mark, 1992).

In this entire study, unless otherwise stated, the following instrument and mathematical procedures were used.

2.2 Instrumentation

NIRS measurements were made using a FOSS NIRSystems (Silver Spring, MD. USA) 6500 Rapid Content Analyser™ (RCA)(Figure 2.1) in diffuse reflectance mode over the wavelength range 1100-2500nm.

The RCA connects onto the monochromator (fitted with a tungsten halogen lamp as a radiation source) and the source radiation is transmitted via fibre optics. The detectors are of lead sulphide and silicon and are arranged around the source. Samples are placed on the sampling stage and centred using an iris diaphragm. A reference measurement for diffuse reflectance is made using a ceramic plate positioned over the sampling stage.

In this study, samples were typically obtained dried and placed in 10mm diameter glass vials (Waters) (Figure 2.2). All seed samples were analysed in both their natural and powdered states, all leaf, root, stem and bark samples were either shredded or powdered, and all resins were powdered. All powdered samples were prepared using a mortar and pestle to an approximate particle diameter of 500μm. These were then placed in the sampling platform of the NIR instrument. A spectrum, being a total of 32 scans, was obtained for each sample. To verify reproducibility, this was done 12 times for each sample, shaking and tapping between each sampling. A typical sampling time was 40 seconds.

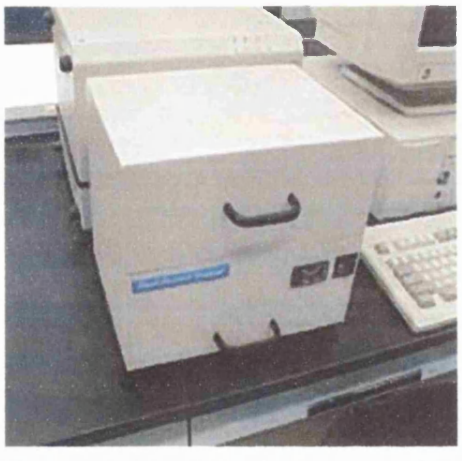


Figure 2.1. FOSS NIRSystems Rapid Content Analyser

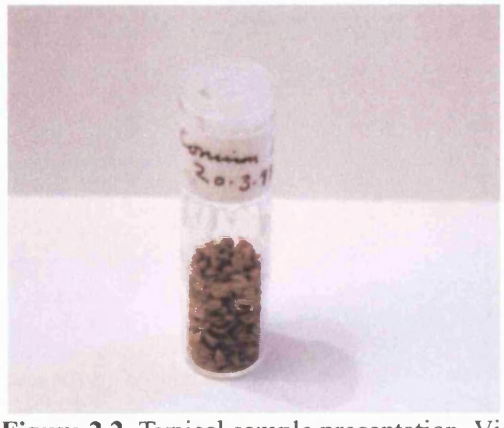


Figure 2.2. Typical sample presentation. Vial is approx. 5cm x 1cm.

2.3 Materials

Most samples of pharmaceutically important natural products were obtained in the dried state from the pharmacognosy archives at the School of Pharmacy, University of London. Twelve different samples of each natural product were obtained in most cases, dating from 1991-1998. Other samples were also obtained from various health food stores and supermarkets. Again, twelve different samples of each product were obtained where possible. Any fresh samples that were obtained were dried in an oven at 35°C until a constant weight was achieved. The relatively low temperature of 35°C was used to avoid the possibility of any volatile oils being lost in the drying process. Table 2.1 shows a list of the samples analysed.

2.4 Mathematical procedures

2.4.1 Data Pre-treatment

In the first instance, spectral data were imported into FOSS Vision® software. To remove the effects of scatter and differences in particle size, and therefore to get reproducible spectra, several spectral pretreatments were used. These included 1^{st}, 2^{nd}, 3^{rd} and 4^{th} derivatives (Osborne et al 1993), DT (Barnes et al 1989), and the use of standard normal variates (SNV) (Barnes et al 1989).

Table 2.1. Natural product samples analysed in the study

Seed	Leaf	Bark	Stem	Root	Flower	Resin	Other
Alfalfa	Belladonna	Cascara	Belladonna	Alkanna	Belladonna	Aloe: Curaçao	Capsicum
Angelica	Buchu (long)	Cassia	Hyoscyamus	Belladonna (English)	Camomile	Aloe: Socotrine	Ephedra
Anise	Buchu (oval)	*Cinchona langsifolia*	*Digitalis purpurea*	Belladonna (Indian)	Elderflower	Aloe: Zanzibar	Ergot
Blue poppy	Buchu (round)	*Cinchona officianalis*		Burdock	Henbane	Cannabis (Indian)	Mace
Halle poppy	Cannabis – South Africa	Cinnamon		Gelsemium	Lime	Cannabis (Lebanon)	Hop
Cannabis	Cannabis – India	Frangula		Gentian	Linden	Cannabis (Pakistan)	Ipecac
Caraway	Cannabis – Thailand	Liquorice		Ginger	Psyllium	Colophony	Juniper
Cardamon	Cannabis - Turkey	Quassia		Horseradish	Pyrethrum	Cutch	Lemon peel
Coriander	Coltsfoot	Slippery elm		Liquorice	Rose	Myrrh	Clove
Cumin	*Digitalis ambigua*	Witch hazel		Orris	*Digitalis mertonensis*	Sterculia	Orange peel
Dill	*Digitalis lanata*	Wild cherry		Podophyllum		Tragacanth	Quillaia
Hemlock	*Digitalis mertonensis*			Sarsaparilla			Saffron
Henbane	*Digitalis orientalis*			Valerian (English)			Senna pod: Indian
Laserpitium	*Digitalis purpurea*			Valerian (Indian)			Senna pod: Alexandrian
Mustard	Eucalyptus			Rhubarb			Senna pod: Tinnivelly
Oenanthe	Eyebright			Rauwolfia			Squill
Star anise	Henbane						
Fennel	Hop						
Ligusticum	Hyoscyamus						
	Lemon verbena						
	Maté						
	Peppermint						
	Raspberry						
	Senna - Alexandrian						
	Stramonium						
	Earl Grey tea						
	English Breakfast tea						
	Assam tea						
	Ceylon tea						
	Traditional Afternoon tea						

2.4.2 Sample Identification

Identification of samples was attempted using several identification methods. These included Maximum Distance in Wavelength Space, Correlation in Wavelength Space, and to some degree, Mahalanobis Distance in Principal Components Space and Residual Variance in Principal Components Space. Correlation Coefficient, two-wavelength analysis, PQS and *k*NN were also used. The first four methods were performed using FOSS Vision software® using the wavelength range 1100-2500nm and a 2^{nd} derivative gap size of 10nm. The other four were performed using spreadsheets written by the author in Microsoft Excel 1997.

Chapter 3: The Use of Near-Infrared Spectroscopy for the Analysis of Herbal Materials - Preliminary Studies

3.1 Introduction

This chapter describes preliminary investigations into the potential of near-infrared spectroscopy for the characterisation of herbal natural products, and in particular, identification of samples. For much of this study, members of the *Umbelliferae* family were used, due to the fact that large numbers of samples were readily available. In addition, as a number of members of the family are in common use, and as the fruits can appear superficially similar, problems may arise due to misidentification and adulteration. This can be of particular concern as toxins, such as coniine in hemlock (*Conium maculatum*) can be present in some members of the family and potentially adulterate other fruits, particuarly those of aniseed.

Later on, NIR techniques were also applied to various other natural products to investigate how successful they were in general. Samples included various roots, barks, flowers, leaves, stems, and seeds of pharmaceutical interest. The overall aim of the study was to compile a reasonable database and set up a routine analysis procedure that would enable the user to successfully identify unknown samples.

3.2 The Umbelliferae family

The *Umbelliferae* family (order *Umbelliflorae*) contains about 275 genera and 2850 species. Most members of the family are herbs with furrowed stems and hollow internodes. Some members are annuals (e.g. coriander), some biennials (e.g. hemlock) and some perennials (e.g. *Ferula*) (Evans, 1989a). The three subfamilies are Hydrocotyloideae, Saniculoideae and Apioideae (Evans, 1989a). The third subfamily is the largest one, and also the source of all the *Umbelliferae*

fruits investigated in this study. These included fennel (*Foeniculum vulgare*), aniseed (*Pimpinella anisum*), dill (*Anethum graveolens*), coriander (*Coriandrum sativum*), hemlock (*Conium maculatum*), angelica (*Angelica archangelica*), *Laserpitium* (*Laserpitium ajacis*), *Ligusticum* (*Ligusticum scoticum*) and hemlock waterdropwort (*Oenanthe crocata*). Among these, the poisonsous plants are hemlock and hemlock waterdropwort, the former containing the toxic alkaloid coniine, and the latter containing oenanthotoxin (Evans 1989a). Some of the fruits investigated in this chapter are discussed briefly below.

3.2.1 Fennel (Foeniculum vulgare)

Fennel consists of the dried ripe fruits of *Foeniculum vulgare*. Although native to southern Europe, it is now naturalised all over the world and is widely cultivated for both flavouring and medicinal use (Polunin and Robbins 1999a). The fruits contain 1-4% volatile oil (Evans 1989b) containing anethole, fenchone, limonene and apiole, as well as flavonoids including rutin, kaempferol, and quertin, and coumarins including bergapten (Polunin and Robbins 1999a). It is mainly used as an aromatic and carminative (Evans 1989b).

3.2.2 Coriander (Coriandrum sativum)

Coriander is the dried, nearly ripe spherical fruit of *Coriandrum sativum* (Evans 1989c), indigenous to Italy but now cultivated world-wide (Polunin and Robbins 1999b). An annual, it has leaves comprising many leaflets and small white flowers in the summer, followed by spherical, brittle fruit (Polunin and Robbins

1999b). The fruits contain up to 1% essential oil, with the distilled product containing 60-70% linalol, 20% monoterpene hydrocarbons (limonene, terpinene, cymene, etc), camphor, geraniol and geranyl acetate (Bisset and Wichtl 2001a). In addition, it contains various flavonoids, coumarins and phenolic acids (Polunin and Robbins, 1999b). Very large quantities of coriander are produced for domestic use as a spice, and for pharmaceutical use as a flavouring agent and carminative (Evans 1989c).

3.2.3 Dill (Anethum graveolens)

Dill consists of the dried, ripe fruits of *Anethum graveolens*, a small annual indegenous to southern Europe, cultivated in Central and Eastern Europe and Egypt and used as a carminative and flavouring agent (Evans 1989d). The volatile oil contains carvone and limone, with the European fruit yielding about 3-4% volatile oil, which should contain from 43-63% of carvone (Evans 1989d).

3.2.4 Aniseed (Pimpinella anisum)

Aniseed consists of the dried, ripe fruits of *Pimpinella anisum*, presumed to be native in the eastern Mediterranean region and western Asia (Bisset and Wichtl 2001b). Aniseed fruits yield 1.5-5% essential oil, with trans-anethol (80-90% of the oil) chiefly responsible for the taste and smell (Bisset and Wichtl 2001b). The oil's major components also include estragole, anise ketone, and β-caryophyllene (Newall et al 1996a). Other constituents include coumarins and flavonoids (Newall et al 1996a).

Occasionally, the highly toxic coniine-containing fruit of *Conium maculatum* (hemlock) are encountered in individual lots of aniseed (Bisset and Wichtl 2001b). In addition, currently, aniseed may be contaminated with up to 1% coriander fruit (Bisset and Wichtl 2001b). Aniseed is used as an expectorant and carminative, as well as a flavouring (Newall et al 1996a).

3.2.5 Angelica (Angelica archangelica)

Angelica is native to northern Europe and is biennial with a thick fleshy root, hollow stems, toothed leaves, and clusters of greenish flowers in late summer (Polunin and Robbins 1999c). Its fruit, leaves, root and rhizome are used, both in the food industry as a flavouring and in the pharmaceutical industry as a carminative, antispasmodic, diuretic, anti-inflammatory and expectorant (Newall et al 1996b). The major constituents of the fruits include 0.3-1% volatile oil containing pinene and limonene, coumarins, Archangelenone (a flavonoid) and sugars (Newall et al 1996b).

3.2.6 Hemlock (Conium maculatum)

Hemlock as a drug consists of the dried unripe fruits of the spotted hemlock, which is a poisonous biennial plant indigenous to Britain and Europe. Used historically to put criminals to death, it is the commonest of British indigenous poisonous plants (Evans 1989e). The fruit is a broad oval shape and about 3mm long. It has five prominent primary ridges that give them a beaded appearance. The endosperm is deeply groved and is surrounded by well-marked alkaloid-

containing layers. When treated with potassium hydroxide, hemlock develops a strong mousy odour caused by the liberation of the toxic alkaloid coniine. This is volatile and may be steam-distilled. It is present to the level of 1-2.5% together with N-methyl coniine, conhydrine, pseudoconhydrine, conhydrinone and γ-coniceine (Evans 1989e)

3.3 Plant structures

3.3.1 Fruit structure

The types of fruits encountered in pharmacognosy include simple, aggregate and collective fruits. Simple fruits are formed from a gynaecium with one pistil, aggregate fruits are formed from more than one pistil and collective fruits are formed from not one flower but from an inflorescence (Evans 1989f).

In the *Umbelliferae*, the fruit is a bicarpellary fruit that splits into two mericarps. The pericarp is thin and bounded by an inner and upper epidermis that resemble those of leaves. The outer epidermis may have stomata and hairs. The internal tissue in dry fruits are in the form of fibres, while secretory tissue (e.g. oil ducts) are commonly present in the pericarps of medicinal fruits. This is the case with the fruits of caraway and dill (Evans 1989f). Figures 3.1 and 3.2 shows typical fruit structures of some *Umbelliferae*.

3.3.2 Seed structure

In the analysis of seeds, care must be taken to distinguish seeds from fruits or parts of fruits containing a single seed (e.g. mericarps of *Umbelliferae* and cereals). Generally, a seed consists of a kernel surrounded by one, two or three seed coats,

with most having two. The kernel may consist of an embryo plant only or and embryo surrounded by tissues containing food reserves (Evans 1989g).

Seed anatomy can vary in terms of numbers of cell layers, structure, arrangement, colour and cell contents depending on species. The epidermis of the outer seed coat is often comprised of highly characteristic thick-celled walls that may have hairs. The storage tissues are composed of uniform cells containing characteristic cell contents such as starch, calcium oxalate, fixed oil and volatile oil (Evans 1989g).

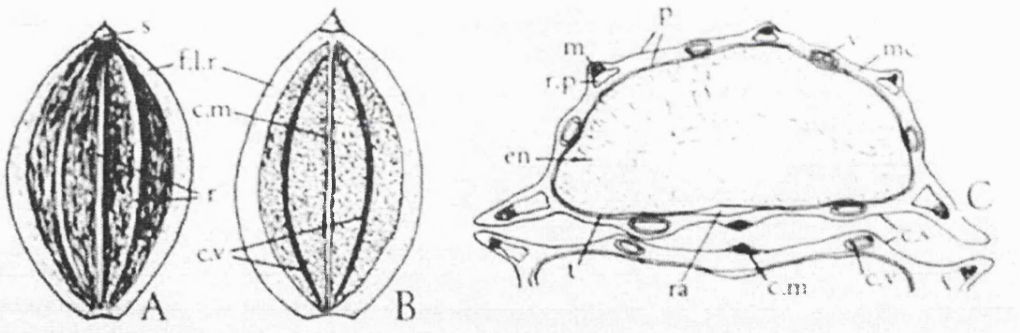


Figure 3.1. Dill fruit. A = dorsal view of a single mericarp. B = commissural surface of mericarp (both x 8). C = transverse section of fruit (x 25). c.m = carcophore meristele; c.s = commissural surfaces; c.v = commissural vitta; en = endosperm; f.l.f = flattened lateral ridges; m = meristele; mc = mesocarp; p = outer and inner epidermis of pericarp; r = ridges; ra = raphe, r.p = reticulate parenchyma, s = stylopod, t = testa, v = vitta. (Evans 1989f.)

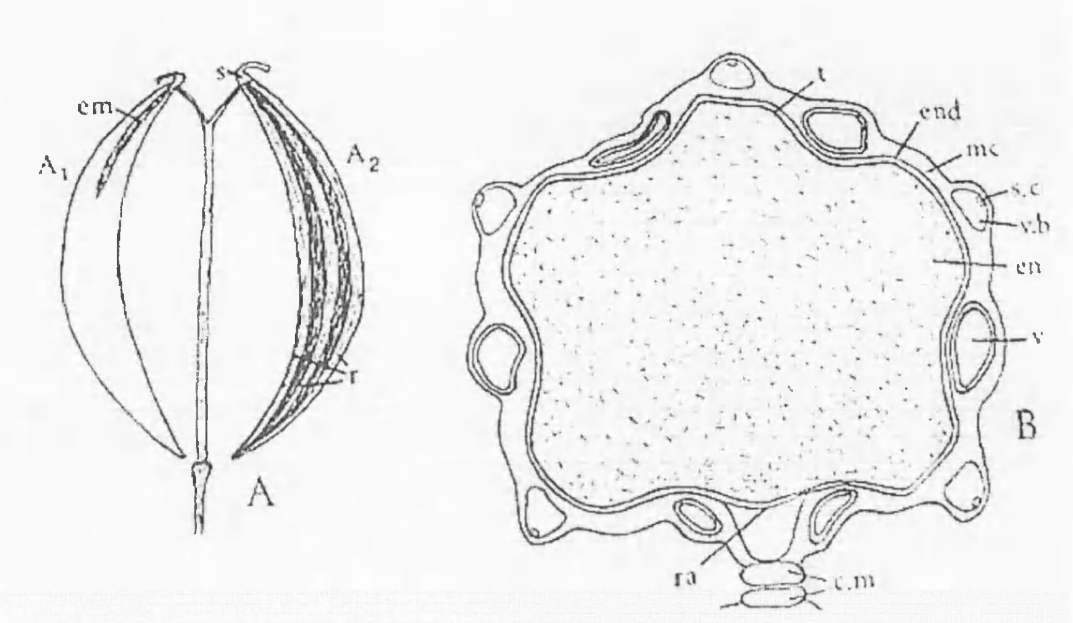


Figure 3.2. Caraway fruit. A = mericarps showing attachments to carpophore. A_1 = mericarp sectioned longitudinally to show position of embryo; A_2 = mericarp side view (x 8); B = transverse section of mericarp (x 50); em = embryo; en = endosperm; end = endocarp; mc = mesocarp; r = three of five primary ridges; ra = position of raphe; s = stylopod; s.c = secretory canal; t = testa; v = vitta; v.b = vascular bundle with associated finely pitted sclerenchyma (Evans 1989g).

The variation in structure between different fruits and seed is considerable. In seeds, carbohydrate reserves and vascular tissue are constantly present. Fruits have similar characteristics but differ in that the amount of vascular tissue is greater and lignified elements of the pericarp are often present (Evans 1989g).

3.3.3 Root structure

The primary root consists of a layer composed of a single layer of thin-walled cells devoid of cuticle and bearing root hairs formed as lateral outgrowths of the cells.

The innermost layer of the parenchymatous cortex is differentiated into an endodermis with a vascular system.

Increase in root diameter is accomplished by secondary thickening, initiated in the zone of 'fundamental parenchyma'. The cylinder of secondary tissues composed of xylem and phloem elements first tend to be arranged in regular radial rows, but often becomes less regular owing to irregular growth of the individual elements (Evans 1989h).

3.4 Aims/Objectives

The main aim of this study was to test the suitability of near-infrared spectroscopy for the characterisation of natural products, using the *Umbelliferae* fruits as a major example. Particular emphasis was placed on species identification within the family and on the importance of sample preparation. It was hoped that any successful techniques could be used on other natural products of various types such as roots and barks. If successful, NIRS could prove to be a useful tool for distinguishing between samples that may often appear very similar.

3.5 Samples

18 fennel, 5 hemlock, 3 aniseed, 3 coriander, 3 dill, 2 angelica, 2 *Ligusticum*, 2 *Laserpitium*, and 2 *Oenanthe* (hemlock waterdropwort) fruit samples were obtained from the pharmacognosy archives at the School of Pharmacy, University of London as well as from the school's botanical gardens at Myddleton House, Middlesex. In addition, some fennel, aniseed, coriander and dill samples were

obtained from various supermarkets and health-food stores. The samples were of various ages dating from 1992-1998.

In addition to the *Umbelliferae* samples, other natural products were also analysed. These included ground samples of roots such as alkanna, belladonna, burdock, gelsemium, gentian, ginger, horseradish, orris, podophyllum and sarsaparilla, as well as bark samples such as cascara, cassia, cinchona, cinnamon, frangula, liquorice, willow (*Salix*), witch hazel and wild cherry. In addition, various plant parts of some natural products were also analysed to investigate if the technique could be used for their identification. These included belladonna flowers, roots, stems, and leaves, henbane flowers, leaves, and seeds, hops and hop leaves, hyoscyamus stems and leaves, and senna leaves and pods. All were obtained from the pharmacognosy archives at the School of Pharmacy.

3.6 Results and discussion

3.6.1 Spectral characteristics

Figure 3.3 shows sample spectra of all the *Umbelliferae* fruit samples prior to any data pretreatment, while Figure 3.4 shows the same set of spectra after they have been SNV-corrected, 2nd derivative transformed. It is clear from these figures that all the spectra for the fruits of the various species are extremely alike, holding the same general shape and peaks, with just a few variations in baseline levels. Looking at just two spectra, such as those of fennel and hemlock, for example, shows that although some slight differences can be detected, it is still difficult to attribute specific spectral characteristics to a species. (Figure 3.5). It is therefore

clear that in the case of herbal materials, identification of samples merely by visual inspection of the spectra can be difficult.

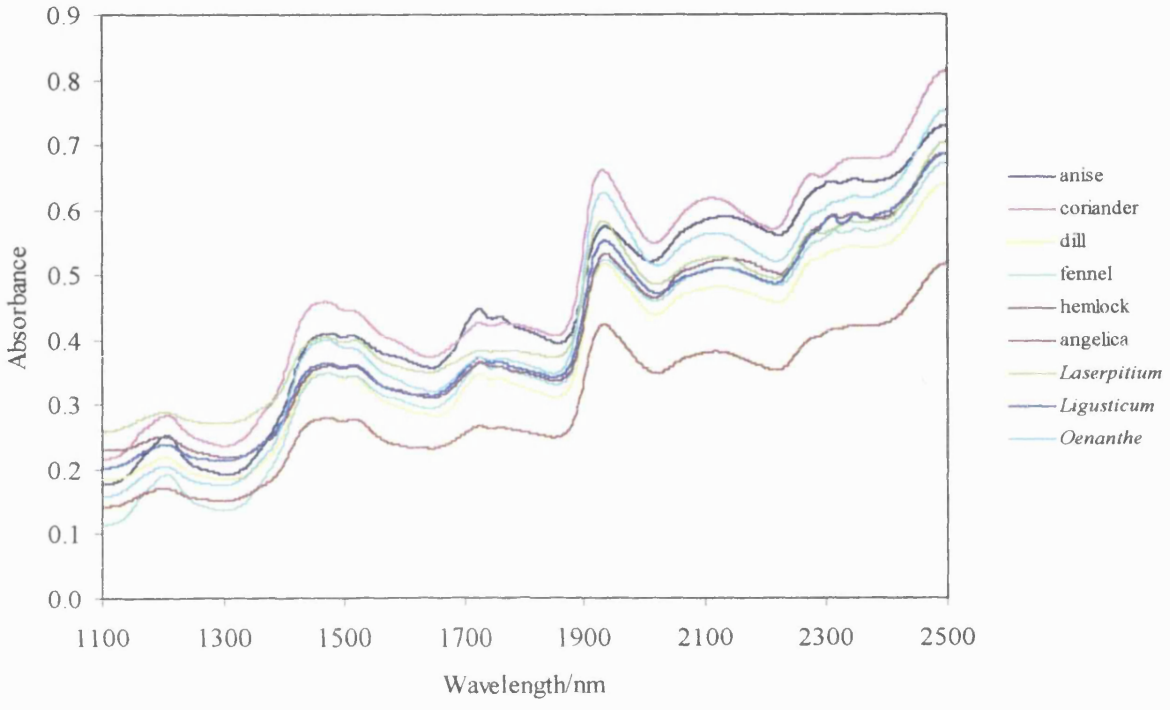


Figure 3.3. Sample spectra of fruit of 9 *Umbelliferae* species in their crude (unground) state

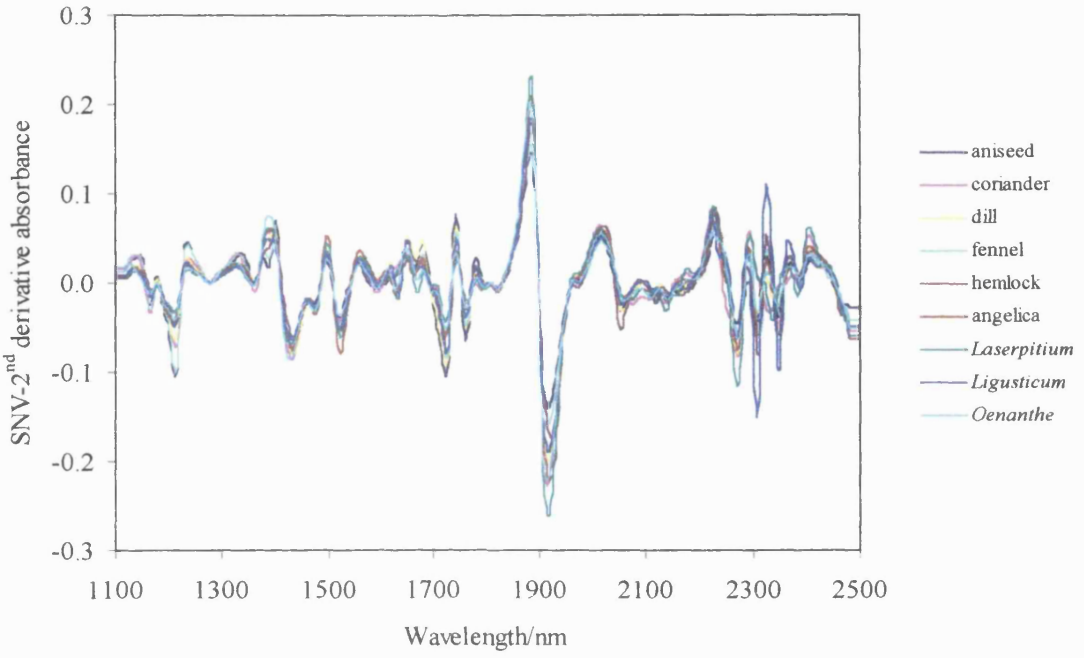


Figure 3.4. SNV-transformed, 2nd derivative corrected spectra of fruit of 9 *Umbelliferae* samples in their crude (unground) state

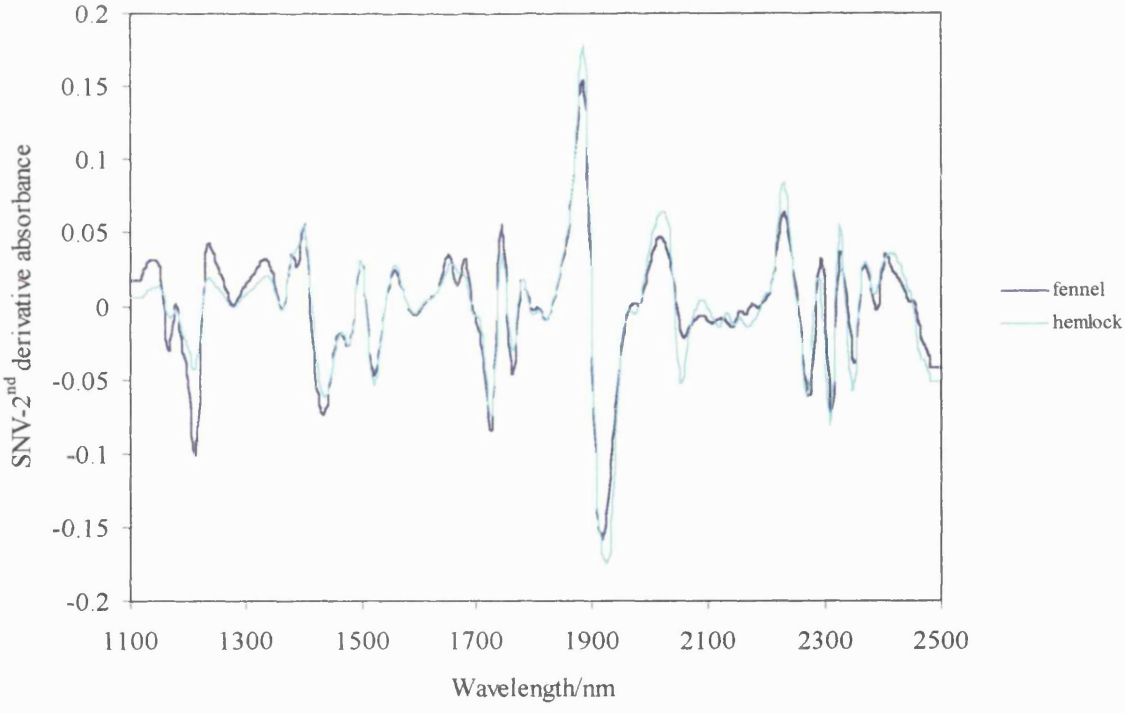


Figure 3.5. SNV-corrected, 2nd derivative spectra of unground fennel and hemlock fruits

Figure 3.6 shows the SNV-corrected, 2nd derivative spectra of the same samples in their prepared (ground) state. As in the case of crude samples, spectral similarities are still seen, making identification by visual methods virtually impossible unless one were to use a modern NIR instrument constructed with a microscope attachment. However, testing this type of instrument on herbal materials was beyond the scope and time frame of this project.

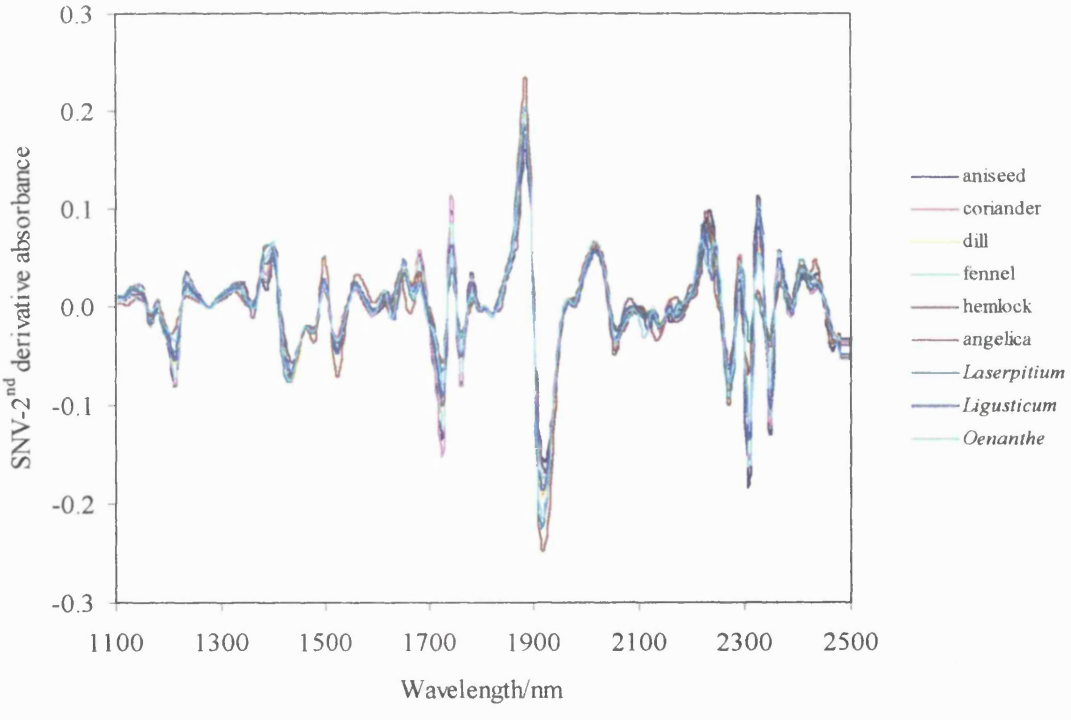
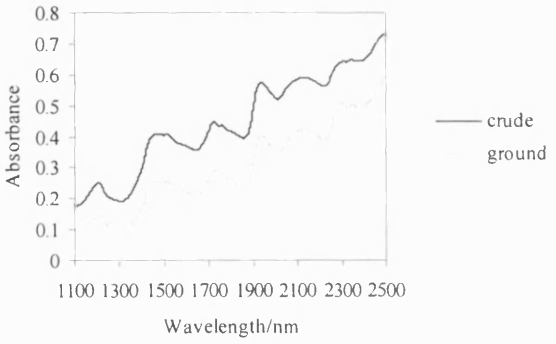
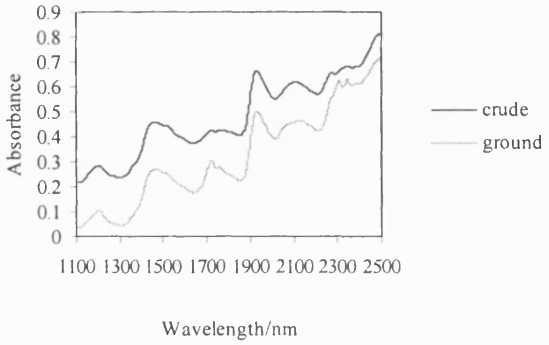


Figure 3.6. SNV-corrected, 2nd derivative spectra of nine prepared (ground) *Umbelliferae* samples

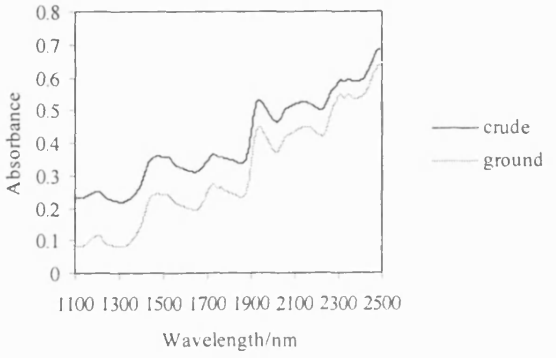
Figure 3.7A-3.7I shows sample spectra of each of the nine *Umbelliferae* species in both their ground and unground states. It can be seen from this that for all cases, the spectra for both sample states are extremely similar, with just a small difference in baseline levels, with the ground spectra having a lower baseline. This is most likely explained by the smaller particle size (approximately 500μm compared to an unground particle size of approximately 3-5mm) of the ground samples. Figure 3.8A-3.8I shows the same spectra after they have been SNV-corrected, 2^{nd} derivative transformed. In these, slight differences between samples are made more visible due to the sharper and more well-defined peaks.



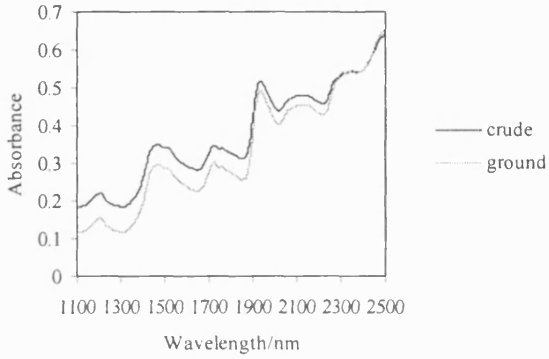
3.7A. Aniseed



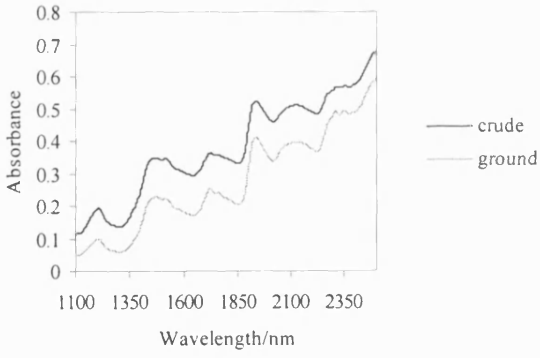
3.7B. Coriander



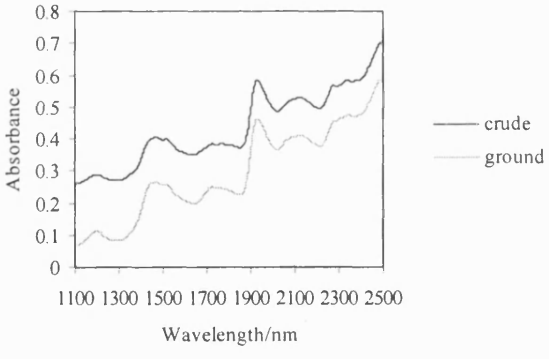
3.7C. Hemlock



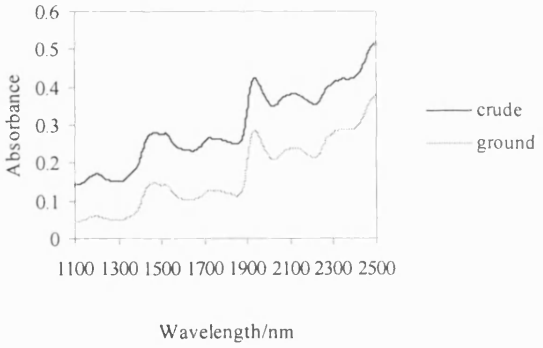
3.7D. Dill



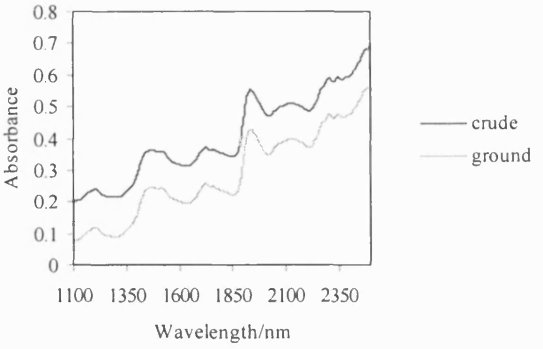
3.7E. Fennel



3.7F. *Laserpitium ajacis*



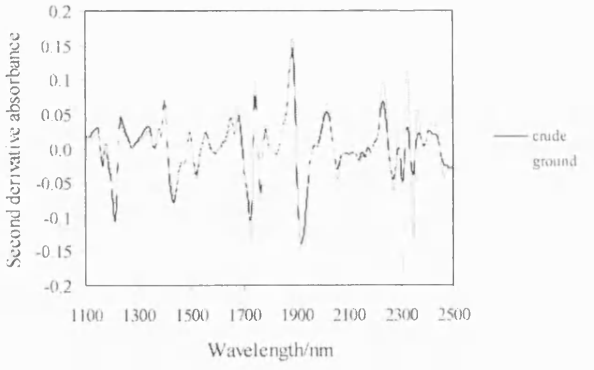
3.7G. Angelica



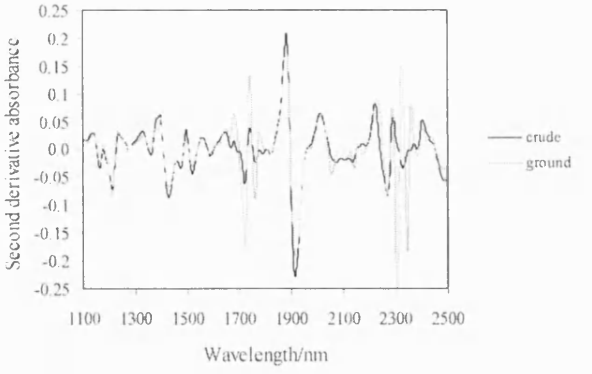
3.7H. *Ligusticum scoticum*

3.7I. *Oenanthe crocata* (hemlock waterdropwort)

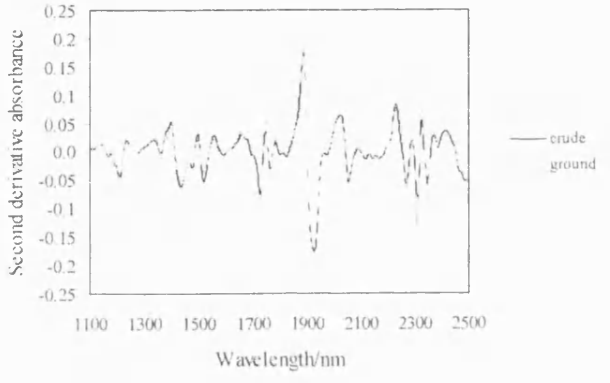
Figure 3.7A-3.7I. Sample spectra of crude and ground Anise (A), Coriander (B), Hemlock (C), Dill (D), Fennel (E) *Laserpitium* (F), Angelica (G), *Ligusticum* (H) and *Oenanthe* (I)



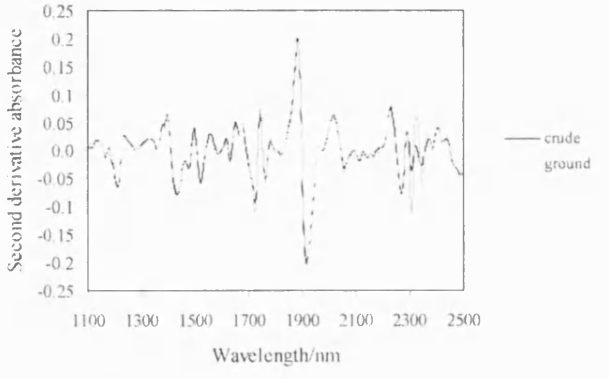
3.8A. Aniseed



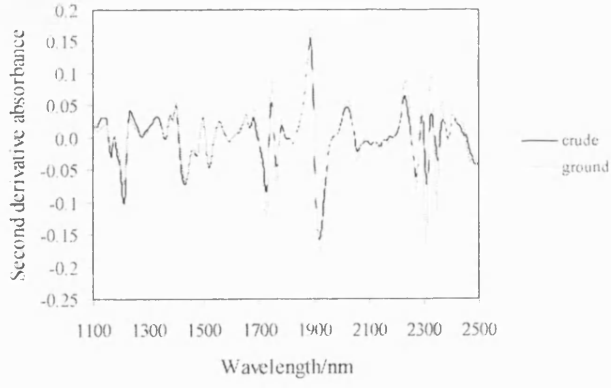
3.8B. Coriander



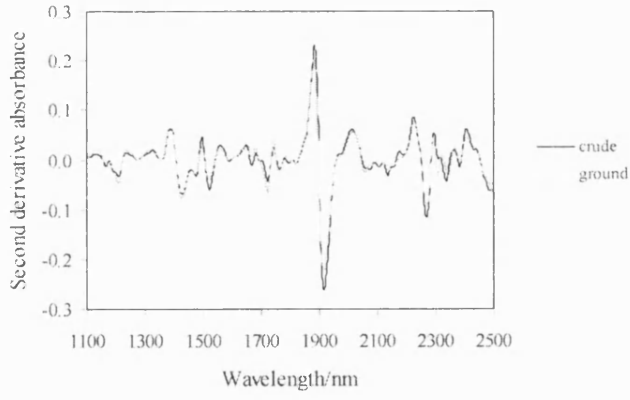
3.8C. Hemlock



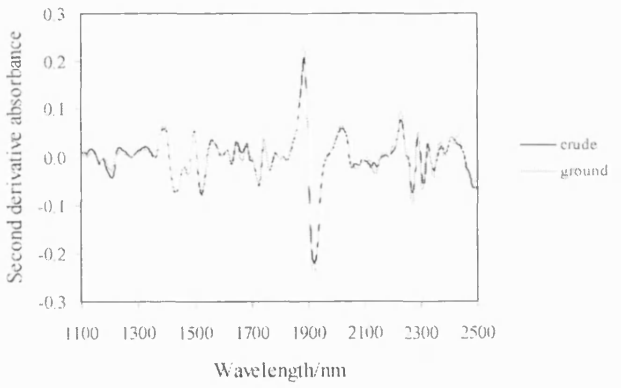
3.8D. Dill



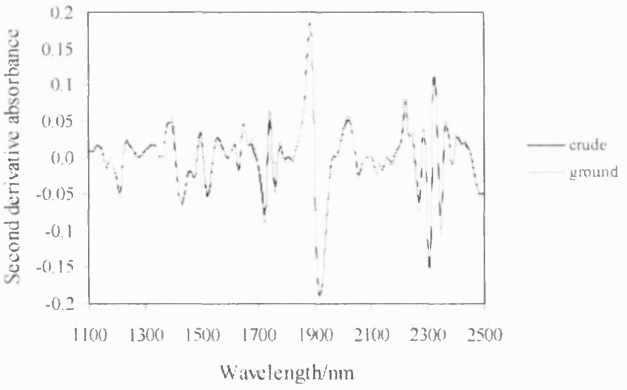
3.8E. Fennel



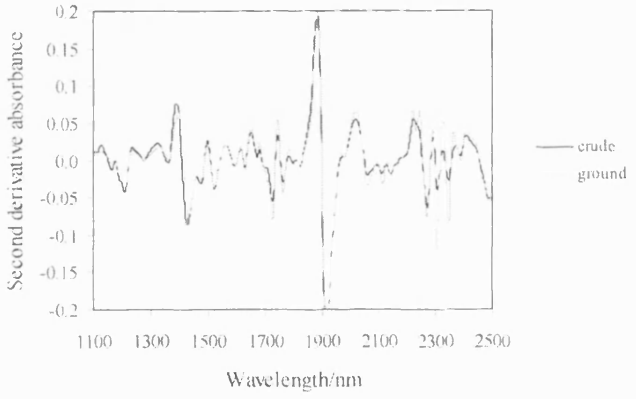
3.8F. *Laserpitium ajacis*



3.8G. Angelica



3.8H. *Ligusticum scoticum*



3.8I. *Oenanthe crocata* (hemlock waterdropwort)

Figure 3.8A-3.8I. SNV-2nd derivative spectra of crude and ground Anise (A), Coriander (B), Hemlock (C), Dill (D), Fennel (E) *Laserpitium* (F), Angelica (G), *Ligusticum* (H) and *Oenanthe* (I)

3.6.2 Data analysis

3.6.2.1 Identification method development- Maximum Distance in Wavelength Space

As the SNV-second derivative spectra resulted in sharper, more well defined peaks and some degree of differentiation between samples, this was used as the data pretreatment of choice. Figures 3.9-3.14 shows spectra of representative fennel and hemlock seeds in their untreated forms and after various data pre-treatments. Among the various data analysis algorithms available in the FOSS Vision® software, the Maximum Distance in Wavelength Space method appeared to be the most successful, and so was used as the predominant method here. Parameters used included a wavelength range of 1100-2500nm and a second derivative gap size of 10nm. The default threshold match value in the software is 4, and this was used initially.

3.6.2.2 Umbelliferae samples

Table 3.1 shows resulting match values for unground *Umbelliferae* samples after using Maximum Distance in Wavelength Space on SNV-corrected, 2nd derivative-transformed spectra. It can be seen from this that the method allowed the correct identification of each type of sample. Match values for correctly matched samples ranged from 1.30 for *Ligusticum* to 3.77 for hemlock. For mismatched samples, values ranged from 10.2 for anise against hemlock to 117 for dill against *Oenanthe*. As there is a substantial gap of 6.43 between the highest correctly matched value and the lowest mismatched match value, it can be said that this method is quite effective in identification of samples. The largest correct match value was rather high, being 3.77. However, as it was below 4.0, the default still

applies. For ground samples, again, the method was able to correctly discriminate the fruits, with correctly matched values ranging from 1.47 for hemlock to 3.84 for fennel (Table 3.2). For incorrectly matched ground samples, values ranged from 7.35 for coriander against hemlock to 90.5 for coriander against *Oenanthe*. Again, although the highest correct match value (3.84) was higher than the ideal, the default threshold value appears to be sufficient.

However, 100% identification was not achieved. That is, in fennel, 3 out of 216 spectra gave rise to failed match values of 4.21, 4.97 and 4.39 for unground samples and 4.28, 4.34 and 4.03 for ground samples. These were all in sample 18, and the reason for the failures occurring in this batch will be investigated and discussed later in the chapter. For the sake of clarity, these failed values were not included in Tables 3.1 and 3.2.

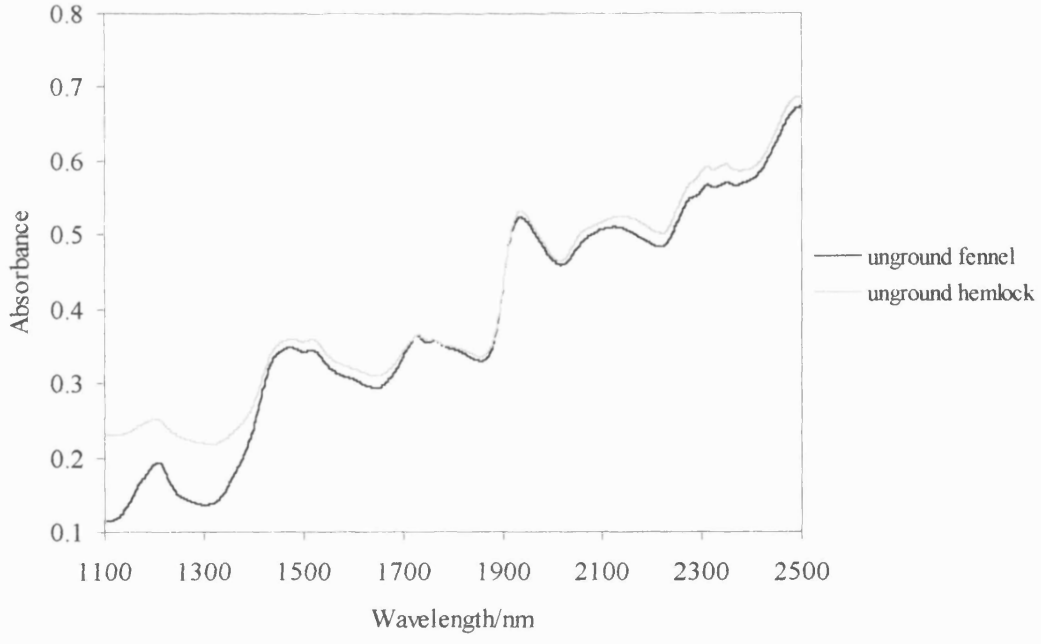


Figure 3.9. Sample spectra of unground fennel and hemlock fruits

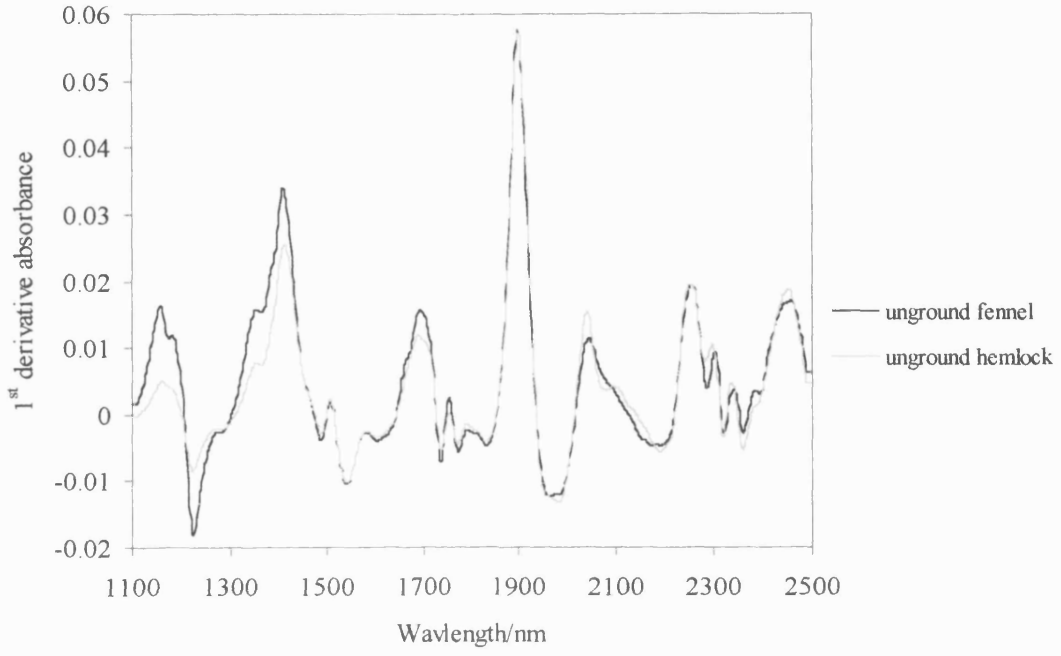


Figure 3.10. 1st derivative spectra of unground fennel and hemlock fruits

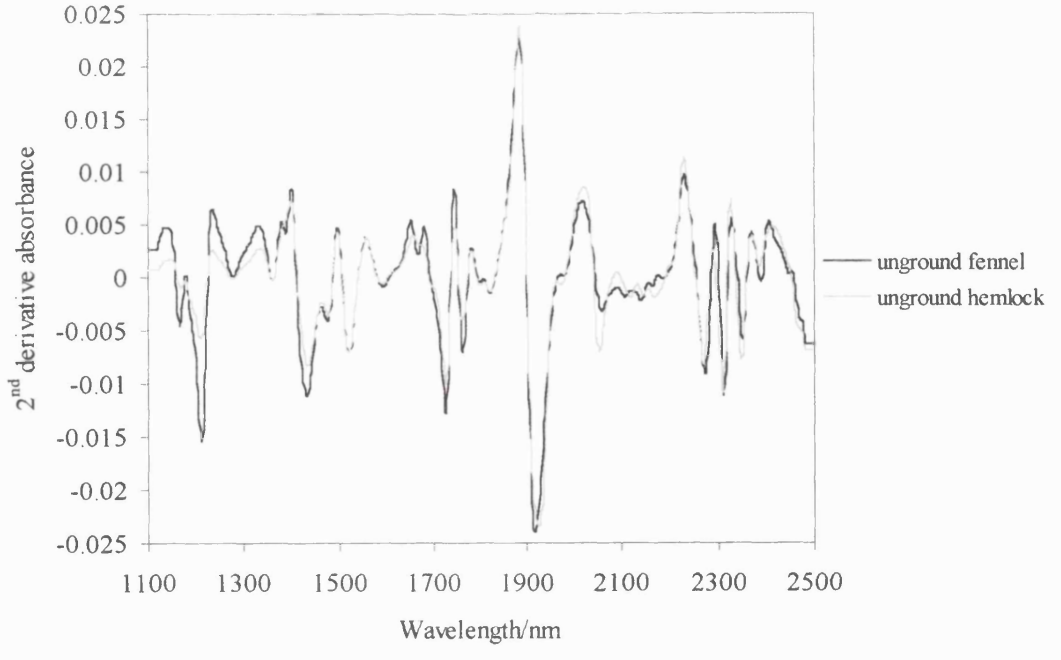


Figure 3.11 2nd derivative spectra of unground fennel and hemlock fruits

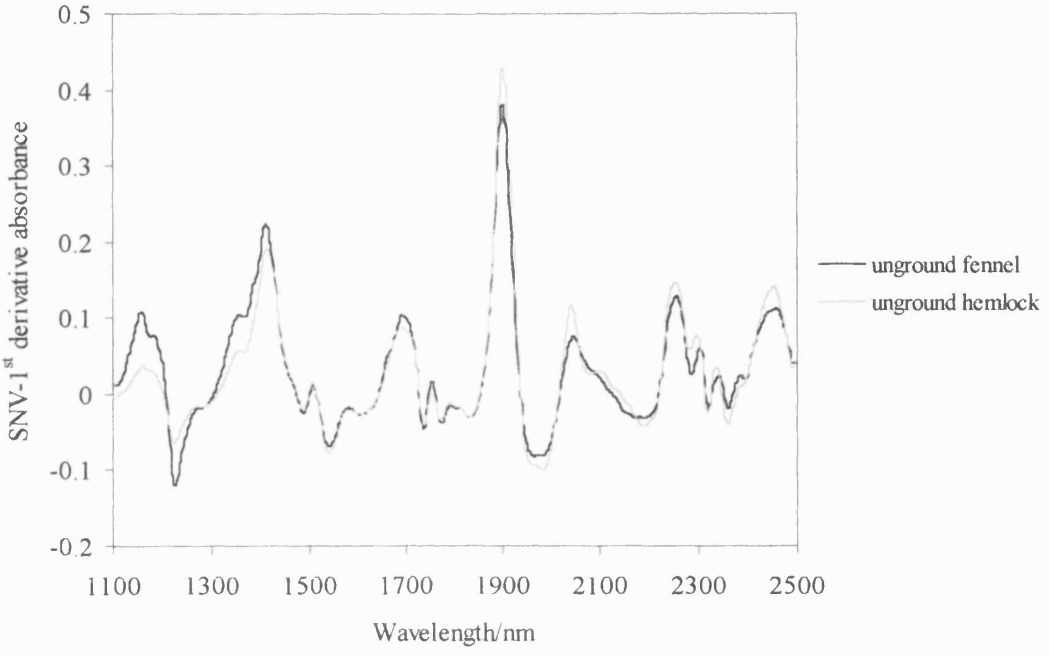


Figure 3.12. SNV-1st derivative spectra of unground fennel and hemlock fruits

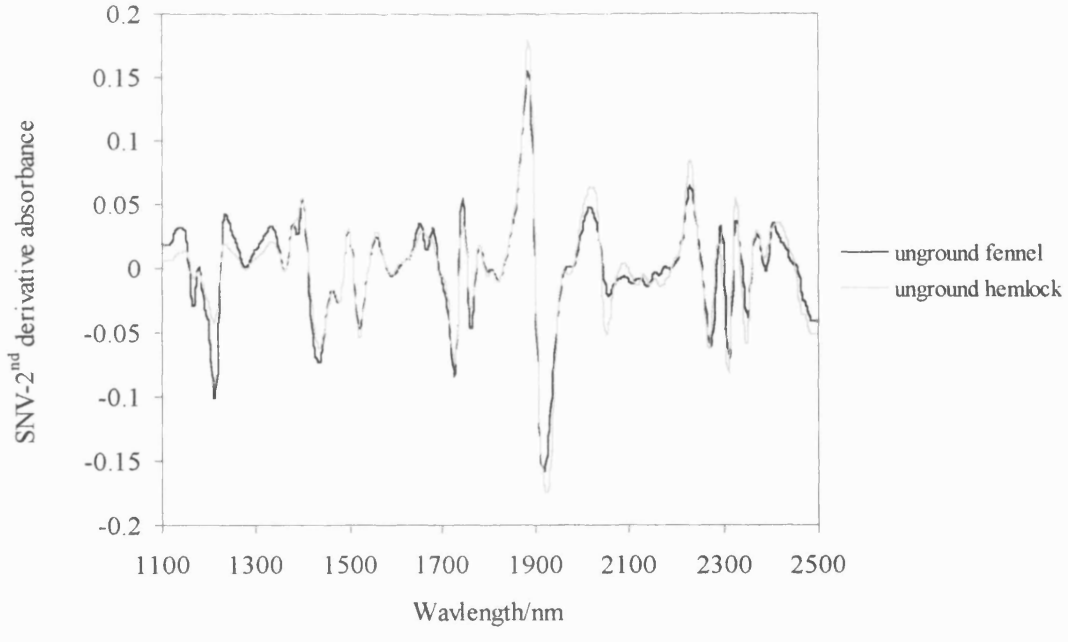


Figure 3.13. SNV-2nd derivative spectra of unground fennel and hemlock fruits

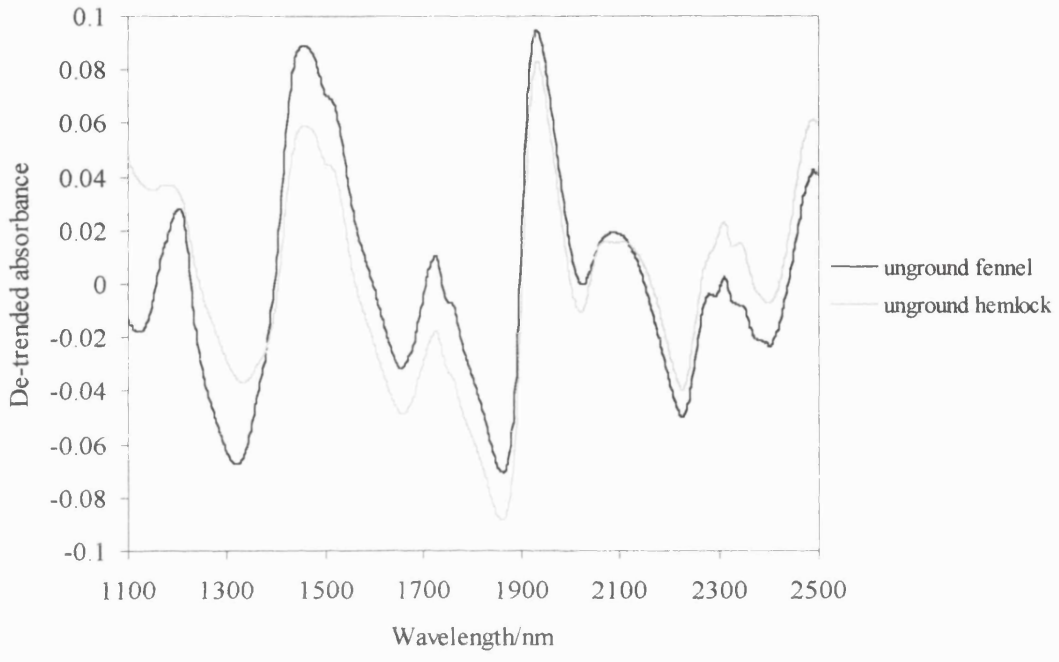


Figure 3.14 De-trended spectra of unground fennel and hemlock fruits

3.6.2.3 Sample preparation

Table 3.3 shows match values using the Maximum Distance in Wavelength Space method for the *Umbelliferae* samples when both ground and unground samples were combined in each category, excluding any 'failures' and 'ambiguous' results. The majority of the samples were correctly identified using this method, with correct match values ranging from 1.20 for *Oenanthe* to 3.99 for fennel and mismatched values ranging from 4.45 for fennel against anise to 57.0 for angelica against *Ligusticum*. However, some failures and ambiguous results resulted in the cases of fennel and hemlock. In fennel, 12 identification failures were observed, ranging from 4.20 to 5.69. In addition, out of the 216 spectra, 18 'ambiguous' results were also obtained, where the program identified the fennel samples as both fennel and hemlock, giving two match values. For example, in one fennel

spectrum, a match value of 2.37 was produced for that sample against fennel, as well as a value of 3.88 against hemlock. However, in all cases, the lower match value gave the correct identity, with the higher value being the incorrect match with hemlock. Lower match values ranged from 1.53 to 2.57 and higher match values ranged from 3.44 to 3.95.

'Ambiguous' results were also observed in 6 hemlock cases, with the program identifying those samples as both hemlock and fennel. Lower match value in these ranged from 1.69 to 2.58 and higher ones ranged from 3.58 to 3.94. As in the fennel samples, all the lower match values were for correct matches (hemlock) and the higher ones for incorrect matches (fennel).

Because of the 7.1% failure rate in fennel and the 5% failure rate in hemlock, it may be concluded that it is not a feasible practice to combine two sample preparations together when attempting identification. Therefore, keeping a sample preparation method consistent is of importance, and this may suggest that a recommended technique would be to grind all samples prior to analysis to an approximate particle size of 500μm, if they are obtained in the crude state. It appears that as well as chemical characteristics, physical ones also play a key part in identification.

Table 3.4 shows match values for ground and unground *Umbelliferae* samples using the Maximum Distance in Wavelength Space method. From this, it can be said that the method can correctly distinguish between ground and unground forms of the same sample, with match values ranging from 5.50 for fennel to 47.8 for *Oenanthe.*

Table 3.1 Match values for unground *Umbelliferae* samples using Maximum Distance in Wavelength Space

	Fennel	Anise	Dill	Coriander	Hemlock	Angelica	*Laserpitium*	*Ligusticum*	*Oenanthe*
Fennel	1.56-2.39-3.56								
Anise	67.2-77.0-84.1	1.74-2.00-2.19							
Dill	23.5-25.7-25.8	24.0-25.6-26.8	1.51-1.81-2.24						
Coriander	41.2-49.8-58.2	17.1-19.5-20.7	29.8-31.1-36.0	1.50-2.11-2.87					
Hemlock	10.9-11.3-11.6	10.2-10.7-17.2	21.6-22.9-23.7	28.0-30.4-31.9	1.53-2.21-3.77				
Angelica	25.9-29.8-31.8	39.4-41.5-44.8	27.8-31.7-35.1	36.2-40.7-43.1	25.2-30.5-38.0	1.43-2.04-2.41			
Laserpitium	29.8-31.1-32.3	36.3-43.1-47.5	32.4-33.8-35.2	77.5-83.7-88.5	32.8-35.2-69.1	48.7-51.4-58.3	1.34-2.06-2.36		
Ligusticum	32.6-36.4-38.2	29.3-31.9-34.0	38.6-41.3-46.7	19.5-21.8-25.2	20.0-25.6-39.0	39.0-41.7-48.3	31.8-33.6-36.4	1.30-2.04-2.27	
Oenanthe	66.3-75.3-80.9	72.1-80.8-91.7	102-111-117	50.6-52.2-56.4	59.1-86.1-86.7	22.2-23.1-24.9	33.5-34.5-38.1	97.3-109-134	1.77-2.07-2.22

Left-hand number = smallest value; right-hand number = largest value; underlined number = median value

Table 3.2 Match values for ground *Umbelliferae* samples using Maximum Distance in Wavelength Space

	Fennel	Anise	Dill	Coriander	Hemlock	Angelica	*Laserpitium*	*Ligusticum*	*Oenanthe*
Fennel	1.19-2.26-3.84								
Anise	43.4-45.4-47.8	1.68-1.87-2.49							
Dill	30.5-32.2-41.3	30.4-34.9-36.2	1.56-1.93-2.43						
Coriander	14.3-15.0-25.2	41.5-53.6-58.4	19.9-38.9-47.3	1.55-1.95-3.02					
Hemlock	8.19-8.22-8.49	9.76-11.8-12.4	11.9-15.5-37.8	7.35-8.87-10.9	1.47-2.16-3.11				
Angelica	47.5-48.8-49.5	31.6-32.7-33.7	25.8-28.2-32.2	33.7-40.5-57.1	23.4-32.5-45.3	1.51-1.70-2.37			
Laserpitium	56.9-59.6-63.6	44.1-46.8-47.7	28.3-31.3-34.3	29.9-37.4-48.6	28.3-32.9-54.7	55.4-63.2-73.7	1.67-1.62-2.21		
Ligusticum	46.4-47.0-71.0	64.1-67.7-68.3	33.0-37.9-40.9	43.3-54.8-73.2	33.9-39.7-59.1	39.0-44.4-51.9	25.4-26.5-28.5	1.58-1.87-2.50	
Oenanthe	49.1-51.7-63.8	49.5-51.9-52.4	37.3-40.5-45.3	56.5-66.7-90.5	39.4-57.5-61.1	49.4-52.8-62.0	66.7-71.8-90.2	52.1-53.4-61.7	1.65-1.88-2.45

Left-hand number = smallest value; right-hand number = largest value; underlined number = median value

Table 3.3 Match values for *Umbelliferae* samples using Maximum Distance in Wavelength Space, combining both ground and unground samples

	Fennel	Anise	Dill	Coriander	Hemlock	Angelica	*Laserpitium*	*Ligusticum*	*Oenanthe*
Fennel	1.42-2.33-3.99								
Anise	4.45-5.44-5.81	1.41-1.72-2.88							
Dill	5.16-6.42-10.3	10.2-10.4-13.1	1.71-2.36-2.72						
Coriander	4.98-6.66-8.13	25.6-42.8-53.6	11.7-17.7-30.9	1.41-2.56-3.41					
Hemlock	4.68-5.32-6.42	8.02-8.96-11.0	10.8-14.0-27.1	6.59-9.24-13.3	1.31-2.67-3.63				
Angelica	16.4-16.4-18.0	19.9-21.2-24.8	15.2-17.4-22.7	21.8-24.2-38.0	13.6-17.6-24.4	1.51-2.30-3.19			
Laserpitium	5.62-7.30-8.05	25.4-26.8-32.1	14.1-18.3-34.9	16.7-22.9-33.4	17.8-20.2-35.0	33.2-35.4-42.2	1.36-2.30-2.89		
Ligusticum	6.56-7.66-10.9	16.1-20.0-27.4	14.3-15.9-19.3	19.4-24.3-45.3	11.9-17.4-35.6	37.8-46.0-57.0	13.3-16.6-23.3	1.25-2.15-2.97	
Oenanthe	10.0-12.6-14.0	28.4-29.5-41.5	19.0-23.2-46.0	25.4-34.9-51.9	18.9-28.5-44.5	19.3-23.4-26.1	14.1-19.0-33.0	28.6-34.7-40.9	1.20-1.77-2.63

Left-hand number = smallest value; right-hand number = largest value; underlined number = median value

Table 3.4 Match values for ground and unground *Umbelliferae* samples using Maximum Distance in Wavelength Space

Unground	Fennel	Anise	Dill	Coriander	Hemlock	Angelica	*Laserpitium*	*Ligusticum*	*Oenanthe*
	5.50-6.42-7.95	33.9-37.9-46.2	13.7-16.4-18.0	23.9-26.5-27.9	9.28-10.3-11.4	10.3-13.7-22.0	10.1-11.0-15.4	15.2-20.1-28.9	39.3-43.0-47.8
Ground	Fennel	Anise	Dill	Coriander	Hemlock	Angelica	*Laserpitium*	*Ligusticum*	*Oenanthe*

Left-hand number = smallest value; right-hand number = largest value; underlined number = median value

3.6.2.4 Failed fennel spectra

As mentioned earlier, initial analysis of the *Umbelliferae* samples yielded three failures in the fennel samples in both the ground and unground states. These were all produced by the same sample (sample number 18), and thus this sample was examined more closely. It was found that this specimen was considerably newer than the other 17 fennel seed batches by over a year. Thus, it was suspected that due to the relative freshness of this sample, a larger amount of moisture was present, causing it to produce some failures when compared to the other older seeds. To test this theory, the sample was taken and dried in an oven at 70°C, taking spectra and weights every half-hour until a consistent weight was reached. At the end of this process, the original set of spectra for fennel number 18 was replaced by the new set of spectra of the drier sample and identification was again attempted.

Figure 3.15 shows the spectra obtained from unground fennel number 18 as it dried in the oven. From this, it can be seen that the peaks appear to decrease with each time interval, suggesting that excess moisture is being driven away. A close-up of the vicinity near the peak for water (approximately 1930nm) (Figure 3.16) shows this more clearly.

When the spectra for the dried fennel number 18 was used instead of the original one, it was possible to achieve 100% identification of all samples, in both the coarse and ground states. Match values were all below 4 for correctly matched samples. Thus, it can be said that the reason for the initial failure was due to different amounts of moisture in the samples. This may suggest that in order for

the method to work efficiently, samples need to be relatively similar in terms of dryness. Indeed, initially, fennel number 18 looked considerable greener in appearance.

Thus, a useful approach would be to dry all new or fresh samples prior to analysis in order to reduce failures caused by different amounts of moisture.

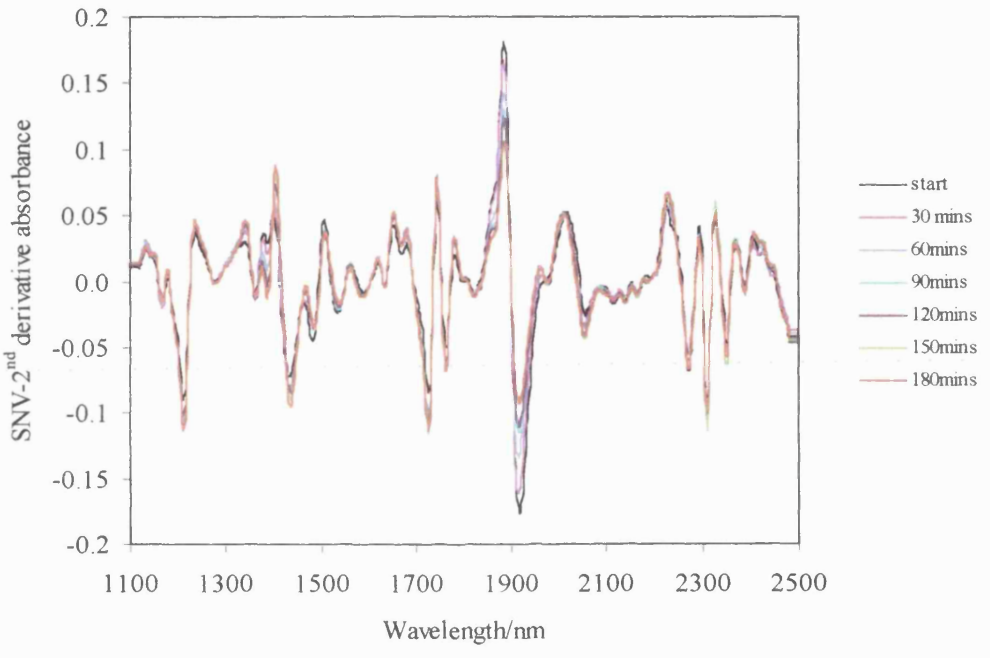


Figure 3.15. SNV-2nd derivative spectra of fennel number 18 (unground) dried over time

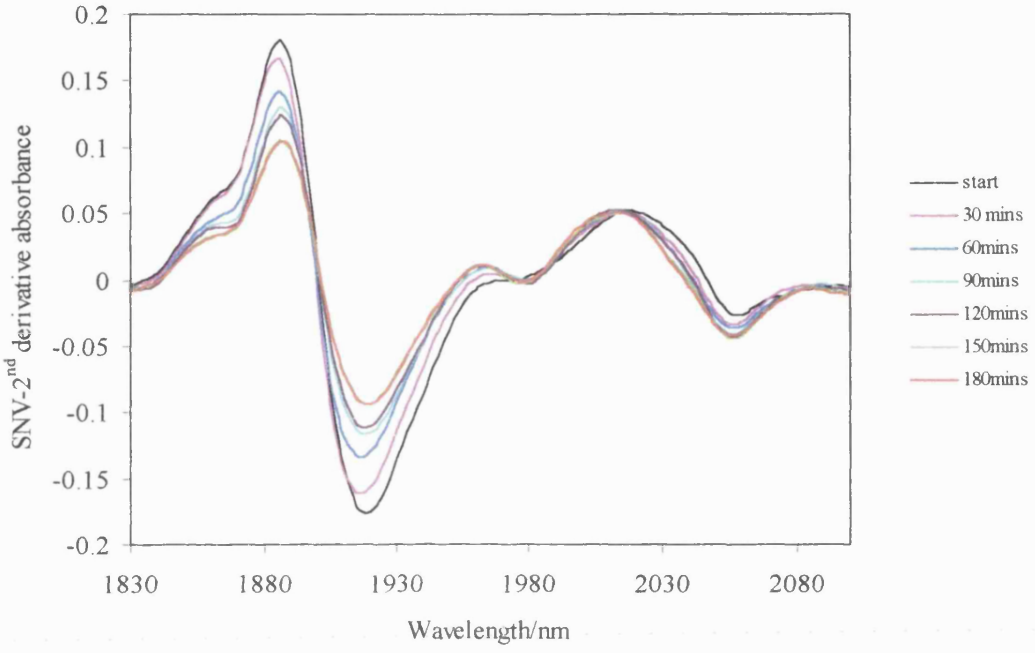


Figure 3.16. SNV-2nd derivative spectra of fennel number 18 (unground) between 1830 and 2100nm, dried over time

3.6.2.5 Other samples

Similar techniques were used on other natural product samples in order to check their range of applicability. Twelve different samples of each natural product were examined in their ground state, which had been prepared to an approximate particle size of 500μm.

3.6.2.5.1 Roots

Table 3.5 shows match values obtained when using the Maximum Distance in Wavelength Space method on various root samples. It is clear from this that the method is effective in correctly identifying all of the roots, with correct match

values ranging from 1.87 for alkanna to 2.47 for burdock. Successful discrimination values ranged from 25.2 for alkanna against horseradish to 163.9 for gelsemium against orris. In this case, the threshold match value of 4 appears to be a reasonable one.

3.6.2.5.2 Barks

Table 3.6 shows match values for various bark samples when using the Maximum Distance in Wavelength Space algorithm. Again, the method appears to be effective in discriminating all the samples, with correctly matched values ranging from 2.15 for *Salix* to 2.62 for cassia. For samples that were correctly distinguished, values were from 9.57 for cascara against cassia to 161.1 for cinnamon against liquorice. Again, as there is a substantial gap of 6.95 between the largest correctly matched value and the smallest mismatched value, the threshold of 4.0 is a reasonable one for identification.

3.6.2.5.3 Plant parts

Table 3.7 shows match values obtained for various plants and their plant parts. From this, it can be seen that although samples may come from the same plant, their different parts can be successfully identified using the Maximum Distance in Wavelength Space method. Correctly matched values ranged from 2.12 for belladonna flowers to 2.40 for henbane seeds and hyoscyamus pods. Mismatches ranged from 22.2 for hyoscyamus pods against leaves to 134.5 for hops and hop leaves. The very high lower limit for successful mismatch values suggests that as in all the other cases, the threshold can be kept at 4.0.

It was also possible to demonstrate that the method can work in a general way as well. Table 3.8 shows match values obtained when all the plant parts from Table 3.7 were combined for each plant. That is, for example, belladonna flowers, roots, stems, and leaves were combined and treated as belladonna in general, while senna leaves and pods were combined and treated as senna in general. Surprisingly, the method was able to correctly identify all the plants even when they were a combination of various parts. Correctly matched values ranged from 1.98 for henbane to 2.32 for hop and mismatched values ranged from 17.6 for henbane against hyoscyamus to 76.8 for hop against senna. Thus the method can be used in both a specific and general way.

Table 3.5. Match values for various root samples using Maximum Distance in Wavelength Space

	Alkanna	Belladonna	Burdock	Gelsemium	Gentian	Ginger	Horseradish	Orris	Podophyllum	Sarsaparilla
Alkanna	1.87									
Belladonna	48.3	1.92								
Burdock	108.8	78.9	2.47							
Gelsemium	50.9	91.3	91.2	2.29						
Gentian	101.2	61.0	63.6	91.6	2.42					
Ginger	128.7	103.1	116.6	154.4	132.6	2.30				
Horseradish	25.2	28.5	30.3	50.4	28.2	27.7	2.28			
Orris	127.1	69.7	72.5	163.9	136.4	53.6	113.8	2.51		
Podophyllum	43.6	36.7	48.4	80.6	66.5	41.3	39.8	27.1	2.23	
Sarsaparilla	50.6	31.4	58.4	28.3	83.5	50.5	75.8	39.3	44.8	2.07

Numbers are average values. n = 144 spectra for each sample

Table 3.6. Match values for various bark samples using Maximum Distance in Wavelength Space

	Cascara	Cassia	*Cinchona langsifolia*	*Cinchona officianalis*	Cinnamon	Frangula	Liquorice	*Salix*	Witch Hazel	Wild Cherry
Cascara	2.24									
Cassia	9.57	2.62								
Cinchona langsifolia	57.8	40.1	2.19							
Cinchona officianalis	58.9	96.0	45.3	2.36						
Cinnamon	63.3	99.7	38.5	51.3	2.27					
Frangula	18.5	22.7	37.7	48.0	33.7	2.29				
Liquorice	45.5	116.4	88.9	62.8	161.1	102.6	2.36			
Salix	22.0	30.9	70.6	58.0	34.4	23.8	61.8	2.15		
Witch Hazel	14.8	22.3	30.2	32.4	17.5	15.5	17.2	16.7	2.25	
Wild Cherry	20.0	25.8	30.8	29.0	17.8	20.5	30.4	18.1	41.0	2.36

Numbers are average values. n = 144 spectra for each sample

Table 3.7. Match values of various plants and their parts using Maximum Distance in Wavelength Space

	Belladonna flower	Belladonna root	Belladonna stem	Belladonna leaf
Belladonna flower	2.12			
Belladonna root	121.4	2.29		
Belladonna stem	34.4	23.5	2.35	
Belladonna leaf	48.0	61.6	90.6	2.19
	Henbane flower	Henbane leaf	Henbane seed	
Henbane flower	2.29			
Henbane leaf	45.1	2.24		
Henbane seed	50.0	51.1	2.4	
	Hyoscyamus pod	*Hyoscyamus* stem	*Hyoscyamus* leaf	
Hyoscyamus pod	2.4			
Hyoscyamus stem	26.1	2.36		
Hyoscyamus leaf	22.2	30.7	2.37	
	Hop	Hop leaf		
Hop	2.23			
Hop leaf	134.5	2.34		
	Senna leaf	Senna pod		
Senna leaf	2.38			
Senna pod	26.6	2.28		

Numbers are average values. n = 12 for each plant part

Table 3.8. Match values of combined plant parts for various plants using Maximum Distance in Wavelength Space

	Belladonna	Henbane	Hop	*Hyoscyamus*	Senna
Belladonna	2.07				
Henbane	22.5	1.98			
Hop	22.8	27.2	2.32		
Hyoscyamus	27.1	17.6	40.9	2.15	
Senna	39.8	50.4	76.8	70.7	2.12

Numbers are average values. n = 12 for each plant part

3.7 Conclusion

This preliminary study has shown that near-infrared spectroscopy in combination with the Maximum Distance in Wavelength Space method can be used as a successful tool in the characterisation of herbal materials. In the case of the *Umbelliferae* samples, successful identification of samples was achieved in both crude and prepared (ground) states, although the libraries contained samples of various ages and conditions. Combining ground and unground samples was also successful to some extent, although the few failures and ambiguous results suggest that when analysing samples, they should all be consistent in the way they were prepared. Thus, a useful suggestion may be to grind all future samples to a similar particle size (500 μm) prior to analysis. It was also found that moisture may affect the successful discrimination of samples, as was the case of fennel number 18. Thus, it may be advisable in the future to dry all relatively new or fresh materials. Driving off any excess moisture appeared to be successful in solving the observed problem in initial identification.

In the case of other plant materials, the method was equally successful, suggesting that NIRS would be a powerful technique in the identification of any unknown sample of vegetable origin. It also appeared successful in specifically discriminating various parts of the same plant. On a more general level, it was also able to categorise various plant libraries made up of different plant parts combined. This may be of particular use in forensic science, when often the main aim is to identify the plant type of an unknown substance that may contain a mixture of two or more plant parts.

Chapter 4: The Identification of *Digitalis purpurea* Using Near-Infrared Spectroscopy

4.1 Introduction

A large number of plants throughout the plant kingdom contain C_{23} or C_{24} steroidal glycosides that exert a slowing and strengthening effect on the failing heart (Evans 1989). *Digitalis purpurea*, more commonly known as the foxglove, has been used to treat cardiac patients since 1785, when the Birmingham physician and botanist William Withering introduced it for the treatment of dropsy, which is now known to be the result of a heart condition (Evans 1989e). Indeed, two of its components, the glycosides digoxin and digitoxin are the most popular treatments for rapid atrial fibrillation (Cox & Balick 1994).

Digitalis consists of the dried leaves of *Digitalis purpurea*, and is required to contain not less than 0.3% of total cardenolides calculated as digitoxin (British Pharmacopoeia 2001a). The prepared form of the drug is a standardised powder, in which the leaf may be diluted to the required strength by the addition of weaker powdered digitalis or powdered grass (Evans 1989e). Present methods for analysing *Digitalis*, which include chromatographic techniques and microscopy (Evans 1989e, British Pharmacopoeia 1999), are time consuming and, in the former case, destructive. In the latter case, the procedure involves meticulous observations including the presence and appearance of stomata and hairs on both leaf surfaces, the shapes of the various cells, and the absence of calcium oxalate crystals. (Evans 1989e) As a result, NIRS may have its advantages as a method for its analysis, as it is rapid, non-destructive, and can provide simultaneous information about the chemical composition and physical state, including moisture content and particle size data (Moffat et al 1997).

This investigation aims to use NIRS for the characterisation of *Digitalis purpurea,* the most widely used species of *Digitalis*, using relatively simple procedures. It is hoped that NIRS in this case can be used to achieve rapid identification and to confirm purity of this medicinally important natural product. *Digitalis purpurea* was chosen as the subject of interest in this study, as it can often be confused with other species of *Digitalis*, such as *ambigua*, *mertonensis*, *orientalis* and *lanata*, which may appear morphologically similar, but can differ in the amount and presence of glycosides. In addition, adulteration by non-digitalis plants has also been recorded. Examples include comfrey (*Symphytum officinale*), primrose (*Primula vulgaris*), elecampane (*Inula helenium*), ploughman's spikenard (*Inula conyza*) and nettle (*Urtica dioica*) (Evans 1989e). Adulteration by these can cause unwanted toxic side effects. For example, the pyrrolizidine alkaloid constituents of comfrey are known to have hepatotoxic properties (Newall et al, 1996).

4.2. Aims

In this investigation, the use of NIRS as a technique for the rapid identification of *Digitalis purpurea*wasexamined. If successful, this would be advantageous, as traditional methods are destructive and time-consuming.

4.3 Materials

4.3.1 Digitalis samples

Twelve samples each of 5 species of *Digitalis* leaf (*Digitalis purpurea, Digitalis lanata, Digitalis mertonensis, Digitalis ambigua, Digitalis orientalis*), all in the

dried state, were obtained from the pharmacognosy archives at the School of Pharmacy, University of London, as were stem samples of *Digitalis purpurea.* Leaf samples of some other pharmaceutically important natural products (twelve of each) were also obtained from the archives. These included belladonna, hyoscyamus, stramonium, buchu, eyebright, senna, peppermint, hop and coltsfoot.

4.3.2 Data analysis

The data analysis methods used are described and explained in Chapter 2.

4.4 Results and discussion

4.4.1 Spectral characteristics

Figure 4.1 shows sample spectra from the five species of *Digitalis*. Figure 4.2 shows the same set of spectra after they have been SNV-corrected, second derivative-transformed. A preliminary investigation on the various data pretreatments showed that the SNV-2nd derivative combination provided most successful results, as well as sharper, more well-defined peaks and the removal of scatter effects and particle size dependency.

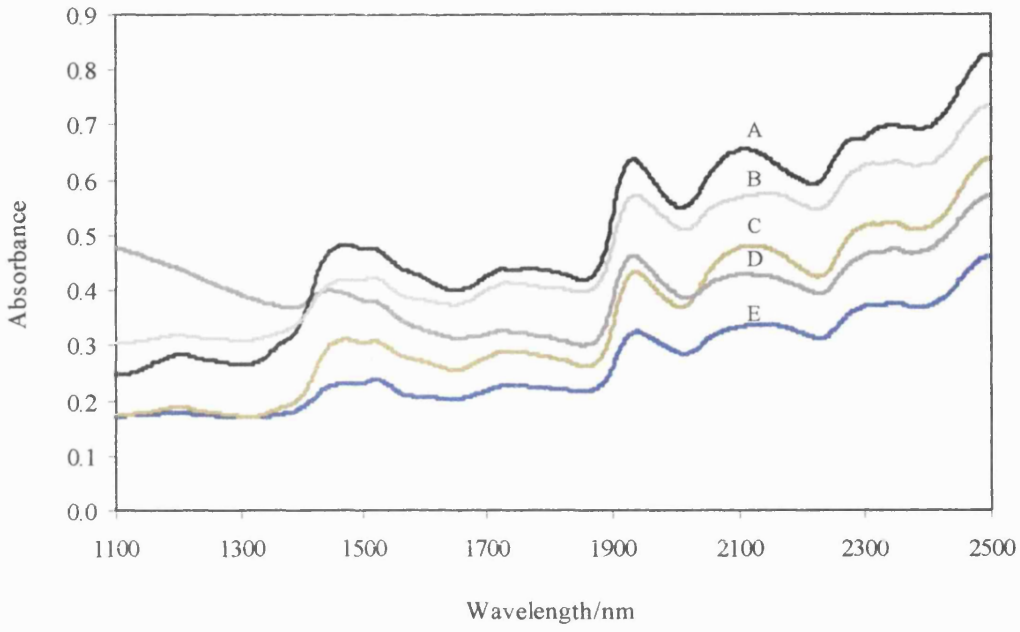


Figure 4.1. Typical spectra of 5 *Digitalis* species. A= *purpurea*, B= *lanata*, C= *mertonensis*, D= *orientalis*, E= *ambigua*.

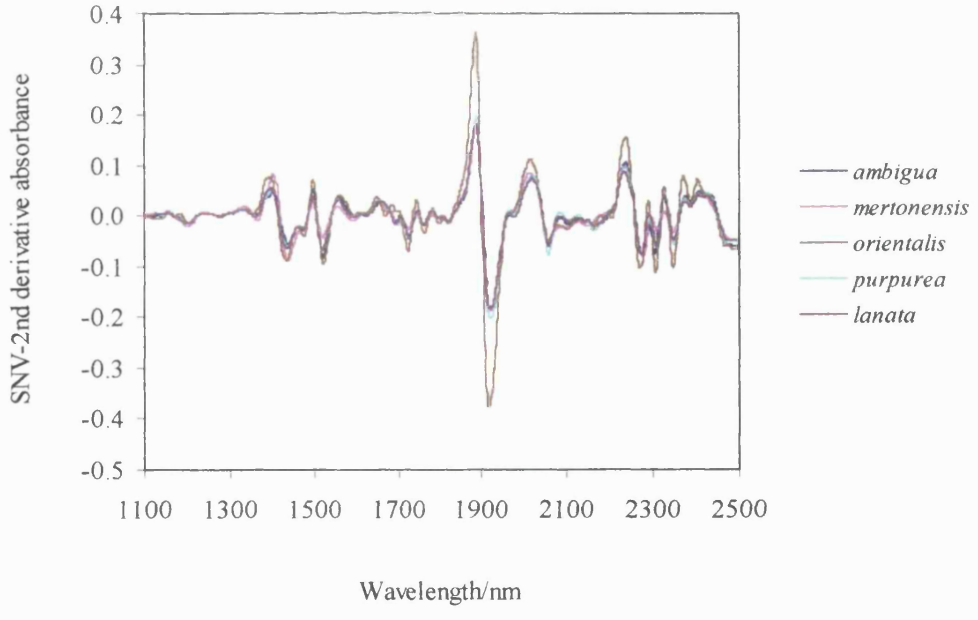


Figure 4.2. SNV-corrected, 2nd derivative transformed spectra of 5 *Digitalis* species.

4.4.2 Identification - Maximum Distance in Wavelength Space

4.4.2.1 Different species

By using the Maximum Distance in Wavelength Space method on SNV-corrected, second derivative-transformed spectra in FOSS Vision software, it was possible to completely categorise *Digitalis* as being distinct from the other leaf samples investigated. Table 4.1 shows resulting match values between the 5 *Digitalis* species and other leaf samples of pharmaceutical interest. Values ranged from 12.1 to 59.5, which is well above a reasonable value expected if classification to a sample had occurred. Between the five species of *Digitalis*, the lowest value observed for matched samples was 1.32 for *Digitalis orientalis* and the highest was 2.37 for *Digitalis ambigua* (Table 4.2.). Figure 4.3 shows the range of match values observed in both matched and mismatched samples. It is clear from this that there is a substantial gap of 5.03 between even the highest correctly matched value and the lowest mismatched number, and so 4.0 appears to be a generous threshold limit. Based on this value, it can be said that the method was successfully able to achieve complete identification between the five *Digitalis* species as well as between *Digitalis* and other genuses. It was also possible to say that based on the results among the different species of *Digitalis*, *Digitalis purpurea* appears to be most similar to *Digitalis lanata*, then to *Digitalis ambigua*, then to *Digitalis orientalis*, and finally to *Digitalis mertonensis*. One reason for its similarity to *Digitalis lanata* could be the fact that their primary glycosides closely resemble each other (Evans 1989e).

To verify reproducibility, the same procedure was carried out on two different batches of the same samples. For the five species of *Digitalis*, results ranged from 1.25 to 2.26 for the first batch and 1.36 to 2.57 for the second batch for correctly matched samples. For mismatched *Digitalis* samples, match values ranged from 7.22 (*purpurea* against *lanata*) to 75.3 (*orientalis* against *ambigua*) for the first batch and 7.67 (*purpurea* against *lanata*) to 80.2 (*orientalis* against *ambigua*) for the second. When comparing *Digitalis* with other leaves, results were also consistent with the initial findings. Values ranged from 11.5 to 57.9 for the first batch and 13.2 to 62.1 for the second.

Table 4.1. Maximum Distance in Wavelength Space match values between leaves of 5 *Digitalis* species and some other leaf samples for SNV-transformed, 2nd derivative spectra.

Leaf sample	*D. purpurea*	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
Belladonna	34.0	15.7	41.3	19.0	15.5
Hyoscyamus	46.5	19.0	52.5	21.9	30.3
Stramonium	41.9	17.9	46.0	26.7	27.5
Buchu	56.5	48.0	27.1	25.9	59.5
Eyebright	27.0	26.6	31.2	32.3	31.2
Senna	32.7	27.3	33.6	17.7	17.6
Peppermint	39.2	16.6	52.5	21.2	19.7
Hop	24.3	12.1	35.7	16.8	12.2
Coltsfoot	55.8	35.2	47.0	40.0	35.3

Numbers are average values

Table 4.2. Maximum Distance in Wavelength Space match values between leaf samples of 5 *Digitalis* species for SNV-corrected, 2nd derivative spectra.

Leaf sample	*D. purpurea*	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
D. purpurea	1.65-1.90-2.26				
D. lanata	7.41-9.04-10.97	1.57-1.96-2.20			
D. mertonensis	27.92-31.36-34.22	12.37-14.41-16.05	1.52-1.91-2.25		
D. orientalis	14.75-15.23-16.63	14.24-16.71-18.25	16.74-19.68-21.19	1.32-1.91-2.31	
D. ambigua	10.03-11.28-12.74	10.64-19.02-21.73	14.11-18.46-20.95	60.71-72.40-78.43	1.62-2.03-2.37

Left hand number = smallest value, underlined number = median value, and right hand number = largest value

Figure 4.3 Match value ranges for correct identification (pass) and incorrect identification (fail) for five species of *Digitalis*

4.4.2.2 Plant parts for Digitalis purpurea

Table 4.3 shows match values between leaf and stem samples of *Digitalis purpurea*. Values between the same plant parts ranged from 1.52 to 2.26, and those between different plant parts were from 29.04 to 67.41. As a result, it can be said that the method was successfully able to distinguish between these different plant parts of the same species. Figure 4.4 shows the sensitivity of the technique for the discrimination of plant parts. Here, identification using Maximum Distance in Wavelength Space was attempted on *Digitalis purpurea* leaf samples that had been successively diluted with *Digitalis purpurea* stem. This was carried out on 100% leaf, 75 % leaf, 50% leaf, 25% leaf and 0% leaf (100% stem). From Figure 4.4, it is clear that a sample with even small amounts of stem (2%) can be detected and identified as distinct from a leaf-only sample.

Table 4.3. Match values between SNV-corrected, 2nd derivative spectra of *Digitalis purpurea* leaves and stems using Maximum Distance in Wavelength in FOSS Vision software.

Plant part	leaves	stems
leaves	1.65-1.90-2.26	48.82-56.28-67.41
stems	29.04-34.29-40.80	1.52-1.77-2.16

Left hand number = smallest value, underlined number = median value, right hand number = largest value.

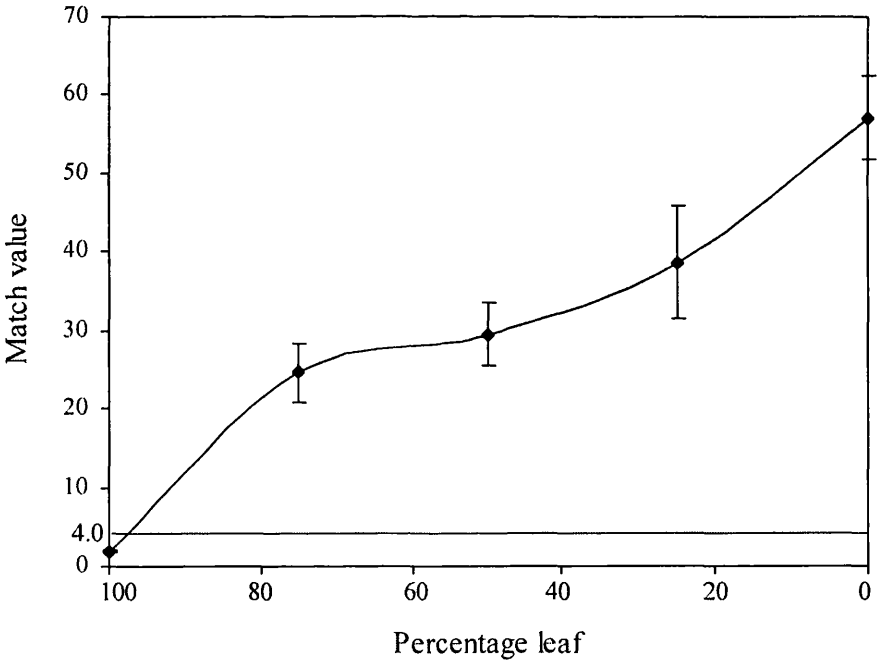


Figure 4.4. Maximum Distance in Wavelength Space match values obtained from attempted identification (as leaf) of successive dilutions of *D. purpurea* leaf (100% leaf → 100% stem). Error bars indicate ± 1 standard deviation.

4.4.3 Correlation in Wavelength Space

Table 4.4 shows Correlation in Wavelength Space values between the five species of *Digitalis* leaf. Values ranged from virtually complete correlation (0.999) to 0.953. Relatively high values were observed when the same procedure was applied between the five *Digitalis* species and other natural products, with the lowest observed value being 0.890 for *Digitalis mertonensis* against Stramonium and the highest being 0.991 for *Digitalis ambigua* against hop leaves (Table 4.5). As resulting values were so high, setting a threshold limit appears rather problematic, with 0.995 being probably the lowest reasonable value. However, when using that threshold value, it can be said that the Correlation in Wavelength Space method can successfully identify *Digitalis* as being distinct from samples from other genuses, and also the five species as being unlike from each other. It can also be seen from Table 4.4 that *Digitalis purpurea* appears to be most like *Digitalis lanata* and least like *Digitalis mertonensis* among the 5 species, with *Digitalis ambigua* and *Digitalis orientalis* in between the two. It was also possible to utilise the method for the correct identification of *Digitalis purpurea* leaves and stems (Table 4.6).

To verify reproducibility, the same procedure was carried out on two more batches of the same samples. For the five species of *Digitalis*, results ranged from 0.999 to 0.950 for the first batch and 0.999 to 0.952 for the second batch. When comparing *Digitalis* with other leaves, results were also consistent with the initial findings. For both batches, the highest results were observed for *Digitalis ambigua* and hop

leaves (0.98 and 0.98) and the lowest were observed for *Digitalis mertonensis* and Stramonium (0.882 and 0.887).

Table 4.4. Correlation values between 5 species of *Digitalis* leaves using the Correlation in Wavelength Space method in FOSS Vision software on SNV-corrected, 2nd derivative spectra.

Species	*D. purpurea*	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
D. purpurea	0.9995				
D. lanata	0.9940	0.9977			
D. mertonensis	0.9644	0.9648	0.9995		
D. orientalis	0.9682	0.9720	0.9356	0.9995	
D. ambigua	0.9873	0.9812	0.9624	0.9524	0.9990

Table 4.5. Correlation values between 5 species of *Digitalis* leaves and other leaf samples using the Correlation in Wavelength Space method in FOSS Vision software, on SNV-corrected, 2nd derivative spectra.

Species	*D. purpurea*	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
Stramonium	0.964	0.955	0.890	0.922	0.952
Coltsfoot	0.951	0.960	0.931	0.952	0.948
Eyebright	0.964	0.973	0.954	0.977	0.943
Hyoscyamus	0.978	0.973	0.914	0.956	0.971
Buchu	0.936	0.935	0.914	0.932	0.940
Belladonna	0.982	0.976	0.920	0.939	0.978
Senna	0.984	0.985	0.953	0.965	0.970
Hop	0.987	0.979	0.945	0.940	0.991
Peppermint	0.990	0.986	0.943	0.964	0.982

Table 4.6. Correlation values between *Digitalis purpurea* leaves and stems using the Correlation in Wavelength Space method in FOSS Vision software on SNV 2nd derivative spectra.

Plant part	Leaves	Stems
Leaves	0.9995	0.9018
Stems	0.9009	0.9985

4.4.4 Correlation coefficients

When looking at *r* values between *Digitalis* and other natural products (Table 4.7) it is clear that values are rather high, ranging from 0.891 to 0.990. Results between the 5 species are also quite variable, with the lowest being 0.936 and the highest being 0.996 (Table 4.8) for mismatched samples. The *r* value between *Digitalis purpurea* leaves and stems was a fairly low 0.903. Again, however, it is clear that in order of similarity to *Digitalis purpurea* among the five species, the same order applies to that found in the previous two investigations. It was also possible to use *r* values to correctly identify an 'unknown' sample of *Digitalis purpurea.* When *r* values were calculated between the 'unknown' and the 5 species, the highest value was observed between the 'unknown' and *Digitalis purpurea* in the library (Table 4.9).

When comparing this method to Correlation in Wavelength Space, it can be seen that results are extremely similar, both being characterised by high values that make the setting of thresholds difficult.

Table 4.7. Table of correlation coefficients, *r*, between SNV-corrected, 2nd derivative spectra of leaf samples of 5 *Digitalis* species and other leaf samples.

Species	*D. purpurea*	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
Belladonna	0.983	0.979	0.922	0.940	0.980
Hyoscyamus	0.979	0.975	0.916	0.957	0.972
Senna	0.984	0.988	0.954	0.971	0.965
Stramonium	0.964	0.957	0.891	0.923	0.951
Hop	0.987	0.980	0.946	0.939	0.992
Peppermint	0.990	0.989	0.945	0.965	0.983
Coltsfoot	0.952	0.962	0.933	0.952	0.949
Eyebright	0.965	0.976	0.955	0.977	0.945
Buchu	0.937	0.938	0.915	0.941	0.933

Table 4.8. Table of correlation coefficients, *r*, between SNV-corrected, second derivative transformed spectra of leaf samples of 5 *Digitalis* species.

Species	*D. purpurea*	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
D. purpurea	1.0				
D. lanata	0.996	1.0			
D. mertonensis	0.965	0.968	1.0		
D. orientalis	0.969	0.973	0.936	1.0	
D. ambigua	0.987	0.983	0.963	0.952	1.0

Table 4.9. Table of *r* values between SNV-2nd derivative spectra of *5 Digitalis* species and an 'unknown' sample of *D. purpurea.*

	D. purpurea	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
'unknown'	0.9985	0.9953	0.9685	0.9735	0.9889

4.4.5 Two wavelength plot

The NIR absorption spectrum of water includes five bands with maxima at 1940, 1450, 1190, 970, and 760 at room temperature (Osborne et al 1993). As these peaks could be used to quantify water content of materials, they were avoided as various wavelength combinations and different data pretreatments were tested on leaves of the five *Digitalis* species. Pre-treatments attempted were: 1^{st} to 4^{th} derivative, SNV-1^{st} to 4^{th} derivative, de-trend, SNV-de-trend, de-trend-SNV. The final combination used was the two wavelengths 1150 and 2160nm, on baseline-corrected (de-trend), SNV-transformed spectra. These wavelengths were chosen empirically, as the resulting plot (Figure 4.5) showed the best visual separation of the five species of *Digitalis* compared to other combinations attempted. Using this method on two new batches of the same samples yielded results (plots) that were highly comparable to the first batch both times.

4.4.6.1 Identification using 'nearest neighbours'

Samples of *Digitalis* leaf that were not used previously were treated as "unknowns" and their absorbances were plotted in the resulting two-wavelength plot. When the "nearest neighbours" method of analysis was then carried out, it was clear that each "unknown" species of *Digitalis* leaf could be identified correctly. That is, it was possible to correctly assign all the species that were treated as unknowns to the groups that gave rise to the smallest Euclidean distances (i.e. nearest neighbours). It is clear from Table 4.10 that the closest neighbours for each unknown group came from the species that it really belonged to. For example the average Euclidean distance between the *Digitalis purpurea*

"unknown" and the *Digitalis purpurea* reference datapoints was 0.071 compared to 0.213, 0.443 and 0.686 for the same unknown against *Digitalis lanata*, *Digitalis mertonensis* and *Digitalis orientalis*, respectively. Figure 4.6 shows the 4 "unknown" *Digitalis* species incorporated into the plot represented in Figure 4.5.

In terms of reproducibility, results were consistent when this method was attempted on two new batches of the same samples. Euclidean distances for correctly matched samples were 0.082, 0.122, 0.187 and 0.057 for the first batch and 0.068, 0.115, 0.162 and 0.050 for the second batch, for *Digitalis purpurea*, *Digitalis lanata*, *Digitalis mertonensis* and *Digitalis orientalis*, respectively. For mismatched samples, distances ranged from 0.210 (*Digitalis purpurea* unknown against *Digitalis lanata*) to 1.181 (*Digitalis mertonensis* unknown against *Digitalis orientalis*) for the first batch and 0.225 (*Digitalis purpurea* unknown against *Digitalis lanata*) to 1.132 (*Digitalis mertonensis* unknown against *Digitalis orientalis*) for the second. A possible threshold for this method therefore would be 0.19. That is, for *Digitalis*, a sample can be identified as a species if its distance from the reference datapoints is no more than 0.19.

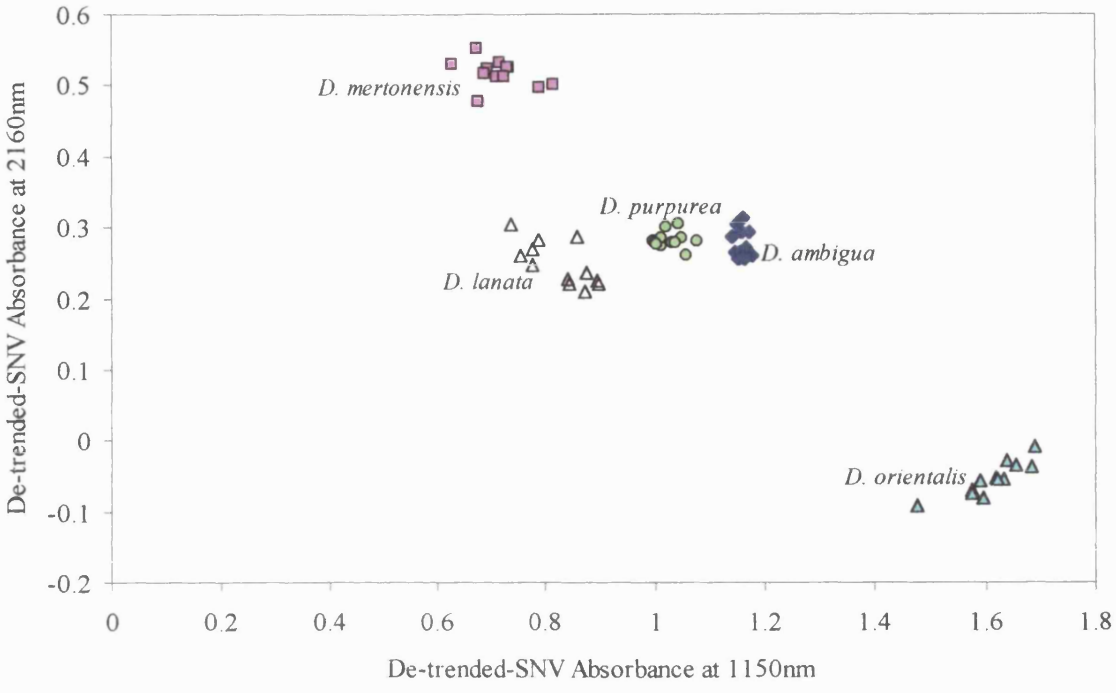


Figure 4.5. A two-wavelength plot for leaves of 5 *Digitalis* species.

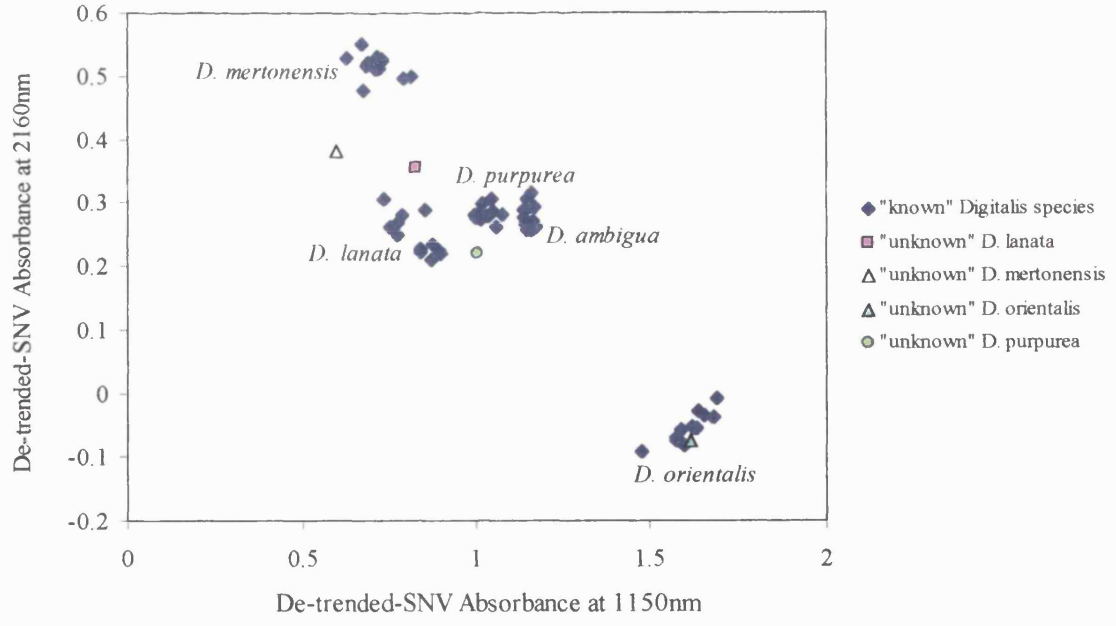


Figure 4.6. A two-wavelength plot for leaves of 5 *Digitalis* species, including 'unknown' samples.

Table 4.10. Average Euclidean distances between 5 *Digitalis* species and 4 'unknown' samples between points on a 2-wavelength plot for de-trended SNV spectra, wavelengths 1150-2160nm.

	D. purpurea	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
D. purpurea	0.071	0.182	0.416	0.667	0.163
(unknown)					
D. lanata	0.213	0.119	0.203	0.884	0.336
(unknown)					
D. mertonensis	0.443	0.265	0.184	1.104	0.567
(unknown)					
D. orientalis	0.686	0.853	1.077	0.052	0.580
(unknown)					

Smallest values are underlined

4.5 Polar Qualification System

The suitability for the PQS method for the discrimination of the five species of *Digitalis* was tested using various wavelength increments. Figure 4.7 shows the best-separated plot that was achieved, using the range 1290 – 1390 nm. It can be seen from this that PQS is successful in separating the samples according to species. However, unlike the two-wavelength method, separation is not as distinct. That is, the datapoints of *Digitalis lanata* and *Digitalis orientalis* appear to drift slightly in a horizontal manner, rather than being formed into a tight cluster.

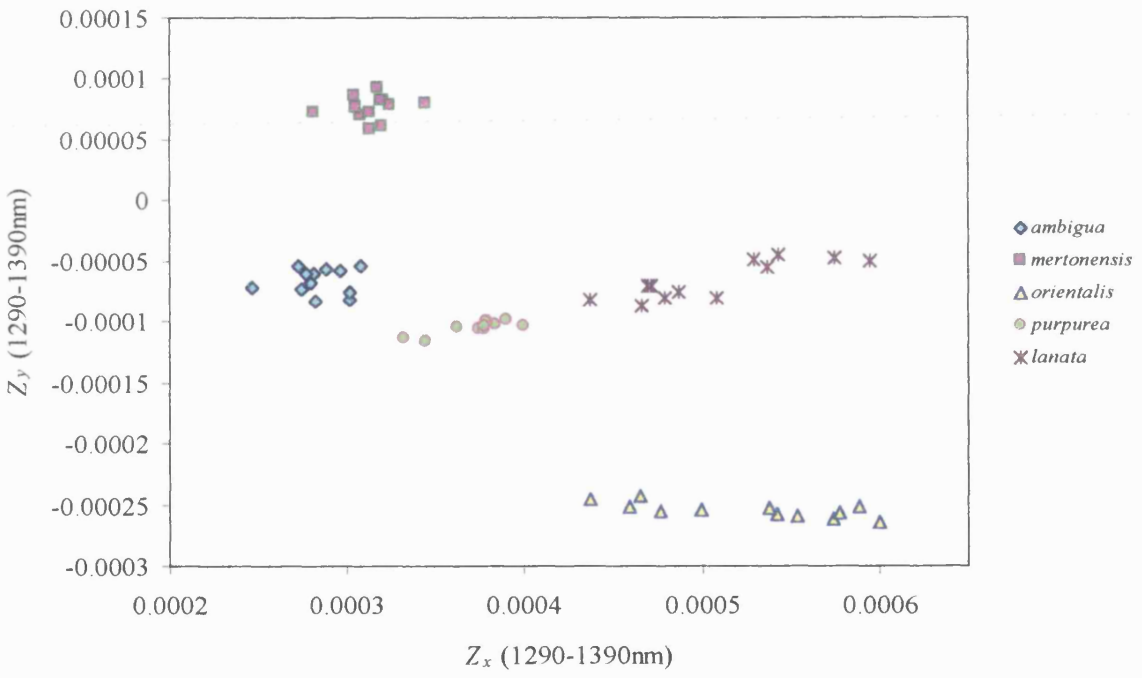


Figure 4.7. A PQS plot for five species of *Digitalis* using the wavelength range 1290-1390 nm

4.6 Conclusion

In the first instance, it was possible to completely identify *Digitalis purpurea* as being distinct from other samples, using a Maximum Distance in Wavelength Space statistical comparison method on Standard Normal Variate corrected, second derivative spectra. Match values ranged from 1.65 to 2.26 for correct identification. It was also possible to discriminate between different plant parts of *Digitalis purpurea,* with match values ranging from 1.52 to 2.26. Although less conclusive, it was also possible to use the Correlation in Wavelength Space method on Standard Normal Variate corrected, second derivative spectra to identify *Digitalis purpurea*, with a resulting value of 0.9995. Again, it was also possible to distinguish between different plant parts. The use of correlation coefficients (*r* values) for the correct identification of an unknown sample of *D. purpurea* was also possible, with a resulting *r* value of 0.9953. The two-wavelength, nearest neighbours analysis carried out for baseline-treated (de-trended), standard normal variate corrected spectra at 1150 and 2160nm resulted in the successful identification of unknown samples.

Table 4.11 shows a summary of various results for *Digitalis purpurea.* Based on these results, it appears that Maximum Distance in Wavelength Space is the most successful method for the identification of *Digitalis* from other plants as well as identifying a single *Digitalis* species, with very distinctly low or high match values being observed depending on matched or mismatched samples. It was also sensitive enough to successfully distinguish between different plant parts. The use of correlation values is also successful to some extent, but setting thresholds is

troublesome, as most of the resulting values are quite high. Therefore, more work is needed in order to make the methods more robust. Although used commonly in the pharmaceutical industry, PCA was not used in this instance, as preliminary results proved inconclusive. The use of two wavelength plots is useful in pattern recognition and giving a good visual idea of the differences between the species. However, this method is still in its preliminary stages, and it is necessary to devise a method to select the most ideal wavelength combinations and data pre-treatments to use, as there may exist other combinations that are potentially more successful than the ones used here.

Like the two-wavelength method, the PQS method provided a good visual separation of the five species as being distinct from each other. However, separation in this instance was not as well defined as that achieved by the two-wavelength method. This may be due to the fact that PQS uses a range of wavelengths, rather than the two-wavelength method, which uses two distinct wavelengths where separation is best observed. A possible extension to this study (which was beyond the time frame of this project) would be to examine the constituents of the various plant leaves and different parts of the plant, and then target specific wavelengths to get the most discrimination.

Therefore, it can be said that NIRS does possess the potential for the rapid characterisation and identification of natural products, in this case, *Digitalis purpurea.* This is of significant importance, as NIRS is faster than the commonly used traditional techniques. While current methods may take hours, analysis by NIRS can be performed in a matter of minutes. However, this study is still a

preliminary one, and the development of new chemometric techniques in the future is aimed, as well as the improvement of any techniques mentioned above. It is also hoped that the effect of various parameters, such as moisture content and particle size, can be investigated.

Table 4.11. Table of various values between *Digitalis purpurea* leaves and other leaf samples including those of other *Digitalis* species.

Species	A. Match value	B. Correlation	C. Correlation coefficient *r*	D. Euclidean Distance
D. purpurea	1.90	1.0	1.0	0.024
D. lanata	9.04	0.994	0.996	0.205
D. mertonensis	31.4	0.964	0.965	0.392
D. orientalis	15.2	0.968	0.969	0.673
D. ambigua	11.3	0.987	0.987	0.127
Belladonna	34.0	0.982	0.983	0.524
Hyoscyamus	46.5	0.978	0.979	0.177
Stramonium	41.9	0.964	0.964	0.277
Buchu	56.5	0.936	0.937	0.352
Eyebright	27.0	0.964	0.965	0.398
Senna	32.7	0.984	0.984	0.149
Peppermint	39.2	0.990	0.990	0.145
Hop	24.3	0.987	0.987	0.222
Coltsfoot	55.8	0.951	0.952	0.323

Column A = average match values for SNV-2nd derivative spectra using Maximum Distance in Wavelength Space in FOSS Vision software, Column B = average correlation values using Correlation in Wavelength Space in FOSS Vision software on SNV-2nd derivative spectra, Column C = average correlation coefficient, *r*, for SNV-2nd derivative spectra, and Column D = average Euclidean distances between points on a 2-wavelength plot for de-trended SNV spectra, wavelengths 1150 and 2160nm.

Chapter 5: The Determination of the Geographical Origins of *Cannabis sativa* and other Natural Products by Near-Infrared Spectroscopy

5.1 Introduction

There are several factors that determine the commercial geographical sources of a natural product; two of them are the suitability of the plant to a particular environment and the economic factors associated with the production of a drug in a particular area (Evans, 1989f). Many plants are capable of growing well in numerous localities that have similar climates, but it must always be remembered that although a plant may thrive in different conditions, it may fail to produce the same constituents (Evans, 1989f). That is, even though herbal medicines come from the same species, the quality and efficacy are somewhat different according to growing conditions based on geographical origin (Woo et al, 1999). For instance, cinchonas growing at high altitude are considerably different to those growing in the plains, and pharmacopoeial ginger, which once came exclusively from Jamaica, has now been replaced by that grown in Africa and China due to the improved quality of these two areas as sources (Evans, 1989f).

Because of the differences that may exist among natural products grown in different geographical origins, a rapid and accurate analytical method to determine the origin would be useful for the correct value estimation of a drug in question, and for the prevention of illegal distribution (Woo et al, 1999). In addition, a rapid technique would be useful in forensic science for identifying the contents and geographical origin of, for example, suspect packages. However, it is not easy to identify the geographical origin of natural products using current existing analytical techniques or through visual inspection. Because there are tens of major components which differ slightly depending on growing conditions such as

geographical origin, it is impossible to select only a few specific components as essential criteria (Woo et al, 1999). Although NIRS has been used successfully to discriminate between natural products of different species as with *Digitalis*, studies that apply it to the classification of samples of the same species but of different geographical origin are relatively rare.

One previously recorded study involved the differentiation of Korean and Chinese samples of *Astragali Radix* (a root of *Astragalus membranaceus*, clinically used in Korea and China to improve a reduced immune response), *Ganoderma* (an oriental fungus, *Ganoderma lucidum*), and *Smilacis Rhizoma* (a rhizome of *Smilax glabra*, used in the Orient for chronic skin disease and syphilis) (Woo et al 1999). Results showed that NIRS does have the potential of discriminating these herbal medicines according to their geographic origin, although in this case, the origins were confined to just two countries (Woo et al 1999).

The determination of geographical origins is also of importance in the tobacco industry. Tobacco types include Virginia, Burley, Maryland, and fire-cured tobaccos. Under certain conditions, tobacco types may visually appear to be the same, when, in fact, they can be chemically different. One problem in the industry is the possible substitution of one type for another at the point of sale (Hana et al, 1997). One study involved the analysis of native Burley tobacco (USA), and non-native Burley tobacco (Korea, Thailand, Mexico, Ecuador, Argentina, Brazil, Chile, Costa Rica, Italy and Malawi). Results showed some potential, although there was a relatively high percentage of misclassified non-native tobaccos possibly due to variations in the spectra caused by soil, weather, irrigation and

fertilisation (Hana et al, 1997). Although promising, no attempt was made in that study to actually identify the country of origin of a sample - it merely sought to distinguish US samples from non US ones.

The determination of local origins can also be of importance in the food industry. For instance, Basmati, a rice grown in the Punjab region of India and Pakistan, accounts for about 20 percent of all rice sales in the UK. Basmati can only be grown once a year with a yield half of that of other rices. As a consequence of its scarcity and its popularity in the UK, Basmati is more expensive compared to other rice varieties. As the eating quality of this particular rice cannot be duplicated by growing the same seed in other regions, the ability to discriminate authentic Basmati from cheaper, inferior ones would be advantageous to the trader (Osborne et al, 1993).

In this investigation, NIRS was used to determine the geographical origins of samples of *Cannabis sativa* flowering heads (Figure 5.1). Cannabis was chosen here, as it is grown in a variety of different countries, and, as a controlled drug, its determination of origin would have massive potential in forensic science. In addition, although it has been used by man for at least 3500 years (Notcutt et al 1997), there has been increased interest in the plant recently due to its possible effectiveness in treating pain, tremor, and muscle spasticity, especially those associated with multiple sclerosis (Pertwee, 1997). Furthermore, the use of cannabis as a possible application in treating epilepsy, glaucoma, and bronchial asthma have been suggested (House of Lords Select Committee on Science and Technology, November 1998). The principal chemical components of cannabis are

the cannabinoids: tetrahydrocannabinol (THC) (which is the most abundant and responsible for its intoxicating properties), cannabidiol (CBD), and cannabinol (CBN) (Holdcroft et al 1997).

This study aims to use NIRS to determine the geographical origins of cannabis flowering heads from four distinct localities, namely India, South Africa, Thailand, and Turkey. The same concept was also applied to cannabis resins from the three countries of origin of India, Pakistan, and Lebanon. In addition, to further test the technique, samples of Belladonna root, Valerian root, and aloe resins were also examined. There were chosen as they were the most readily available natural products on hand, which were grown in a variety of different countries. If successful, NIRS would be beneficial over traditional methods due to its speed, non-destructiveness, and reliability.

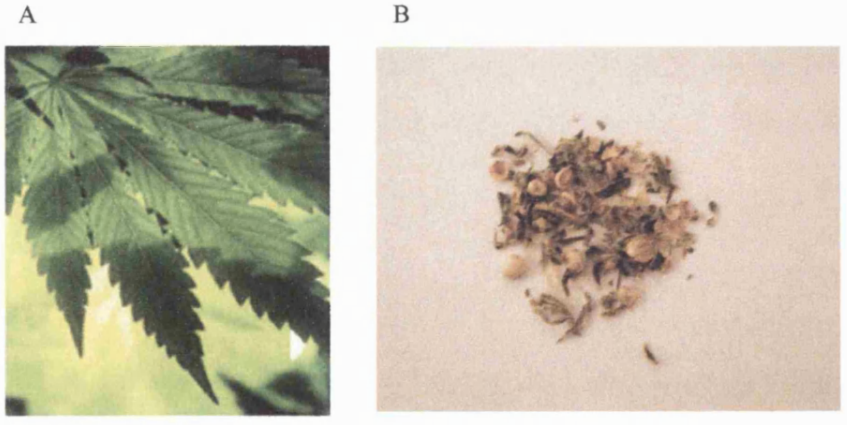


Figure 5.1. A *Cannabis sativa* leaf (A) and flowering heads (B)

5.2 Materials and methods

5.2.1 Instrumentation

See chapter 2

5.2.2 Cannabis samples

Twelve samples each of *Cannabis sativa* flowering heads from India, South Africa, Thailand, and Turkey were obtained from the phamacognosy archives at The School of Pharmacy, University of London. In addition, ten cannabis resin samples from Lebanon, eight from India, and eight from Pakistan, were also obtained from the same archives.

Other natural products were also examined. These included twelve samples each of English and Indian Valerian and Belladonna root, and twelve samples each of Cape, Socotrine, and Curaçao aloe resins. Again, these were obtained from the pharmacognosy archives at The School of Pharmacy.

All samples were powdered to an approximate particle size of 500μm using a mortar and pestle and placed in 10mm diameter glass vials (Waters).

5.2.3 Data analysis

In FOSS Vision® software, three different identification procedures were used on all the samples. These were Maximum Distance in Wavelength Space, Correlation in Wavelength Space, and Residual Variance in Principal Components Space. All used the wavelength range 1100-2500nm and a second derivative gap size of 10nm. Correlation coefficients, two-wavelength plot, *k*NN and PQS were also used.

5.3 Results and discussion

5.3.1 Spectral characteristics – Cannabis

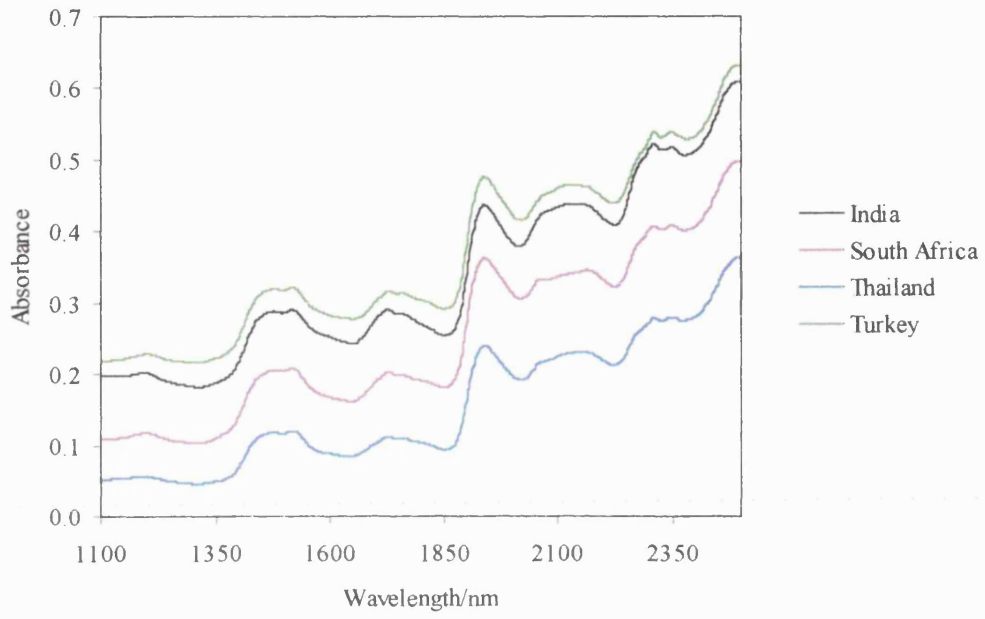
Figure 5.2 shows sample spectra of Cannabis flowering heads and resins from various geographical origins. Figure 5.3 shows the same set of spectra after they have been SNV-corrected, 2nd derivative transformed. It can be seen from these that the sample spectra for each category are extremely similar in shape, with their main differences being in their in baselines. Using SNV-corrected 2nd derivative spectra removes these baseline effects and produces sharper, more well-defined peaks. However, the spectra are still difficult to distinguish from each other by mere visual examination. It is also worth noting here that while the untreated spectra for flowering heads and resin are reasonably similar to each other (Figure 5.2), the SNV-2nd derivative spectra show some major differences between flowering heads and resin (Figure 5.3). For example the spectra for the resins show a peak in the vicinity of 1650nm which is not seen in the spectra for flowering heads.

5.3.2 Identification

For the Maximum Distance in Wavelength identification method, the Correlation in Wavelength Space Method, the Residual Variance in Principal Components Space Method, and the correlation coefficient method, SNV-corrected, second derivative spectra were used. An initial look at the various data pretreatments showed that this combination provided most successful results, as well as sharper, more well-defined peaks and the removal of scatter effects and particle size dependency. Using other data pretreatments (first derivative, SNV-first derivative,

second derivative, third derivative, SNV-third derivative fourth derivative, SNV-fourth derivative, de-trend, SNV-de-trend) all produced at least three failures in the data analysis process.

a



b

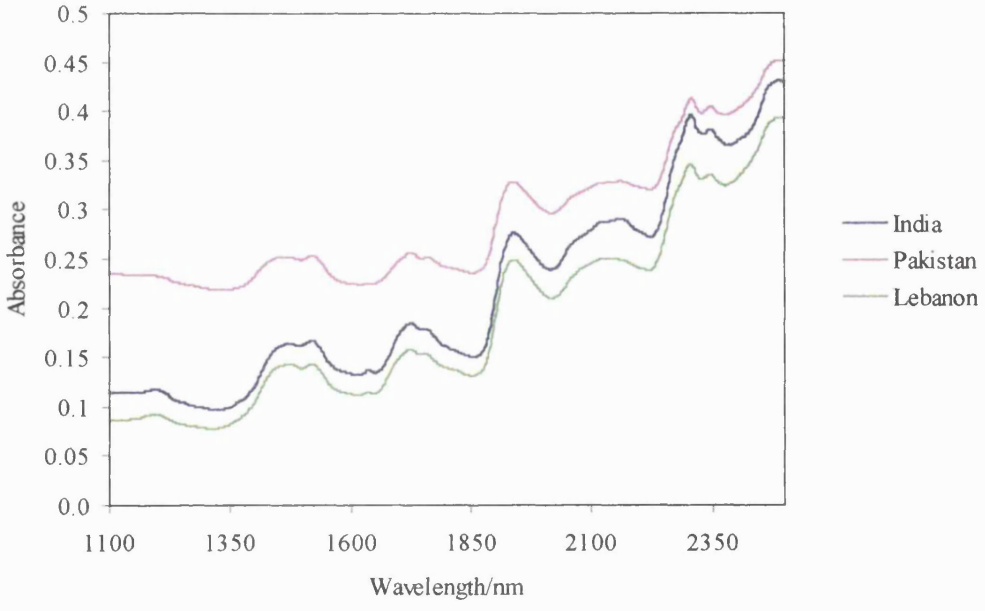


Figure 5.2 Sample spectra of **(a)** cannabis flowering heads and **(b)** cannabis resin from various geographical origins.

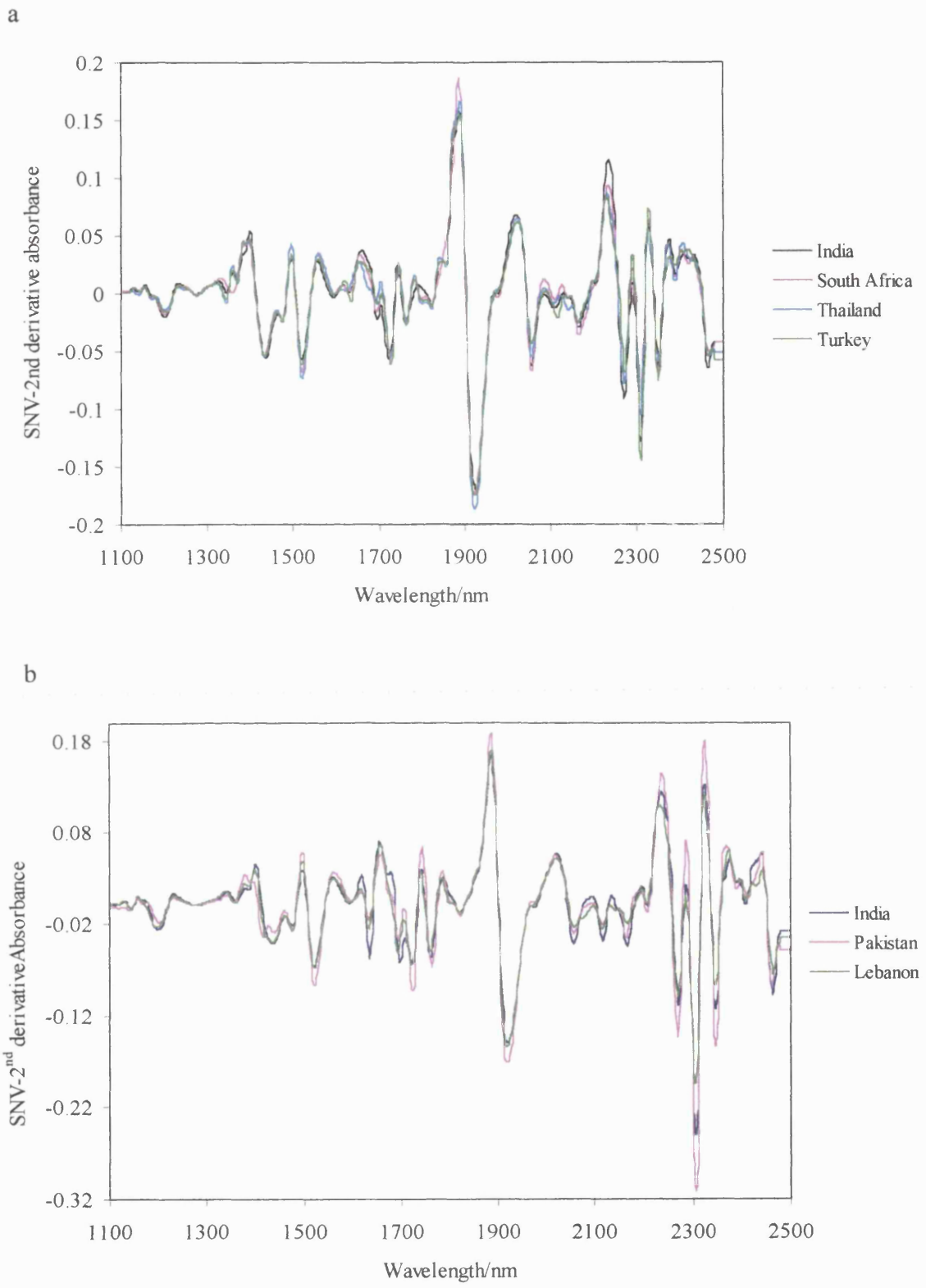


Figure 5.3. SNV-corrected, 2nd derivative spectra of **(a)** cannabis flowering heads and **(b)** cannabis resin from various geographical origins.

5.3.3 Maximum Distance in Wavelength Space

5.3.3.1 Cannabis flowering heads

It was possible to correctly allocate each sample of cannabis flowering heads to its place of origin using the Maximum Distance in Wavelength Space method on SNV-corrected, second derivative-transformed spectra in FOSS Vision software. Table 5.1 shows that the resulting match values between samples of the same geographical origin are well below the Vision® threshold limit of 4.0. Values ranged from 0.23 to 2.42 for correctly matched samples and 8.21 to 46.1 for unmatched samples. The latter two values are well above the value of 4.0 (the threshold limit in the software) which would be expected if correct classification had occurred. The generous gap of 5.79 between correctly and incorrectly matched samples thus suggests that 4.0 is a reasonable threshold value.

5.4.3.2 Cannabis resins

Results were similar in the case of cannabis resins from three countries of origin. Table 5.2 shows the match values obtained when using Maximum Distance in Wavelength Space on samples of resin from India, Pakistan, and Lebanon. Here, values ranged from 1.40 to 2.47 between samples from the same geographical origin, and 46.7 to 67.5 for samples from different countries. In this case, the very large gap of 44.23 between the largest 'correct' match value and the smallest 'incorrect' match value suggests that the threshold is more than efficient, and that

this particular method is a powerful technique for the correct allocation of samples to the country in which they were grown.

Table 5.1. Maximum Distance in Wavelength Space match values for cannabis flowering heads from 4 different geographical origins

	India	South Africa	Thailand	Turkey
India	0.43-0.83-1.97			
South Africa	8.21-8.74-11.68	0.23-0.46-2.23		
Thailand	36.4-37.1-46.1	13.1-16.0-21.9	1.25-0.91-1.47	
Turkey	14.3-16.0-23.0	20.6-23.7-25.8	22.2-24.5-26.2	1.57-1.92-2.42

Left hand number = smallest value, underlined number = median value, and right hand number = largest value

Table 5.2. Maximum Distance in Wavelength Space match values for cannabis resin from 3 different geographical origins

	India	Pakistan	Lebanon
India	1.40-1.94-2.39		
Pakistan	46.7-52.5-57.7	1.68-1.91-2.29	
Lebanon	55.9-62.7-67.5	52.6-61.5-64.7	1.41-2.06-2.47

Left hand number = smallest value, underlined number = median value, and right hand number = largest value

5.3.4 Correlation in Wavelength Space - Cannabis flowering heads and resins

Table 5.3 shows correlation values between cannabis flowering heads from four geographical origins using Correlation in Wavelength Space. In correctly matched samples, values ranged from 0.998 to 0.999, and in incorrectly matched samples, from 0.980 to 0.988. Values were also high in the case of cannabis resins, ranging from 0.997 to 0.999 in correctly matched samples and 0.863 to 0.965 in samples from differing countries (Table 5.4). As the resulting correlation values are so high in both cases, setting a feasible threshold value appears problematic, with perhaps the lowest reasonable value being 0.995 for flowering heads and 0.990 for resins. Although the method will work to correctly identify the country of origin of samples if these thresholds are used, the fact that they have to be so high suggests that the Correlation in Wavelength Space method has only limited success compared to the Maximum Distance in Wavelength Space method.

Table 5.3. Correlation in Wavelength Space values for cannabis flowering heads

	India	South Africa	Thailand	Turkey
India	0.999			
South Africa	0.987	0.999		
Thailand	0.981	0.983	0.999	
Turkey	0.982	0.980	0.988	0.998

Values are mean results

Table 5.4. Correlation in Wavelength Space values for cannabis resin

	India	Pakistan	Lebanon
India	0.999		
Pakistan	0.863	0.999	
Lebanon	0.965	0.959	0.997

Values are mean results

5.3.5 Residual Variance in Principal Components Space - Cannabis flowering heads and resins

When using the Residual Variance in Principal Components Space identification method on cannabis flowering heads and resins, it was possible to allocate each sample to its correct country of origin. In the case of correctly identified flowering heads, resulting values ranged from 0.473 for Thai samples to 0.654 for Turkish samples. That is, the variance was within the threshold of 0.84. For mismatched samples, values ranged from 0.995 for Thailand vs. India and 1.0 for India vs. South Africa, India vs. Turkey, and South Africa vs. Turkey. As a result, it can be said that when using the default software threshold of 0.84 residual variance, this method is also a successful one for the correct identification of the geographical origin of the samples. Tables 5.5 and 5.6 show resulting values for flowering heads, and resins, respectively.

Table 5.5. Residual Variance in Principal Components Space values for cannabis flowering heads from 4 different countries of origin

	India	South Africa	Thailand	Turkey
India	0.585			
South Africa	1.0	0.610		
Thailand	0.995	0.994	0.473	
Turkey	1.0	1.0	0.999	0.654

Values are largest observed results

Table 5.6. Residual Variance in Principal Components Space values for cannabis resin from 3 different countries of origin

	India	Pakistan	Lebanon
India	0.549		
Pakistan	1.0	0.680	
Lebanon	1.0	1.0	0.547

Values are largest observed results

5.3.6 Correlation Coefficients - Cannabis flowering heads and resin

Tables 5.7 and 5.8 show the resulting *r*-values obtained when calculating correlation coefficients for cannabis flowering heads and resins. Results between different geographical origins were very high for the flowering heads, ranging from *r*-values of 0.980 for South Africa vs. Turkey to 0.988 for Thailand vs. Turkey. In the case of cannabis resins, results were still high between different geographical origins, although slightly lower compared to those of flowering heads. *r*-values ranged from 0.943 for Pakistan vs. Lebanon to 0.966 for Lebanon vs. India.

As resulting *r* values are so high, even for mismatched samples, it may be suggested in this case that although correlation coefficients may be used for correctly identifying the geographical origins of cannabis flowering heads and resin, it is not altogether the safest method to use, as the setting of thresholds would be problematic.

Table 5.7. Correlation coefficients for cannabis flowering heads from 4 geographical origins

	India	South Africa	Thailand	Turkey
India	1.0			
South Africa	0.986	1.0		
Thailand	0.981	0.984	1.0	
Turkey	0.983	0.980	0.988	1.0

Values are mean results

Table 5.8. Correlation coefficients for cannabis resin from 3 geographical origins

	India	Pakistan	Lebanon
India	1.0		
Pakistan	0.961	1.0	
Lebanon	0.966	0.943	1.0

Values are mean results

5.3.7 Two-wavelength plots

5.3.7.1 Two-wavelength plot - Cannabis flowering heads

Various data pretreatments and wavelength combinations were attempted on the spectra of cannabis flowering heads before the absorbances at the wavelengths 1624 nm and 2326 nm were used on de-trended (baseline corrected) spectra (Figure 5.4). This was because the resulting plot (Figure 5.5) showed the best visual separation of the four geographical origins compared to other combinations attempted. In addition, a visual inspection of de-trended spectra showed some separation between the four countries under investigation.

5.3.7.2 Two-wavelength plot - Cannabis resin

In the case of cannabis resin, de-trended spectra were again used (Figure 5.6). However, in this case, when the previous wavelength combination used on the flowering heads was applied, although there was some separation of the different geographical origins (Figure 5.7), it was found that some improvement was possible by using a different combination. For instance, using the combination of 2380 nm and 1120 nm results in a tighter set of plots (Figure 5.8) and thus this was used in the end.

Again, in this case, applying the same technique to different batches of the same samples produced highly comparable results with similar degrees of separation.

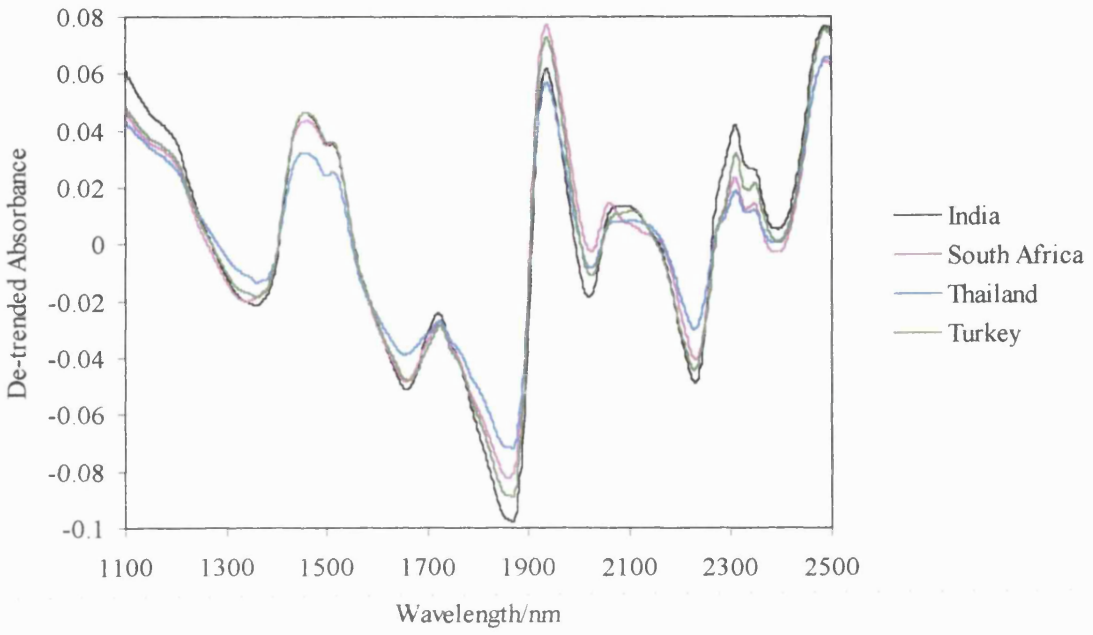


Figure 5.4. Baseline corrected (de-trended) spectra of cannabis flowering heads from 4 geographical locations

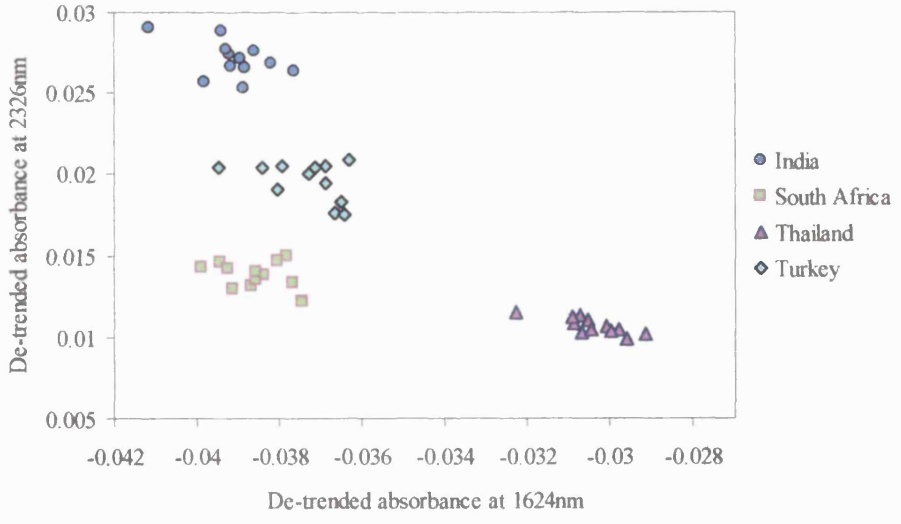


Figure 5.5. A two-wavelength plot of cannabis flowering heads from four geographical locations

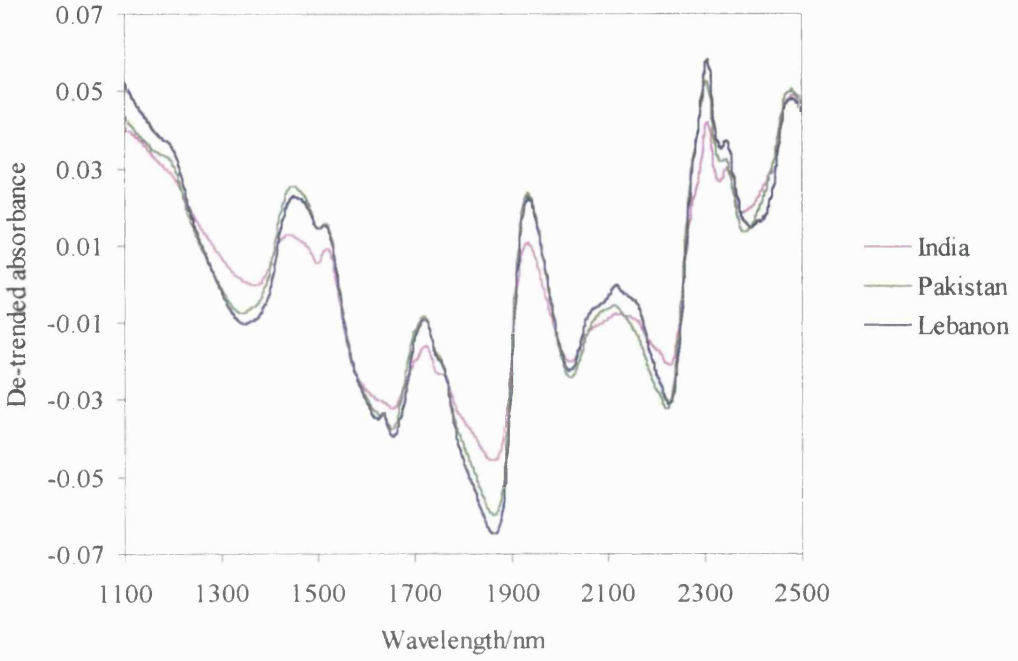


Figure 5.6. Baseline corrected (de-trended) spectra of cannabis resin from 3 geographical origins

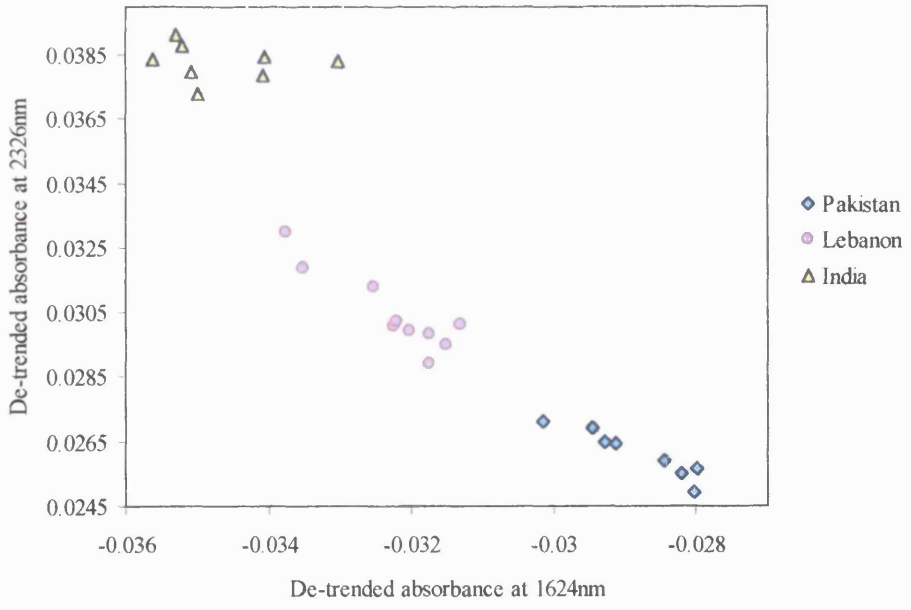


Figure 5.7. Absorbance values at 1624nm against absorbance values at 2326 for cannabis resin from 3 geographical origins

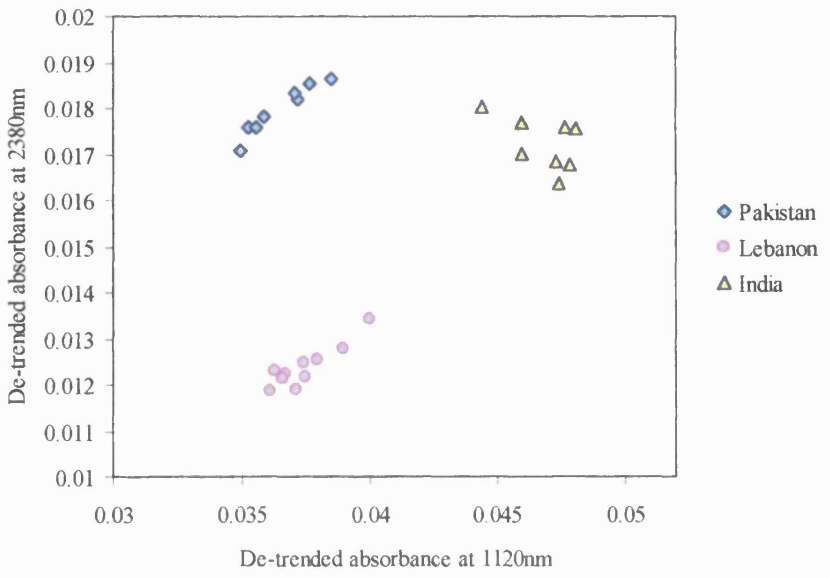


Figure 5.8. Absorbance values at 2380nm against absorbance values at 1120nm for cannabis resin from 3 geographical origins

5.3.7.3 Identification using 'nearest neighbours' (kNN) - Cannabis flowering heads and resin

The resulting plots from the two-wavelength method could be used to carry out a *k*NN or 'nearest neighbour' analysis on 'unknown' samples. Here, different samples of cannabis of known provenance were treated as if their origins were unknown. Incorporating the datapoints of these 'unknowns' into the plots showed that in both cases, visually, all the unknowns appeared to be closest to the cluster of points belonging to the group it theoretically belonged to (Figures 5.9 and 5.10). When Euclidean distance calculations were carried out on the resulting plots, it was clear that each 'unknown' species could be identified correctly. That is, it was possible to correctly assign all the samples that were treated as unknowns to the

countries that gave rise to the smallest Euclidean distances (i.e. 'nearest neighbours') (Tables 5.9 and 5.10). In the case of flowering heads, the differences between the smallest Euclidean distance and the next smallest was 0.0073, 0.0046, 0.0052 and 0.0052 for India, South Africa, Thailand, and Turkey, respectively. For the resin samples, they were 0.0033, 0.0026, and 0.0102 for Pakistan, Lebanon, and India, respectively. Thus, it could be said that the 'unknowns' without question were from the countries they were theoretically from. In all cases, the 12 smallest Euclidean distances were produced by the countries the 'unknowns' supposedly originated from.

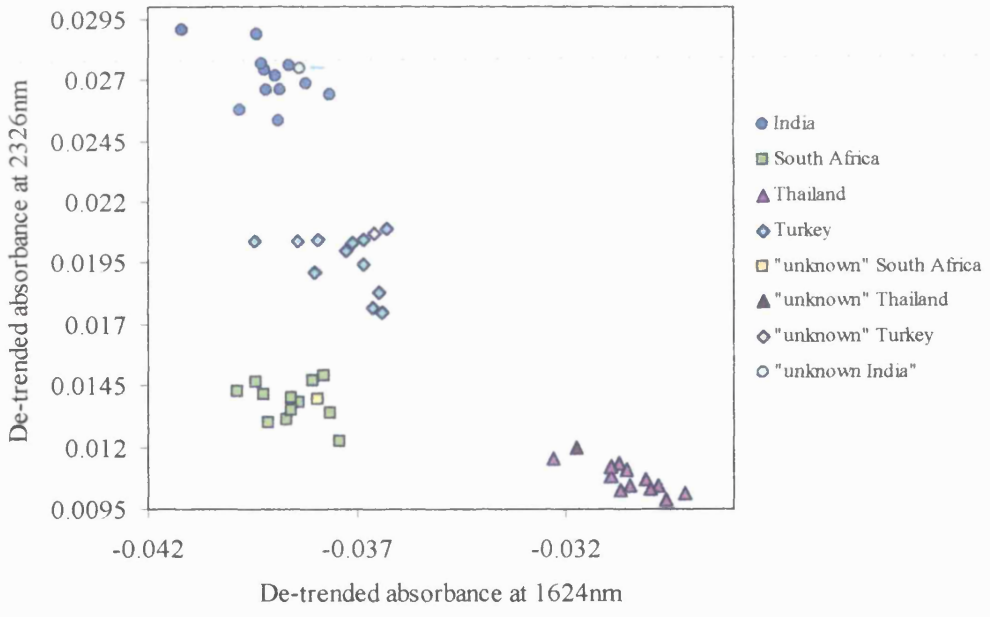


Figure 5.9. A two-wavelength plot of cannabis flowering heads incorporating 'unknown' samples

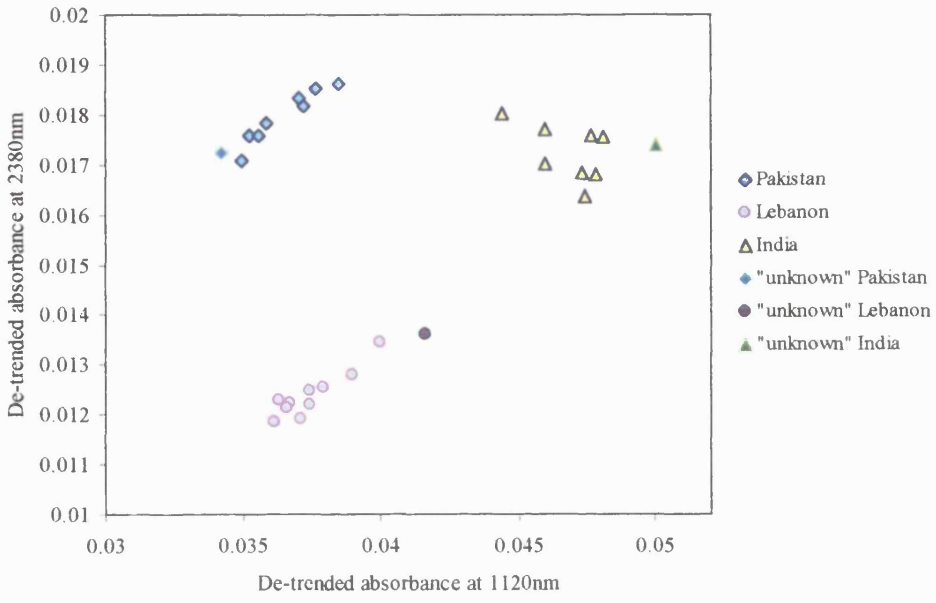


Figure 5.10. A two-wavelength plot of cannabis resin incorporating 'unknown' samples

Table 5.9. Euclidean distances between 'unknown' and 'known' cannabis flowering heads

	India	South Africa	Thailand	Turkey
'unknown' India	0.0028	0.0132	0.0168	0.0069
'unknown' South Africa	0.0158	0.0011	0.0071	0.0073
'unknown' Thailand	0.0206	0.0082	0.0019	0.0118
'unknown' Turkey	0.0101	0.0057	0.0095	0.0017

Values are averages. Smallest values in each column are underlined

Table 5.10. Euclidean distances between 'unknown' and 'known' cannabis resin

	India	Pakistan	Lebanon
'unknown' India	.0030	0.01290	0.0066
'unknown' Pakistan	0.0132	0.0028	0.0067
'unknown' Lebanon	0.0132	0.0061	0.0040

Values are averages. Smallest values in each column are underlined

5.3.8 Polar Qualification System

The PQS system was used on the samples of *Cannabis* flowering heads and resins to test its suitability for the discrimination of their geographical origins. Figures 5.11 to 5.12 show the plots obtained from second derivative spectra using the most successful wavelength ranges (in 100 nm increments) for each of these products. It can be seen from the clustering pattern observed from these that PQS is successful in separating the samples according to their geographical origins. However, while the clusters resulting from PQS analysis of resins were relatively tight, those resulting from analysis of flowering tops showed a tendency to spread horizontally. This may indicate that the suitability of this method depends on the type of material analysed. That is, while flowering tops are a complicated mixture of parts botanically and chemically, resin is an exudate and morphologically much simpler and consistent in character.

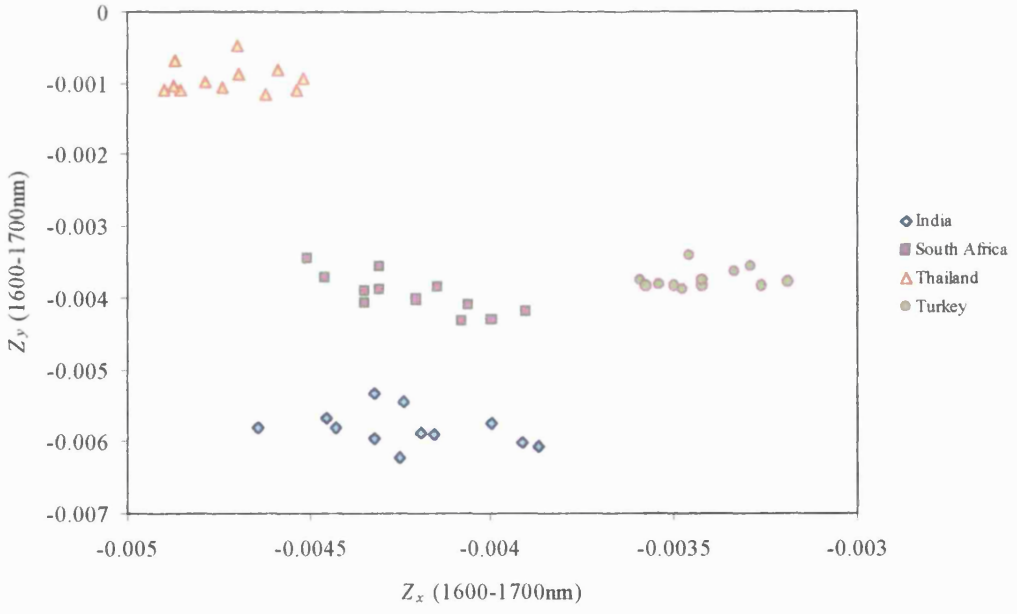


Figure 5.11 A PQS plot for cannabis flowering heads from four geographical origins using the wavelength range 1600-1700nm

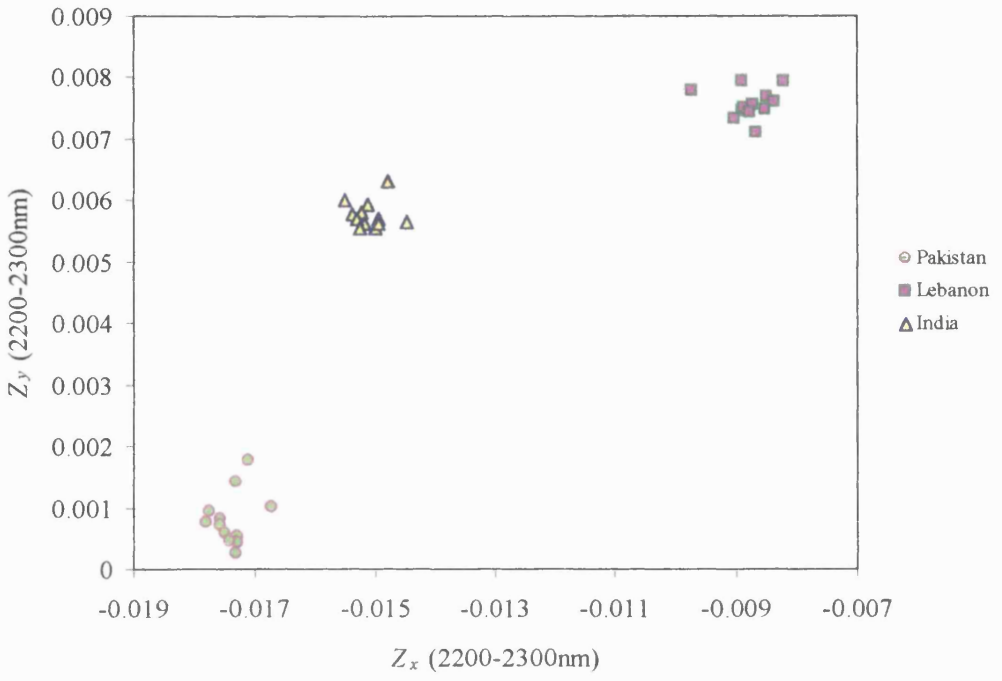


Figure 5.12 A PQS plot for cannabis resins from three geographical origins using the wavelength range 2200-2300nm

5.3.9 Other natural products

5.3.9.1 Identification- Maximum Distance in Wavelength Space

Tables 5.11 and 5.12 show resulting match values obtained when using the Maximum Distance in Wavelength Space method on 12 samples each of belladonna root and valerian root grown in England and India, and 12 samples each of aloes from 4 localities. These samples were chosen as they are grown in a variety of countries and were readily available from the archives at The School of Pharmacy. It is clear from the results that, as in the identification of the geographical origins of cannabis flowering heads and resin, this method is a successful one for these samples as well. In the valerian samples, the largest observed correct match value was 2.47, and the smallest was 1.40. Match values of

mismatched samples ranged from 23.05 to 34.91. For belladonna, the largest match value for correctly matched roots was 2.32, and the smallest 1.46. Mismatched samples produced match values ranging from 5.35 to 7.27. Thus, for these two samples, it can be said that the software default threshold match value of 4.0 enables the correct identification of the origins of these roots.

Results were just as successful in the case of the aloe resins. The largest match value observed in the correctly identified samples was 2.51, and the smallest was 1.43. In mismatched samples, match values were very high, ranging from 37.25 to 97.79. Thus, in this case as well, the threshold match value of 4.0 was found to be more than sufficient. However, it must be noted in the case of aloes that although they are named according to geographic localities, they can, in fact, often come from different species. The Cape aloes are from South Africa and Kenya, and are usually obtained from *Aloe ferox* and its hybrids; the Curaçao ones are from the West Indian Islands of Curaçao and are mainly taken from *Aloe barbadensis*; and Socotrine and Zanzibar varieties from *Aloe perryi* (Evans 1989g). However, adulteration and substitution among these types can occur (Evans 1989g) and thus it is important to be able to identify each type and its origin.

Table 5.11. Maximum Distance in Wavelength Space Match Values for samples of English and Indian belladonna and valerian root

Belladonna root	**English**	**Indian**
English	1.46-1.92-2.30	
Indian	5.35-6.41-7.27	1.54-1.95-2.32
Valerian root	**English**	**Indian**
English	1.66-1.95-2.47	
Indian	23.05-29.91-34.91	1.40-1.98-2.20

Left hand number = smallest value, Right hand number = largest value, Underlined number = median value
n = 12 for each sample

Table 5.12. Maximum Distance in Wavelength Space Match Values for samples of aloe resins

	Cape	Curacao	Socotrine	Zanzibar
Cape	1.55-1.94-2.42			
Curacao	74.02-88.69-96.99	1.54-1.99-2.41		
Socotrine	37.25-52.41-61.03	80.28-84.44-86.95	1.43-1.95-2.51	
Zanzibar	81.11-93.74-97.79	51.83-62.35-71.95	75.24-89.69-91.33	1.49-1.81-2.38

Left hand number = smallest value, Right hand number = largest value, Underlined number = median value
n = 12 for each sample

5.3.9.2 Correlation in Wavelength Space

As in the case of cannabis flowering heads and resins, the values obtained when carrying out the Correlation in Wavelength Space method were very high. Table 5.13 shows resulting values for valerian root. It can be seen from this that even between mismatched samples, the correlation value was as high as 0.97, which is well above the software default threshold value of 0.85. Correctly matched samples produced the value of 0.998 in both cases. To be able to correctly identify the origin of the samples, therefore, a high threshold value of 0.99 would need to be applied.

In the case of belladonna root and aloe resins, it was found when applying the method that the threshold value would need to be above 0.999 and 0.995, respectively. As these values are so high, it can be said that the Correlation in Wavelength Space method is not a feasible one for these products.

Overall, the Correlation in Wavelength Space method was not an incredibly successful one, as setting threshold values proved to be so problematic.

Table 5.13. Correlation in Wavelength Space results for valerian root from India and England

	English	Indian
English	0.998	
Indian	0.970	0.998

Values are mean results

5.3.9.3 Correlation Coefficients

Table 5.14 shows *r*-values obtained when calculating correlation coefficients for spectra of belladonna and valerian, and Table 5.15 shows those from the aloe resins. For the roots, values were high, being 0.995 and 0.970 for mismatched belladonna and valerian samples, respectively. For the aloe resins, values were also high, although slightly lower than those observed for belladonna and valerian. The high *r*-values, especially in the cases of belladonna and valerian, show how alike the samples are and therefore how sensitive a technique must be in order to correctly assign them to their country of origin.

Table 5.14. *r*-values for belladonna and valerian root from Indian and England

Belladonna root	**English**	**Indian**
English	1.0	
Indian	0.995	1.0
Valerian root	**English**	**Indian**
English	1.000	
Indian	0.970	1.000

Values are mean results

Table 5.15. *r*-values for aloe resins

	Cape	Curacao	Socotrine	Zanzibar
Cape	1.000			
Curacao	0.938	1.000		
Socotrine	0.897	0.963	1.000	
Zanzibar	0.890	0.936	0.896	1.000

Values are mean results

5.3.9.4 Residual Variance in Principal Components Space

The results obtained when using the Residual Variance in Principal Components Space method on belladonna and valerian root samples are shown in Table 5.16. Those resulting from the aloe resin samples are summarised in Table 5.17. It can be seen from these results that this method can be used for the assignation of each sample to their correct group. Variance from correctly matched belladonna and valerian samples did not exceed 0.7, with the lowest being 0.519 for Indian belladonna and 0.687 for Indian valerian. In the case of the aloes, values for correctly matched samples ranged from 0.65 for Socotrine and 0.71 for Zanzibar. Thus, in all cases, the software default threshold probability level of 0.84 appears to be a reasonable one, enabling correct identification of the geographical origins of the samples.

Table 5.16. Residual Variance in Principal Components Space results for belladonna and valerian

Belladonna root	**English**	**Indian**
English	0.655	
Indian	0.999	0.519
Valerian root	**English**	**Indian**
English	0.604	
Indian	1.000	0.687

Values are largest observed results

Table 5.17. Residual Variance in Principal Components Space results for aloe resins

	Cape	Curacao	Socotrine	Zanzibar
Cape	0.70			
Curacao	1.0	0.69		
Socotrine	1.0	1.0	0.65	
Zanzibar	1.0	1.0	1.0	0.71

Values are largest observed results

5.3.9.5 Two-wavelength plot

As in the cases of cannabis flowering heads and resin, the two-wavelength plots for belladonna, valerian, and aloes were attempted on de-trended spectra. Trying various combinations showed that although the absorbances at the wavelengths 2380 and 1120 separated the valerian samples well (Figure 5.13), results were not so successful for belladonna (Figure 5.14) nor for the aloes (Figure 5.15). In fact, it appeared to be impossible to produce successful separation between the groups for the belladonna roots and aloes. For belladonna, although there is some linear pattern discernible in the two-wavelength plot, there is no clear separation between the two groups as seen in valerian. For the aloes, only the Socotrine variety showed clustering behaviour.

As the valerian samples produced a reasonable two-wavelength plot, a *k*NN nearest neighbours analysis was carried out on an 'unknown' sample. Figure 5.16 shows the datapoints of the 'unknowns' incorporated into the plot. It can be seen that the 'unknowns' are comfortably nearest the groups which they theoretically belong to. Table 5.18 shows average Euclidean distances between the 'unknowns' and the 'knowns', and it is clear from this that the smallest values were observed for the group that each unknown was expected to belong to. Thus, with a successful two-wavelength plot, it is possible to use the *k*NN method for the correct identification of country of origin of a sample, in this case, valerian.

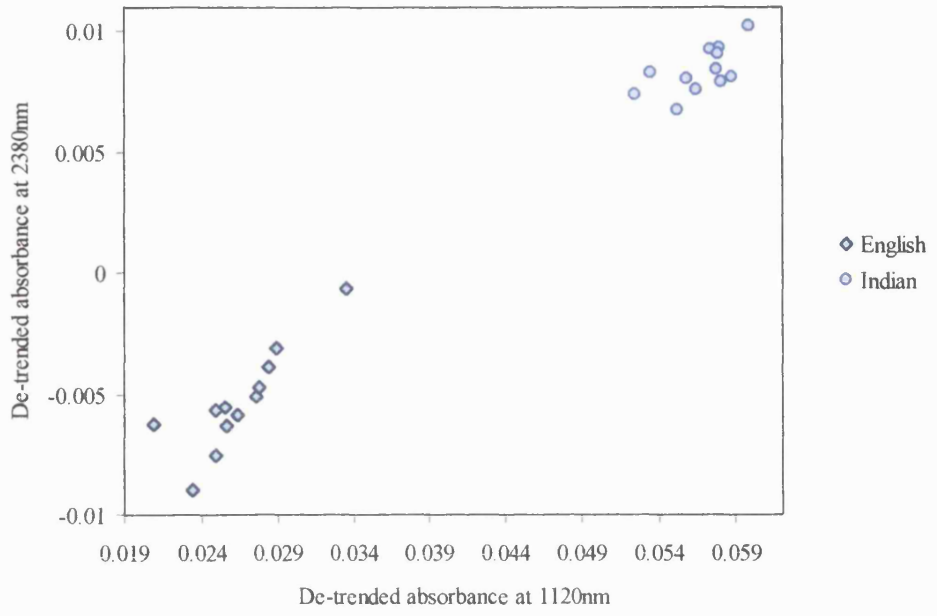


Figure 5.13. A two-wavelength plot for valerian root samples

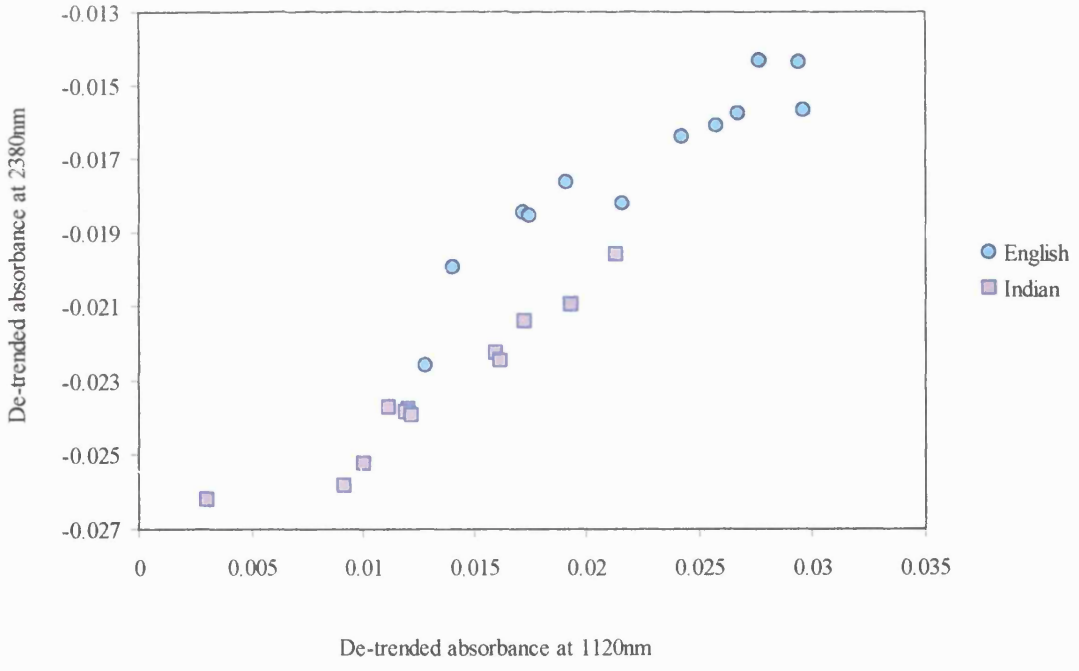


Figure 5.14. A two-wavelength plot for belladonna root samples

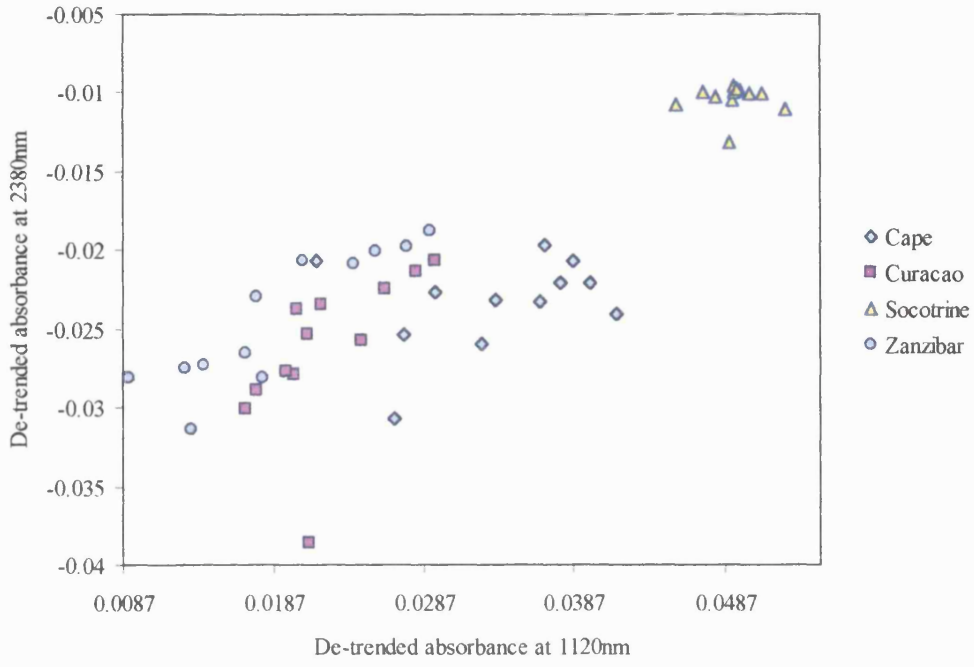


Figure 5.15. A two-wavelength plot for aloe resin samples

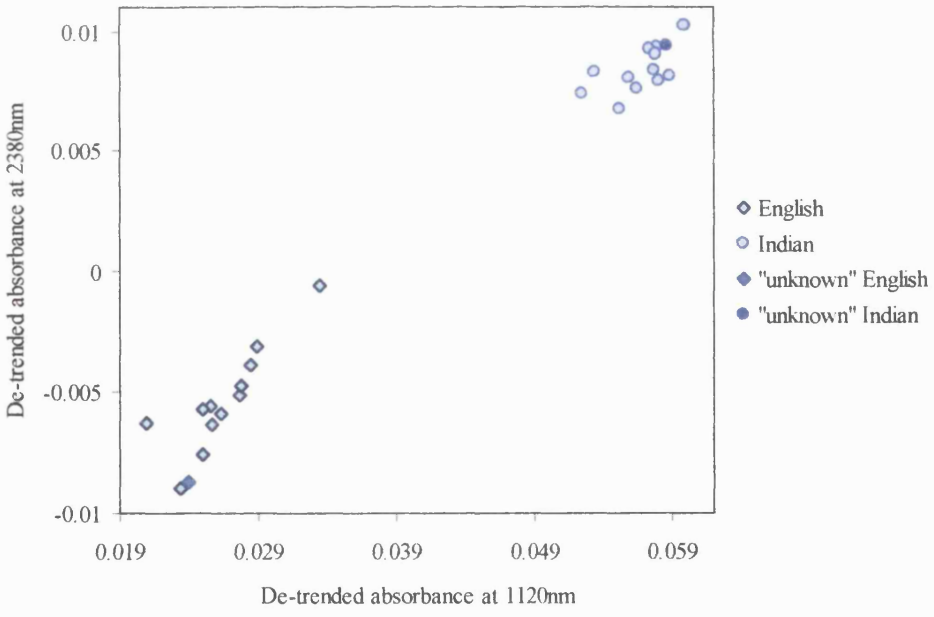


Figure 5.16. A two-wavelength plot for valerian root samples, incorporating 'unknowns'

Table 5.18. *k*NN Euclidean distances for valerian root samples

	English	Indian
English 'unknown'	<u>0.0047</u>	0.0353
Indian 'unknown'	0.0370	<u>0.0025</u>

Values are mean results. Smallest values in each column are underlined

5.3.9.6 Polar Qualification System

When using the PQS method on the samples of valerian, belladonna and aloe, it was possible to achieve a greater degree of separation among the various geographical origins compared to the two-wavelength method. Figures 5.17 to 5.19 show the plots obtained from second derivative spectra using the best wavelength ranges (in 100 nm increments) for each of these products. It is clear from these that PQS is successful in separating the samples according to their origin, and in the cases of aloes and belladonna, more successfully than the two-wavelength method. All samples show a distinct clustering behaviour. The different wavelength ranges may represent differences in active constituents or other components. However, examination of these was beyond the time frame of this project.

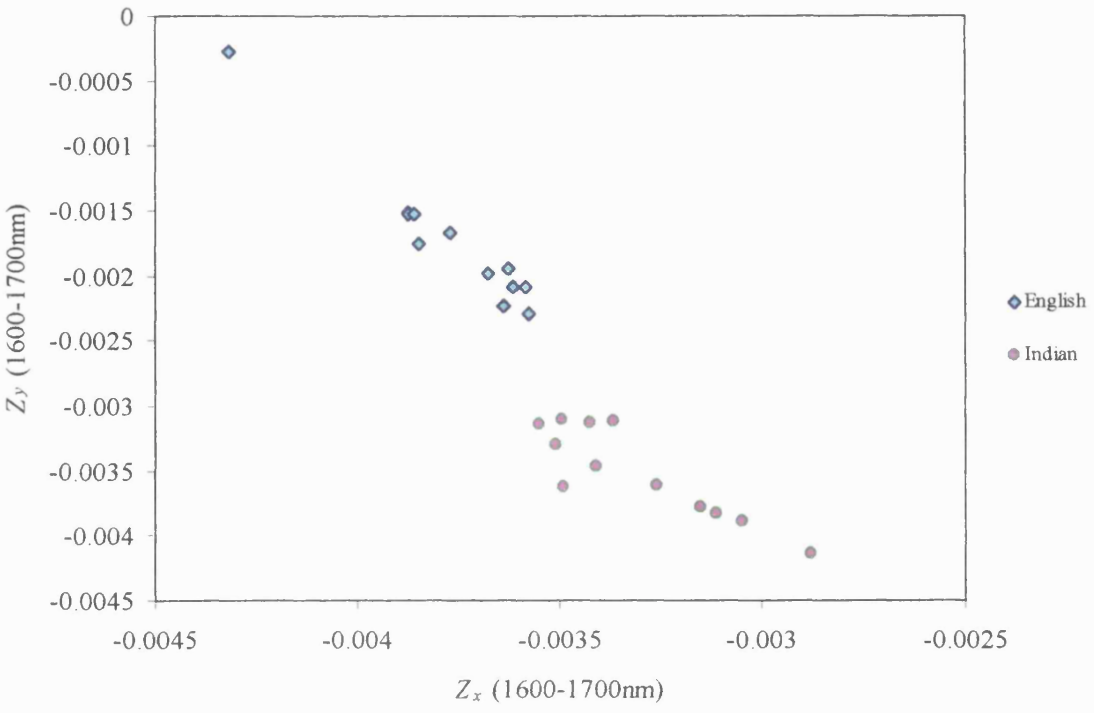


Figure 5.17. A PQS plot of belladonna root from India and England using the wavelength range 1600-1700nm

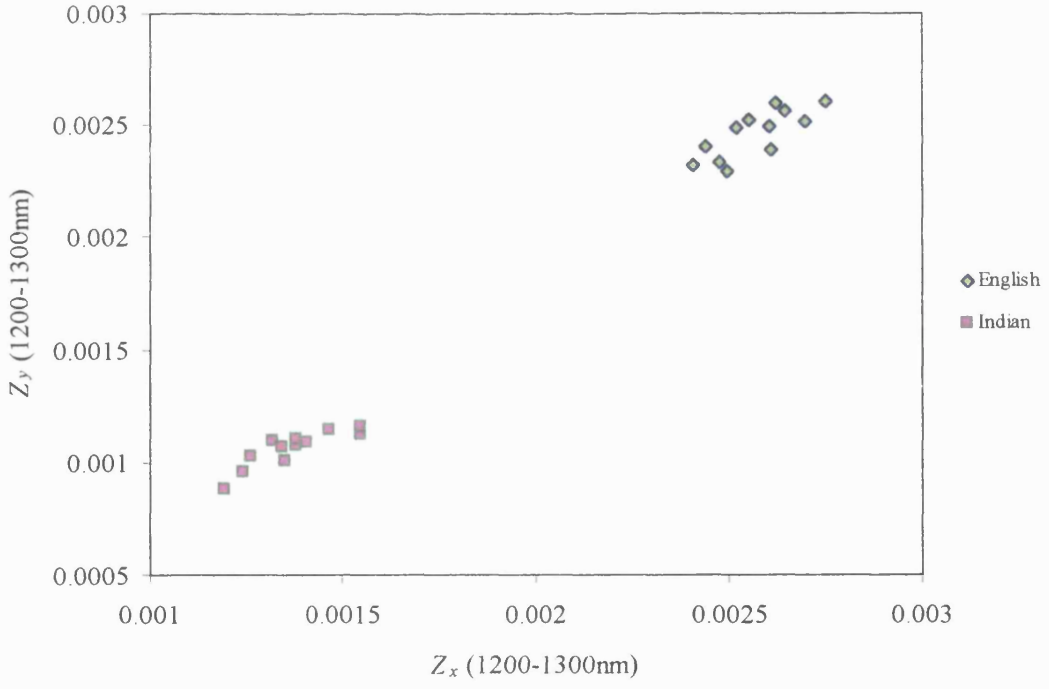


Figure 5.18. A PQS plot of Valerian root from England and India using the wavelength range 1200-1300nm

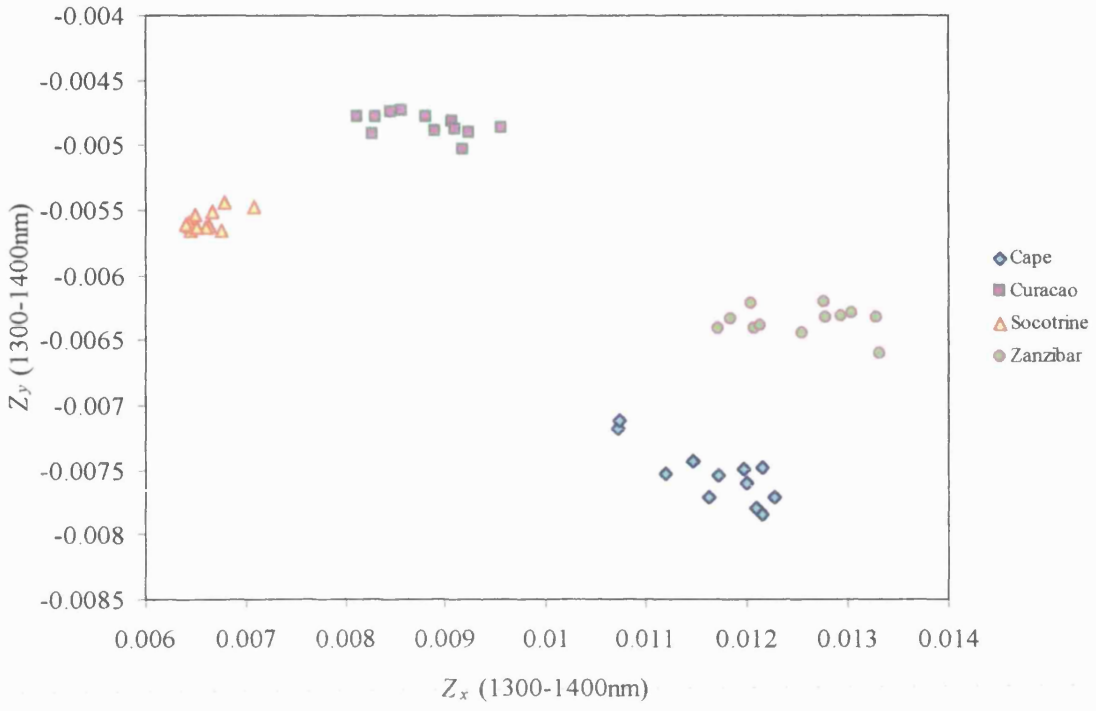


Figure 5.19. A PQS plot for four types of aloes using the wavelength range 1300-1400nm

5.4 Conclusion

Tables 5.19 and 5.20 show a summary of results obtained from the various identification techniques used for cannabis flowering heads and resins, respectively. From these data, it can be concluded that overall, the Maximum Distance in Wavelength Space method appears to be the most successful method for allocating each sample to its place of origin. The match values that were calculated were very distinctly low or high depending on whether samples were matched or mismatched. The use of the Correlation in Wavelength Space method and also of *r*-values for identification was also possible, although results were not as clear-cut as in Maximum Distance in Wavelength Space. That is, the values were often very high, and the setting of thresholds was somewhat problematic. The high values, however, reflect on how similar the samples are, and therefore to some extent how sensitive a method Maximum Distance in Wavelength Space is. The Residual Variance in Principal Components Space method was also successful, with all correctly matched samples falling within the threshold. With the successful plotting of a two-wavelength plot, it was also possible to carry out correct identification of unknown samples by using the *k*-Nearest Neighbours method.

In the case of other natural products, the conclusions reached after the cannabis investigations were partly confirmed. Again, Maximum Distance in Wavelength Space appeared to produce distinctly high or low match values depending on whether samples were correctly matched or not. Correlation in Wavelength Space proved problematic, and in the cases of belladonna and aloes, a reasonable

threshold could not be reached, as values were so high. This appears to confirm the observation that perhaps this is not the most ideal method overall for identification. Correlation coefficients were also rather high, as was also observed in the cannabis samples. The use of Residual Variance in Principal Components Space was, as in the cannabis study, also successful, although results did not appear as clear-cut as in Maximum Distance in Wavelength Space. The use of two-wavelength plots and *k*NN met with very limited success here, as although successful in valerian, it was not possible to successfully produce clear two-wavelength plots for belladonna and aloes. This therefore illustrates how the *k*NN method using the two-wavelength plot can only be carried out if a plot can be produced in the first place, and thus its high success rate is dependent upon groups being able to separate distinctly between two axes.

The PQS method showed mixed results in that while plots were comparable to those produced by the two-wavelength method for cannabis samples, those obtained for other natural products (and in particular belladonna and aloe) showed a significant improvement. That is, unlike the two-wavelength method, it was possible to produce a plot showing clear clustering of points for both belladonna and aloe. This suggests that both the two-wavelength method and the PQS method have varying degrees of success, often depending on the type of material analysed.

Overall, by looking at the results from cannabis and other natural products, it can be concluded that NIRS can be used for the successful determination of their geographical origins and probably those of numerous other plant materials. This would be advantageous in that it is significantly more rapid than traditional

methods, and as the plants are of identical species (except in the case of aloe resins), it would be difficult for traditional techniques to pick out the minute differences that may occur between plants grown in varying locations. The factors that contribute to the differences, which may be aspects such as climate, soil, rainfall and fertiliser, would be an interesting continuation of this study.

Table 5.19. Summary of results for cannabis flowering heads

India	**A. Match value**	**B. Correlation**	**C. *r*-value**	**D. Res. variance**	**E. *k*NN**
India	1.08	0.999	1.0	0.585	0.0028
South Africa	11.68	0.987	0.986	1.000	0.0158
Thailand	39.9	0.981	0.981	0.995	0.0206
Turkey	17.8	0.982	0.983	1.000	0.0101
South Africa	**A. Match value**	**B. Correlation**	**C. *r*-value**	**D. Res. variance**	**E. *k*NN**
India	11.68	0.987	0.986	1.000	0.0132
South Africa	0.97	0.999	1.0	0.610	0.0011
Thailand	17.0	0.983	0.984	0.994	0.0082
Turkey	23.4	0.980	0.980	1.000	0.0057
Thailand	**A. Match value**	**B. Correlation**	**C. *r*-value**	**D. Res. variance**	**E. *k*NN**
India	39.9	0.981	0.981	0.995	0.0168
South Africa	17.0	0.983	0.984	0.994	0.0071
Thailand	3.63	0.999	1.0	0.473	0.0019
Turkey	24.3	0.988	0.988	0.999	0.0095
Turkey	**A. Match value**	**B. Correlation**	**C. *r*-value**	**D. Res. variance**	**E. *k*NN**
India	17.8	0.982	0.983	1.000	0.0069
South Africa	23.4	0.980	0.980	1.000	0.0073
Thailand	24.3	0.988	0.988	0.999	0.0118
Turkey	1.97	0.998	1.0	0.654	0.0017

A = Average match value when using Maximum Distance in Wavelength Space, B = Average correlation value when using Correlation in Wavelength Space, C = Average *r*-value resulting from the correlation coefficient method, D = Highest probability level when using Residual Variance in Principal Components Space, and E = Average Euclidean distance when using the *k*NN method. Correctly matched results are underlined.

Table 5.20. Summary of results for cannabis resin

India	**A. Match value**	**B. Correlation**	**C. *r*-value**	**D. Res. variance**	**E. *k*NN**
India	1.91	0.999	1.000	0.549	0.0030
Pakistan	52.3	0.863	0.943	1.000	0.0132
Lebanon	62.0	0.965	0.961	1.000	0.0132
Pakistan	**A. Match value**	**B. Correlation**	**C. *r*-value**	**D. Res. variance**	**E. *k*NN**
India	52.3	0.863	0.943	1.000	0.0129
Pakistan	1.96	0.999	1.000	0.680	0.0028
Lebanon	59.6	0.959	0.966	1.000	0.0061
Lebanon	**A. Match value**	**B. Correlation**	**C. *r*-value**	**D. Res. variance**	**E. *k*NN**
India	62.0	0.965	0.961	1.000	0.0066
Pakistan	59.6	0.959	0.966	1.000	0.0067
Lebanon	1.98	0.997	1.000	0.680	0.0040

A = Average match value when using Maximum Distance in WavelengthSpace, B = Average correlation value when using Correlation in Wavelength Space, C = Average *r*-value resulting from the correlation coefficient method, D = Highest probability level when using Residual Variance in Principal Components Space, and E = Average Euclidean distance when using the *k*NN method. Correctly matched results are underlined.

Chapter 6: Controlling the Drying Process of Peppermint Leaves Using Near-Infrared Spectroscopy

6.1 Introduction

The British Pharmacopoeia places moisture limits on twelve leaf materials ranging from 6% for *Digitalis* (British Pharmacopoeia 2000a) to 12% for *Senna* (British Pharmacopoeia 2000b). Because these limits exist, it is clear that the quantification of water in plants of pharmaceutical interest is of utmost importance. In conjunction with a suitable temperature, moisture will lead to the chemical decomposition of active ingredients, activation of enzymes leading to degradation, and to the proliferation of living organisms. As most vegetable drugs contain all the essential food requirements for moulds, insects, and mites, deterioration can be rapid once infestation has taken place (Evans 1989h). Thus it is uneconomical to purchase drugs that contain excess moisture. Living plant material has a high moisture content: leaves can contain 60-90%, roots and rhizomes 70-85%, and wood 40-50%. The lowest percentage (5-10%) is found in seeds (Samuelsson, 1999).

A large number of methods are available for the determination of moisture in natural products. These include distillation, microwave drying, oven-drying techniques, and Karl Fischer titration (Burns and Ciurczak, 1992). The most commonly used techniques, however, are determination by loss on drying and Karl Fischer titration (Evans 1989a).

Loss on drying is a method described in the British Pharmacopoeia (British Pharmacopoeia 2000c). This method assumes that the loss in weight in the sample during the procedure is mainly due to water. However, small amounts of other volatile materials can contribute to the weight loss. In fact, differences in moisture

values between results obtained by the oven-drying technique and those from non-thermal procedures have been observed (Burns and Ciurczak, 1992). This method is suitable where large numbers of samples are handled and where a continuous record of loss of weight with time is required (Evans 1989h).

The Karl Fischer procedure is also specified in the British Pharmacopoeia (British Pharmacopoeia 2000d) and is an extensively used and accurate method for moisture determination. It is ideal to use on expensive drugs and chemicals containing small quantities of moisture. The process involves a titration using a reagent (a solution of sulphur dioxide, iodine and pyridine in dry methanol) against a sample containing water. This causes a loss of the dark brown colour. At the end point, when there is no more water, the colour of the reagent persists. More modern Karl Fischer instruments, such as the one used in this investigation, operate using a potentiometric method. Although accurate, the main drawbacks of the method are the instability of the reagent and the possibility of substances in the sample, other than water, which may react with the reagent (Evans 1989h). In addition, long extraction times are often necessary in order to confirm complete removal of the contained water from the sample matrix.

NIRS is a novel method for the quantification of water in materials. The NIR absorption spectrum of water at room temperature includes 5 bands with maxima at 1940, 1450, 1190, 970, and 760nm (Osborne et al 1993). These peaks can therefore be used to determine the moisture content of materials. It is possible to design NIR instruments which are capable of accurately measuring changes in reflected intensity from the product and comparing intensities of the absorption

wavelengths with the intensities of reference wavelengths. In many cases, the resulting ratio will provide an output proportional to the concentration of moisture, or other constituent under investigation. Mathematical algorithms can then be applied to convert the computed information into the measurement of interest, such as percentage of moisture, fat, or protein (MM710 User's Manual 2000a).

It is important that the drying process for natural products be performed under optimum conditions. Although fairly rapid drying helps aromatic drugs to retain their aroma and leaves and flowers their colour, the temperature used in the drying process must be governed by the constituents and physical nature of the drug (Evans 1989i). This is particularly true of materials such as peppermint leaves, which contain volatile oils. In general, leaves, flowers, and herbs may be dried between 20 and 40°C, and barks and roots between 30 and 65°C. The drying process therefore must be a balancing act between the necessity for quick drying and the sensitivity to heat of the constituents (Samuelsson, 1999).

In the herbal industry, there is generally a moisture loss of 80 to 90 % from freshly harvested materials during the drying process. Material is usually considered dry once the moisture level has fallen to 8 to 10 %. At this stage, leaves will crumble and stems will snap (Evans 1989i).

The evaluation of moisture content in natural products is not only confined to the pharmaceutical industry, however. In the food industry, moisture analysis is also important in that it can affect the storage quality and even the value of natural foodstuffs. For example, the ideal moisture content for rice is near but below 22%. As the optimal moisture content for harvesting rice is too high for safe storage,

rice must be dried. If drying occurs too rapidly, the rice will crack, as a result of the internal stresses exceeding the tensile strength of the kernel. As cracked kernels are more susceptible to breakage during the milling process, this results in a reduced yield and hence reduced profits. In fact, broken white rice is worth only about half as much as whole grain rice in world markets. If rice is delivered with a moisture content above 22%, extra costs are incurred for separate handling and drying, and can result in farmers being penalised (Blakeney et al, 1994).

In addition to the pharmaceutical and food industries, the measuring of water content in natural products is also of importance in the textile industries. In the US cotton industry, fibre strength has been included in a grading system of the product. Small changes in moisture content can produce large changes in strength, and so measurement accuracy depends on the moisture level at the moment of testing. Although there are elaborate humidity controls and extensive sample conditioning, moisture among test cottons can differ greatly. Moisture history, chemical composition, and structural differences all contribute to these differences. In addition, contaminants that may be produced during harvest (i.e. leaf, grass, bark, etc) make moisture measurements difficult, but necessary (Taylor, 1994).

In the tobacco industry, it is of utmost importance that the water content be measured at all stages of processing, as moisture can influence various characteristics of the product. For instance it can affect the ageing of tobacco, the use of stemming equipment and quality of the leaf, the efficiency of the cigarette making machines and the uniformity of the finished cigarette. In addition, it can

influence the filling capacity of the tobacco, the shelf life and the smoking characteristics of the finished cigarette (Burns and Ciurczak, 1992).

The use of NIRS for the determination of moisture in natural products has been well documented. For example, Burns and Ciurczak (1992) describe an NIR calibration that was developed for moisture on 13 tobacco blends over the concentration range 9-15%. An oven-volatile method was used as the laboratory analysis procedure. Results showed that values were acceptable for all validation sets (Burns and Ciurczak). In the food industry, the on-line monitoring of moisture during the flour milling process using an NIR instrument has also been documented (Chalmers 2000), and Martens and Naes (1989) mention the potential of NIRS over traditional methods for the determination of major wheat constituents, including moisture. Another investigation involving dried Korean ginseng (*Panax ginseng*) showed that NIRS measurements for moisture were highly comparable to the traditional oven drying method, with correlation values being 0.899 and 0.996 for white ginseng powder and extract, respectively (Cho and Lee, 1994).

However, studies that document in detail the use of NIRS constantly over a production process are very rare, especially those that describe the continuous measurement of moisture levels during the drying process of a product. In order to minimise the loss of important volatile materials, it is necessary to dry a sample to a specific required moisture percentage and not any further. Therefore, a process that could be used to continuously monitor the drying procedure would be of considerable value in the pharmaceutical and herbal industries.

This study compared the most commonly used methods of moisture analysis (loss on drying and Karl Fischer) to NIRS to determine how it could potentially be used to control the drying process of peppermint (*Mentha piperata*) leaves down to the British Pharmacopoeia limit of 11% (British Pharmacopoeia 2000e). Although the BP reference method for water content in peppermint is distillation, this was not used in this investigation due to time constraints. In addition, it was not thought necessary, as the NIR instrument used had initially been calibrated using the Karl Fischer technique. Although the instrument was a prototype calibrated to industry standards for tea, it was considered suitable for other leaf materials including peppermint. Other leaves (*Digitalis lanata*, *Digitalis ambigua*, *Digitalis mertonensis*, *Digitalis lutea*, coriander, coltsfoot, eyebright, tarragon, thyme) were also analysed by both NIRS and Karl Fischer titration to show the general applicability of the method. If successful, NIRS would be an advantageous technique, as it is rapid and non-destructive (Moffat et al 1997).

6.2 Materials and Methods

6.2.1 Instrumentation

NIR spectra were measured using a FOSS NIRSystems (Silver Spring, MD, USA) 6500 Rapid Content Analyser (RCA) in diffuse reflectance mode over the wavelength range 1100-2500nm. NIR measurements for water were also made using an MM710 NIR backscatter gauge (NDC Infrared Engineering, Maldon, Essex, UK) pre-calibrated using SpeedCal® (MM710 User's Manual 2000b) (Figure 6.1). Karl Fischer measurements were made using a 701 KF Titrino titrator (Metrohm, Buckingham, UK).

6.2.2 Materials

Fresh whole peppermint (*Mentha piperata*) leaves were purchased from a local supermarket, as were fresh samples of coriander, tarragon and thyme leaves. Dried samples of coriander, coltsfoot and eyebright leaves were puchased from a health food store, and dried specimens of *Digitalis lanata*, *Digitalis ambigua*, *Digitalis mertonensis* and *Digitalis lutea* leaves were obtained from the pharmacognosy archives at The School of Pharmacy, University of London. Only four *Digitalis* species were used in this study as these were the ones that were available in a large enough supply to make measurements by the MM710 feasible.

The Karl Fischer analysis used Karl Fischer reagent (Merck, Dorset, UK) and 200mg sodium tartrate (Fluka Chemicals, Dorset, UK) as a standard.

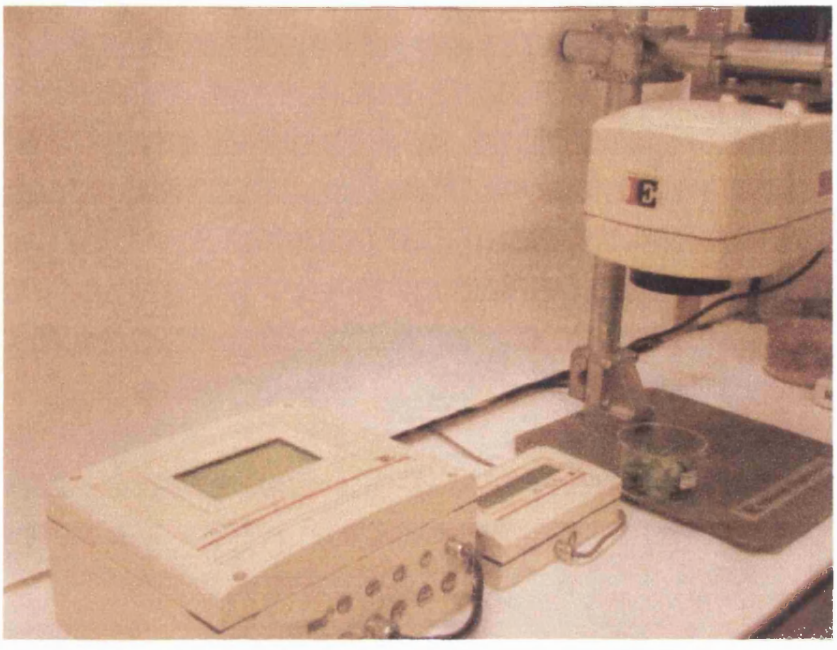
6.2.3 Method

Fresh peppermint leaves were placed in a circular 10cm glass dish under the beam of the MM710 NIR gauge and readings for percentage moisture were taken. Karl Fischer titration was also used to determine percentage moisture. Three different preparations of the leaves (whole, coarsely chopped, and finely chopped) were used to determine which sample preparation produced the best correlation between the two methods. The samples were then dried in an oven at 35°C and further analysed for water by NIRS, by Karl Fischer titration, and by weight loss every 20 minutes until readings indicated that they had reached below the moisture percentage specified in the British Pharmacopoeia (11%). In the Karl Fischer procedure, an extraction time of 10 minutes was allowed to verify complete

extraction of water from the sample. NIR spectra were also taken during the drying process, using the FOSS NIRSystems Rapid Content Analyser, with samples placed in 10 mm diameter glass vials. Each spectrum was an average of 32 scans. Resulting data were processed and plotted in Microsoft Excel 97.

The other leaf samples (*Digitalis lanata*, *Digitalis ambigua*, *Digitalis mertonensis*, *Digitalis lutea*, coriander, coltsfoot, eyebright, tarragon, thyme) were left whole and tested for water content using the MM710 gauge and Karl Fischer titration.

A



B

C

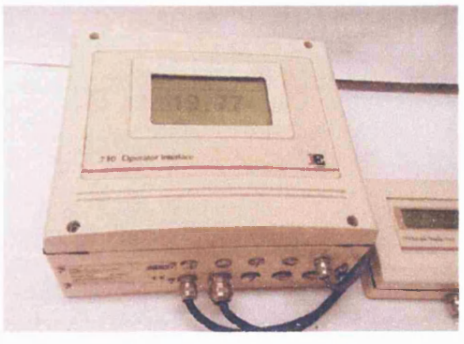


Figure 6.1. The MM710 NIR gauge. Figure 6.1A shows the whole instrument set-up; Figure 6.1B shows the part of the instrument housing the NIR spectrometer and Figure 6.1C shows the gauge displaying a resulting moisture percentage. Note the glass dish containing peppermint leaves under the beam path of the instrument in 6.1A and 6.1B.

6.3 Results and discussion

6.3.1 Sample preparation

The NIR gauge irradiated a relatively small area (60 mm diameter), and as a result, it was suspected that the samples might appear to be non-homogeneous and therefore give imprecise values. Different forms of sample preparation were therefore tested to determine which produced the most accurate measurements. Figure 6.2 compares the moisture percentages obtained using Karl Fischer titration and the NIR gauge for three different sample preparations of fresh peppermint leaves. It can be seen that the best correlation between the two methods was obtained when the leaves were chopped coarsely and the worst when they were chopped finely. The finely chopped leaves appeared visually to be considerably moister compared to the other sample preparations. This suggests that chopping the leaves very finely caused the release of large amounts of water to the surfaces, allowing the water to be detected more readily by the NIR gauge, resulting in larger values for the MM710 compared to Karl Fischer titration. Coarsely chopped leaves may have produced the best results due to the fact that the sample was made more uniform throughout. As NIR gauge readings were most representative of the area within the beam patch of the instrument (60 mm diameter) and of the top layers of the sample as a whole, slightly varying results were seen when using whole leaves. Indeed, NIR gauge readings were seen to range from 68.7% to 73.9% water content for fresh whole leaves. However, as Karl Fischer and NIR values were within a reasonable range (69.3%-75.0%), all further experiments were carried out on whole leaves. This was done in consideration of the fact that in

practice, in the production process, peppermint leaves are dried whole and intact. The British Pharmacopoeia states that whole peppermint leaves should have no less than 12 ml/kg volatile oil and cut leaves no less than 9 ml/kg (British Pharmacopoeia 2000e). This suggests that cutting the leaves ultimately results in loss of volatile oil, which was another reason to leave the leaves whole.

6.3.2 Spectral characteristics

Figure 6.3 shows the spectra of whole peppermint leaves at various stages in the drying process taken using the FOSS NIRSystems Rapid Content Analyser. From these spectra, it can be seen that the spectral peaks, particularly those peaks representing water, decrease in height as the percentage of moisture decreases. For instance, the peak in the vicinity of 1934nm can be seen to decrease in height, as the sample becomes drier. Thus, taking spectra during the process can show how spectral characteristics change as moisture is driven away from the sample.

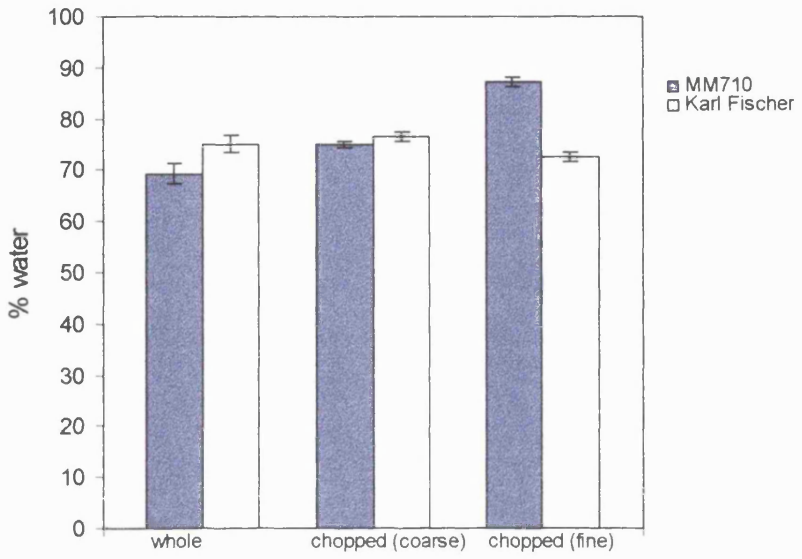


Figure 6.2. Percentage water using Karl Fischer and MM710 for three different sample preparations of fresh peppermint leaves. Error bars indicate ± 1 standard deviation (n=6).

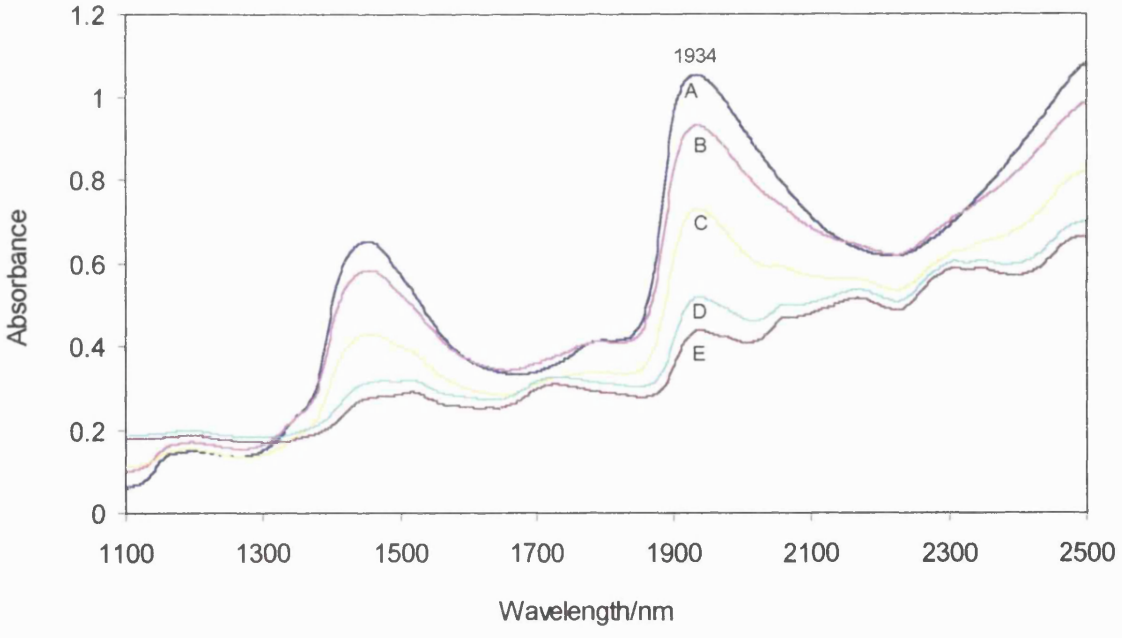


Figure 6.3. Sample spectra of whole peppermint leaves dried in an oven at 35°C. Water content: A=72.0%, B=61.5%, C=56.3%, D=14.2%, and E=3.53% according to Karl Fischer titrations.

6.3.3Drying of whole leaves

Water measurements for the whole leaves from the previous investigation were taken by NIRS, Karl Fischer titration, and loss on drying every 20 minutes as they were dried in an oven at 35°C. Figure 6.4 shows a comparison of moisture percentages obtained from loss on drying and NIRS, while Figure 6.5 shows one from Karl Fischer titrations and NIRS for whole leaves. Although results appear linear in both cases, a higher correlation (R^2= 0.992) was observed when NIRS results were compared to Karl Fischer titration results compared to loss on drying (R^2=0.986). R^2 values using the 6500 RCA data for absorbance values at 1934 nm against Karl Fischer, MM710, and loss on drying data were slightly lower, being 0.946, 0.943, and 0.963, respectively. These lower R^2 values from the 6500 RCA data may be due to the fact that measurements were made from an area of 10 mm diameter, as opposed to 60 mm for the MM710.

In the case of coarsely chopped leaves, results were similar, although a slightly higher correlation was observed compared to that obtained with whole leaves. Again, there was a better correlation between NIRS and Karl Fischer titration values (R^2=0.995) (Figure 6.6) than between NIRS and loss on drying values (R^2=0.986)(Figure 6.7). Differences between readings were observed to be as small as 1.23% for NIRS and loss on drying values and 0.32% for NIRS and Karl Fischer titrations. The smallest differences in the whole leaves, were higher, with values between NIRS and Karl Fischer and between NIRS and loss on drying being 0.61% and 5.34%, respectively, but still practically useable (Table 6.1).

Thus, the small percentage differences quoted above suggest that the accuracy of the procedure is quite high. Precision of the method was also shown to be high, by the small values for the standard deviations of measurement for Karl Fischer titration and NIRS, as displayed in Figure 6.2.

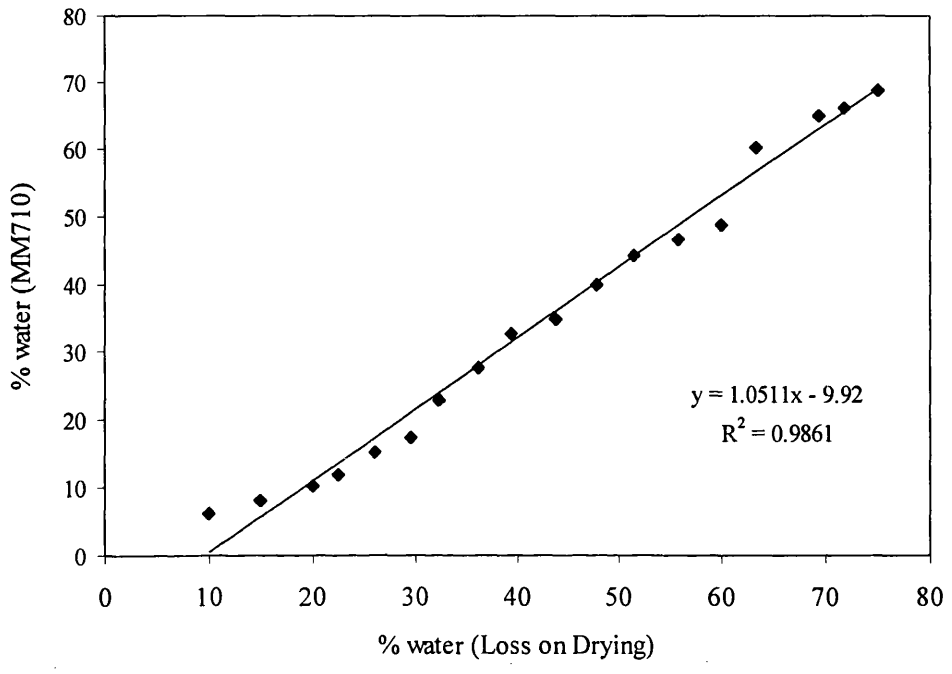


Figure 6.4. Percentage water using MM710 and Loss on Drying for whole peppermint leaves dried at 35°C and sampled every 20 minutes

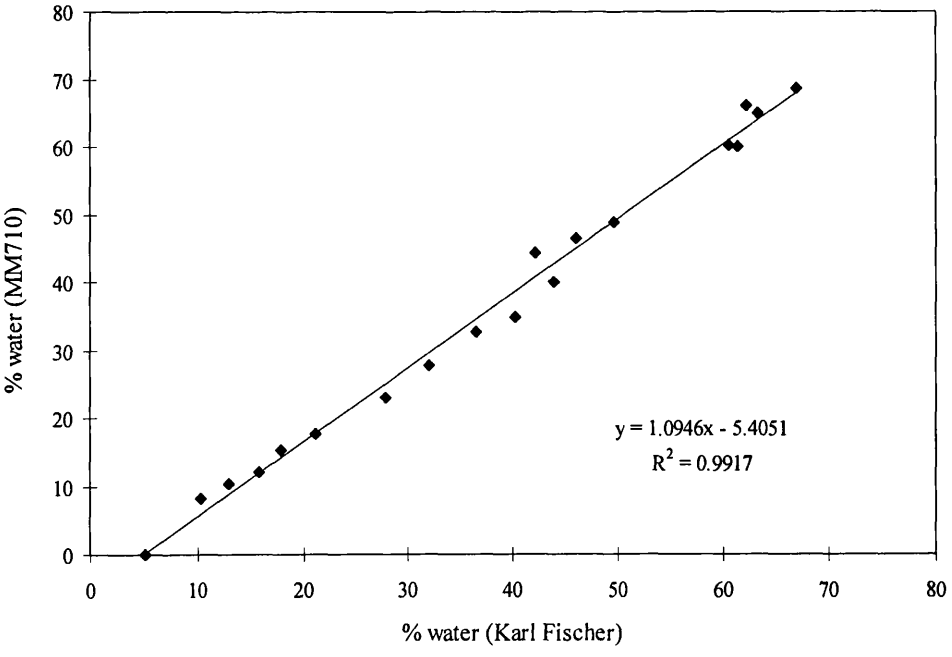


Figure 6.5. Percentage water using Karl Fischer and MM710 for whole peppermint leaves dried at 35°C and sampled every 20 minutes

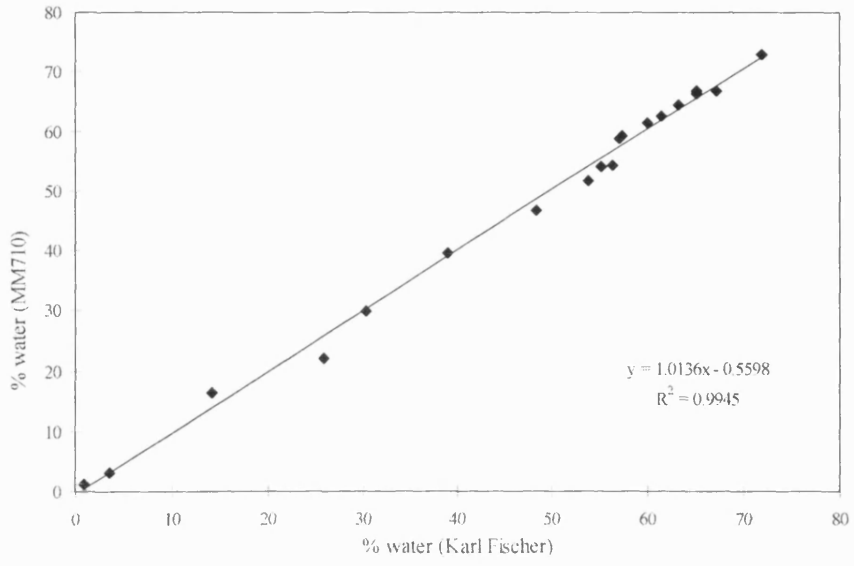


Figure 6.6. Percentage water using Karl Fischer and MM710 for coarsely chopped peppermint leaves dried at 35°C and sampled every 20 minutes

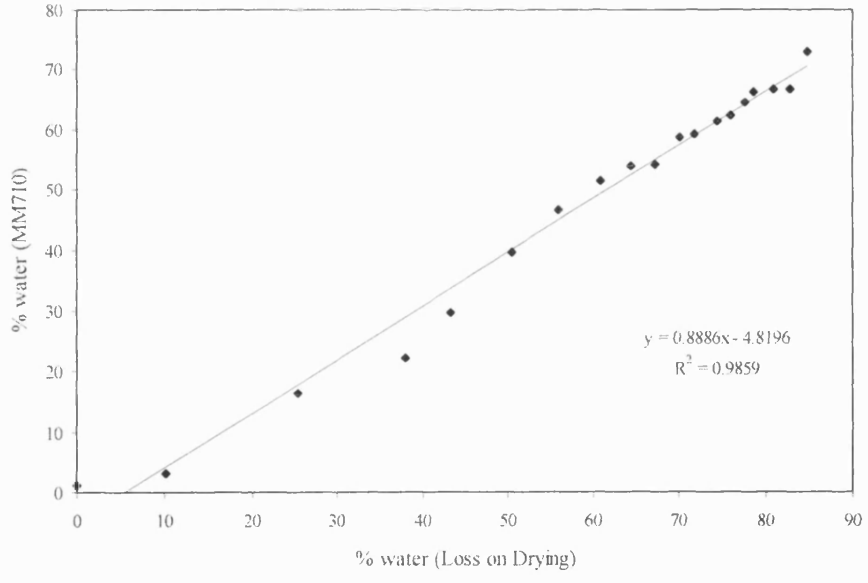


Figure 6.7. Percentage water using MM710 and Loss on Drying for coarsely chopped peppermint leaves dried at 35°C and sampled every 20 minutes

Table 6.1. Differences observed between NIR and Karl Fischer and between NIR and loss on drying for whole and coarsely chopped leaves

	Smallest difference (%)	Largest difference (%)
Whole (KF and NIR)	0.33	5.34
Coarsely chopped (KF and NIR)	0.32	3.8
Whole (Loss on drying and NIR)	3.12	12.04
Coarsely chopped (Loss on drying and NIR)	1.23	16.01

6.3.4 Drying curve

Figure 6.8 shows the drying curve for whole peppermint leaves for the three methods (NIRS, Karl Fischer titration, and loss on drying). These were obtained by plotting water content against time as the sample dried in the oven. Although the curves obtained for NIR gauge readings and Karl Fischer results were similar and overlapping, water percentages obtained from loss on drying data were significantly higher. Differences between loss on drying and NIRS readings and between loss on drying and Karl Fischer results were 10.9% and 10.3%, respectively. In the case of chopped leaves, differences were observed to be as high as 16.01% and 17.7%. Figure 6.9 shows the drying curve of coarsely chopped peppermint leaves for the three methods. Because of the large differences observed, it can be suggested that loss on drying is not an ideal method for the analysis of water content alone. Other materials present in the leaves (e.g. volatile oils and sugars), which may also be lost in the drying process could account for the

apparently larger values. Chopped leaves may have produced higher values in loss on drying due to peppermint oil being lost more readily.

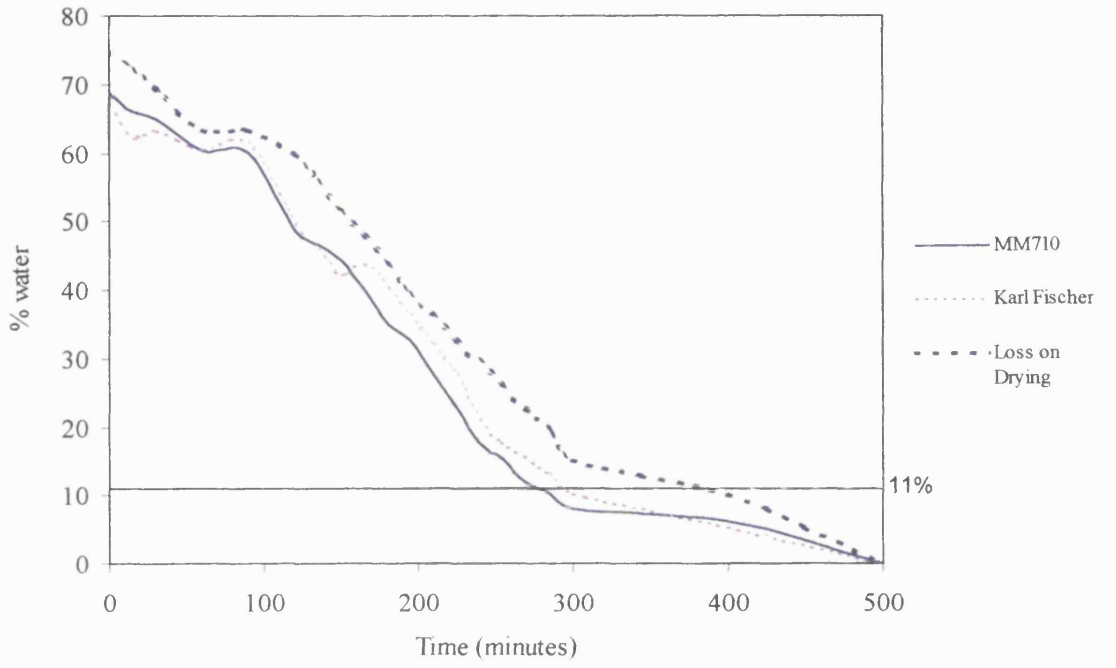


Figure 6.8. Drying curve for whole peppermint leaves in an oven at 35°C over time

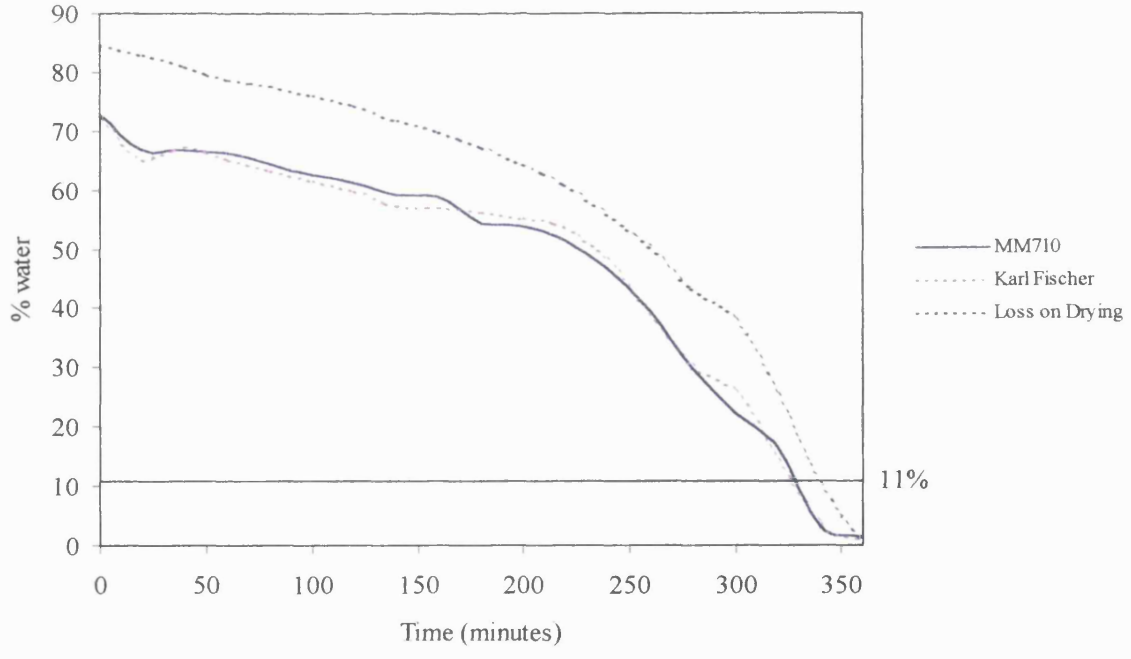


Figure 6.9. Drying curve for coarsely chopped leaves in an oven at 35°C over time

6.4 General application

Table 6.2 shows values obtained using the MM710 gauge and Karl Fischer titration for ten other leaf samples. It can be seen that results between the two methods are highly comparable, suggesting that they are applicable to most types of leaves.

The transferability of the method to a wide variety of leaf samples can possibly be explained by the fact that the water peak in an NIR spectrum of a leaf is very strong in comparison with the absorption of other components. Thus, when looking at absorbances of samples ranging from 80% to 11% water, the water peaks are still by far the strongest ones.

Table 6.2. Moisture contents for ten different leaf samples using the MM710 NIR gauge and Karl Fischer titration

Sample	% water MM710	% water Karl Fischer titration
Digitalis lanata (dried)	2.68	3.05
Digitalis ambigua (dried)	2.10	1.97
Digitalis mertonensis (dried)	1.06	1.46
Digitalis lutea (dried)	1.25	1.42
Coltsfoot (dried)	6.07	5.62
Eyebright (dried)	4.31	4.66
Coriander (dried)	1.28	1.36
Coriander (fresh)	79.4	78.7
Tarragon (fresh)	87.6	86.4
Thyme (fresh)	75.2	74.8

6.5 Rapid Content Analyser Vs MM710

When comparing the NIR spectrometers, it must be remembered that while the 6500 RCA is a laboratory instrument, the MM710 is a process instrument. For precise values, the 6500 RCA may have its advantages in that laboratory procedures can be designed for specific purposes. In addition, it produces whole spectra that can be analysed, and may potentially be used to simultaneously measure various components (e.g. water, sugar, and starch). However, the MM710 has its advantages in that it has been designed for use *in situ*, and could be used in the production process to monitor moisture levels until the desired moisture level is reached.

6.9 Conclusion

It can be concluded that NIRS can be used for the real-time monitoring of the drying process of peppermint leaves and therefore probably of most other leaf materials. Although the MM710 gauge was initially pre-calibrated for water content in tea leaves (MM710 Users guide), it appeared to be successful in a variety of leaf samples, suggesting that the instrument has a wide applicability. NIR gauge results were highly comparable to Karl Fischer and loss on drying results, producing good linearity for data ranging from 70% to 0% water. Specificity of the method for water was shown by the NIR spectra and by the correlation between Karl Fischer and NIR data. Using NIRS to control the drying procedure has the added advantage in that it can be done in a matter of seconds

and, unlike Karl Fischer titrations, is non-destructive. It is also easy to run and, due to the robustness and portability of the gauge instrument, it can be carried out on-line, allowing the moisture content of a sample to be measured continuously at regular intervals while a production process (e.g. drying) is being carried out. Due to the relatively straightforward operation of the instrument, the MM710 gauge also has its advantages in that it is easy to train staff to carry out measurements in the production environment.

Chapter 7: An Investigation into Some Leaf Components Using Near-Infrared Spectroscopy

7.1 Introduction

The study of leaf anatomy shows that a basic structural pattern exists which allows the presence of a leaf to be detected in a powdered sample. Detailed anatomical characteristics will ultimately allow the identification of the genus and species of a leaf under investigation. Knowledge of the diagnostic characteristics of any leaf permits the detection of contaminants and substitutes (Evans 1989j).

A leaf (Figure 7.1) is built up of a protective epidermis, a paranchymous mesophyll, and a vascular system. Factors such as the shape, wall structure and size of the epidermal cells, and the form, distribution and abundance of epidermal trichomes are all of diagnostic importance (Evans 1989j).

In addition to the basic leaf structure, however, leaves also contain a variety of substances including cellulose, lignin, chlorophyll, mucilage, tannin, cutin, volatile oil, calcium oxalate and calcium carbonate (Evans 1989j). These may also vary due to differences in leaf chemistry between plants, and may thus also be used for diagnostic purposes.

In this study, four of the above leaf components, which were readily available, were investigated in relation to NIR spectra of selected leaf materials. The aim was to see if any of these components could in fact be detected in any of the plant samples used, and therefore account for differences observed between species in earlier studies (e.g. chapter 4). The components studied were cellulose, tannic acid (tannin), calcium oxalate and calcium carbonate. These will be discussed briefly below:

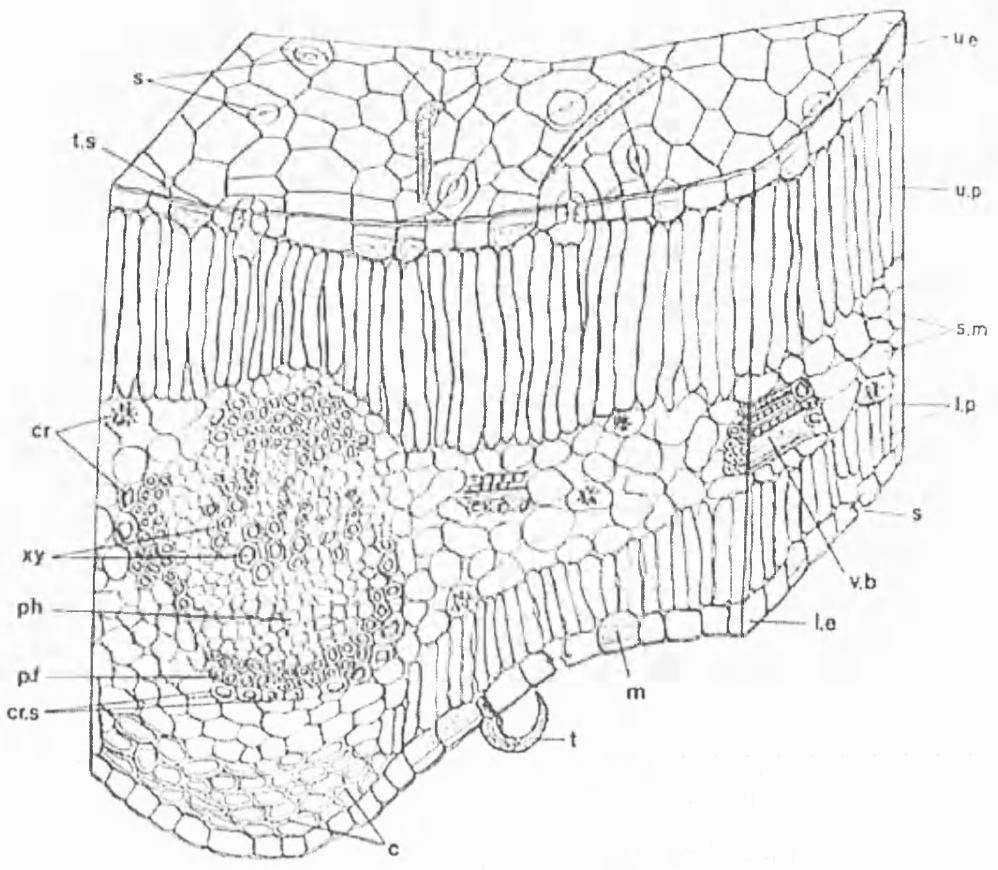


Figure 7.1. Morphology of a typical leaf. c = collenchyma; cr = calcium oxalate crystals; cr.s = crystal sheath; l.e = lower epidermis; l.p = lower palisade; m = mucilage cell; ph = phloem; p.f = pericyclic fibre; s = stomata; s.m = spongy mesophyll; t = trichome; t.s = trichome scar; u.e = upper epidermis; u.p = upper palisade; v.b = vascular bundle; xy = xylem vessels. Taken from Evans 1989.

7.1.1 Cellulose

Carbohydrates are among the first products to arise as a result of photosynthesis, and constitute a large proportion of the plant biomass. As cellulose, they make up a large percentage of the rigid cellular framework (Evans 1989k). In the cell wall, the cellulose is deposited in the form of fibrils that are visible through a microscope. The fibrils are composed of smaller microfibrils, which are themselves bundles of cellulose molecules held together by weak bonds (Evans 1989l). Cellulose molecules consist of long chains of glucose units and their structure is reflected in the macro form of the fibril (Evans 1989l) (Figure 7.2). The weight of a dry plant leaf would be expected to contain about 40% cellulose. (Kokaly and Clark, 1999).

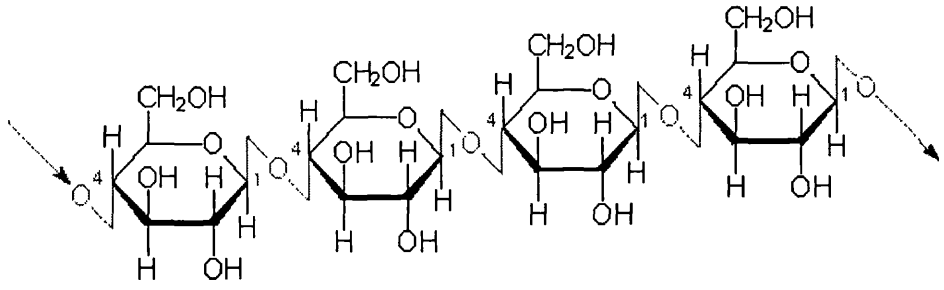


Figure 7.2. Chemical structure of cellulose

7.1.2 Tannins

Tannins (tannic acid) ($C_{14}H_{10}O_9$) are widely distributed in plants and occur in solution in the cell sap, often in distinct vacuoles. As tannins are soluble in water and alcohol, samples must be cut dry if one desires to study the distribution of tannins in a plant material (Evans 1989m). Tannins are usually found in greatest quantity in dead or dying cells, and come from two major groups - the hydrolysable tannins (which may be hydrolysed by acids or enzymes such as

tannase) and the condensed tannins (which are more resistant to breakage than hydrolysable tannins (Evans 1989n). Some plants, for example, tea (*Camellia sinensis*), can contain both hydrolysable and condensed tannins (Evans 1989n). In general, tea can be expected to contain 10-24% tannin (Evans, 1989o).

7.1.3 Calcium oxalate

Oxalic acid occurs very rarely in the free state in plants, but is extremely common as its calcium salt (CaC_2O_4) in the form of crystals. The most common crystalline forms include prisms, rosettes, single acicular crystals, bundles of acicular crystals and microspheroidal or sandy crystals. The cells that contain calcium oxalate may differ from those surrounding them in size, form, or contents, and are referred to as idioblasts. Calcium oxalate may be of considerable diagnostic importance. For example, belladonna can be distinguished by its sandy crystals, stramonium and senna by its cluster crystals, and henbane by its single and twin prisms. Calcium oxalate is usually present to the extent of 1% in plants (Evans 1989m).

7.1.4 Calcium carbonate

Calcium carbonate ($CaCO_3$) may be found embedded in, or incrusted in, the cell walls, and concretions of calcium carbonate formed on outgrowths of the cell wall are termed cystoliths (Evans 1989m). Well-formed cystoliths are seen in the enlarged upper epidermal cells and in the clothing hairs of the lower epidermis of the leaf of *Cannabis sativa* and related plants (Evans 1989m). The amount of calcium carbonate found in plant leaves was not available in the literature.

7.2 Materials and methods

7.2.1 Leaf components

Tannic acid, calcium carbonate and calcium oxalate were obtained from Fluka (Dorset, U.K), while cellulose was obtained from Merck (Dorset, UK). All materials were placed in 10 mm glass vials and placed in the Rapid Content Analyser™ and scanned. 12 sample spectra were obtained for every sample, and to verify reproducibility, the vial was shaken and tapped between each measurement.

7.2.2 Leaf samples

Samples of tea (Traditional Afternoon) were obtained from a local supermarket. Dried samples of belladonna, stramonium, senna, henbane and cannabis leaf were obtained from the pharmacognosy archives at The School of Pharmacy, University of London.

All samples were powdered using a mortar and pestle and placed in 10 mm diameter glass vials (Waters). They were then scanned in the same way as for the leaf components.

7.2.3 Data analysis

All spectra were transported into FOSS Vision® software. Here, the appearance of the plant spectra in relation to those of the chemicals were noted, in both their sample spectra form and as SNV-transformed, second derivatives. In addition, similarities between the plant materials and each of the chemicals were investigated using the identification methods of Maximum Distance in

Wavelength Space and Correlation in Wavelength Space. Particular interest was focused on tannic acid and tea, to calcium oxalate and belladonna, stramonium, henbane and senna, and to calcium carbonate and cannabis. This was due to the fact that these plant materials are characterised by the presence of these particular components. That is, the solanaceous leaves (belladonna, stramonium, henbane) may be diagnosed by the different shapes of calcium oxalate crystals present (sandy, cluster and prism) and senna is characterised by rosette-shaped crystals. Cannabis is characterised by well-formed cystoliths (concretions of calcium carbonate) in the epidermal cells and the clothing hairs of the lower leaf epidermis (Evans 1989m). All used the wavelength range 1100-2500nm and a second derivative gap size of 10nm.

7.3 Results and discussion

7.3.1 Spectral characteristics

Figure 7.3 shows sample spectra of the four leaf components, while Figure 7.4 shows the same set of spectra after they have been SNV-corrected, 2^{nd} derivative transformed. Although SNV-2^{nd} derivative spectra showed the greatest differences between the leaf components, sample spectra were used for most of the spectral observations. This was because similarities in general shapes between the four leaf chemicals and actual leaves could be more easily pointed out in sample spectra. When looking at Figure 7.3, it is clear that while the spectra for tannic acid and cellulose share a similar general shape, the spectra of calcium oxalate and calcium carbonate are rather different, although they do share some similarity in the area

1100-1850nm. Figure 7.5 shows sample spectra of belladonna, henbane, senna and stramonium leaf in relation to the original spectrum of calcium oxalate. From this, it can be seen these plant materials share a similar spectral shape to calcium oxalate from the wavelength 1100nm up to the vicinity of 1900. It can therefore be suggested that the small amount of calcium oxalate in these leaves (about 1%) can possibly be detected to some extent in the NIR spectra. Figure 7.6 shows the sample spectrum of cannabis leaf in relation to that of calcium carbonate. Again, although these spectra appear quite different, they do share some degree of similarity in shape in the area of 1100-1560nm. This pattern was also observed when calcium carbonate was compared to belladonna, henbane, senna, and stramonium leaf (Figure 7.7). In particular, it could be suspected that the peak observed in all these samples in the vicinity of 1520nm could be to some extent due to this material, as they both peaked in that area in the same way as the leaves. Figure 7.8 shows the sample spectrum of tannic acid in relation to that of tea. From this, it can be seen that these spectra also show some similarity up to the vicinity of 2200nm, suggesting that tannic acid does contribute to some extent to the appearance of the spectrum of tea, which is known to be rich in tannins (10-24%).

Figure 7.9 shows the sample spectrum of cellulose with those of all the leaf materials investigated. It is clear from this that the spectrum of cellulose is very similar to those of the other leaf materials, excepting a few peaks in the area of 2100 and 2280nm that were undetected in the leaves. This observation is not

entirely surprising, as a significant amount (40%) of plant material is made up of cellulose (Kokaly and Clark, 1999).

In a final investigation in the NIR spectra, a combined NIR spectrum of the four leaf components was obtained by averaging all the spectra together. When comparing this resulting spectrum with the leaf materials, it can be seen that the combined spectrum is very much like the spectra of the plants, showing peaks and troughs in a very similar manner to the leaves (Figure 7.10). This might suggest that indeed, spectra of plant materials are contributed to some extent by all these leaf components.

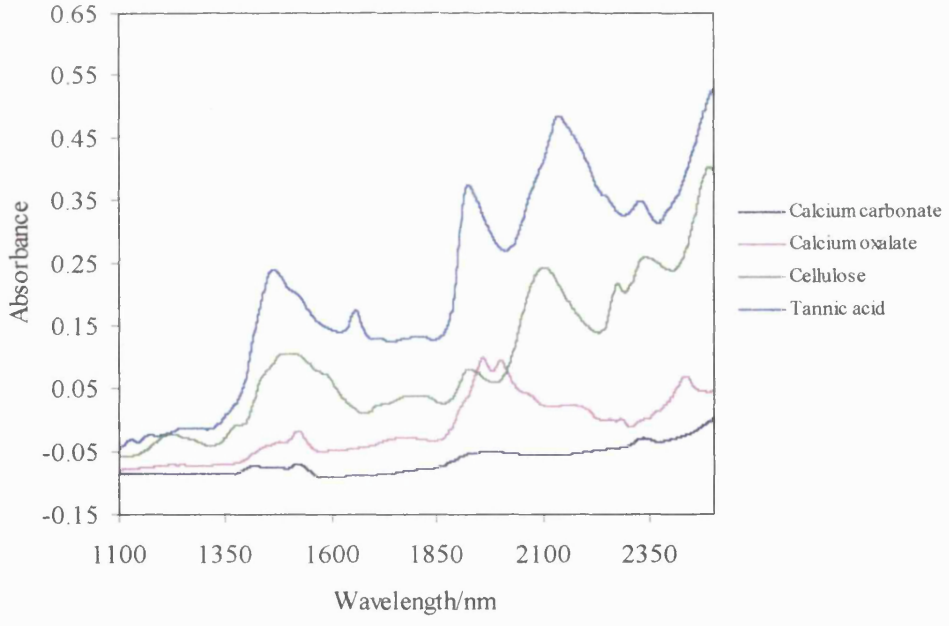


Figure 7.3. Spectra of four leaf components

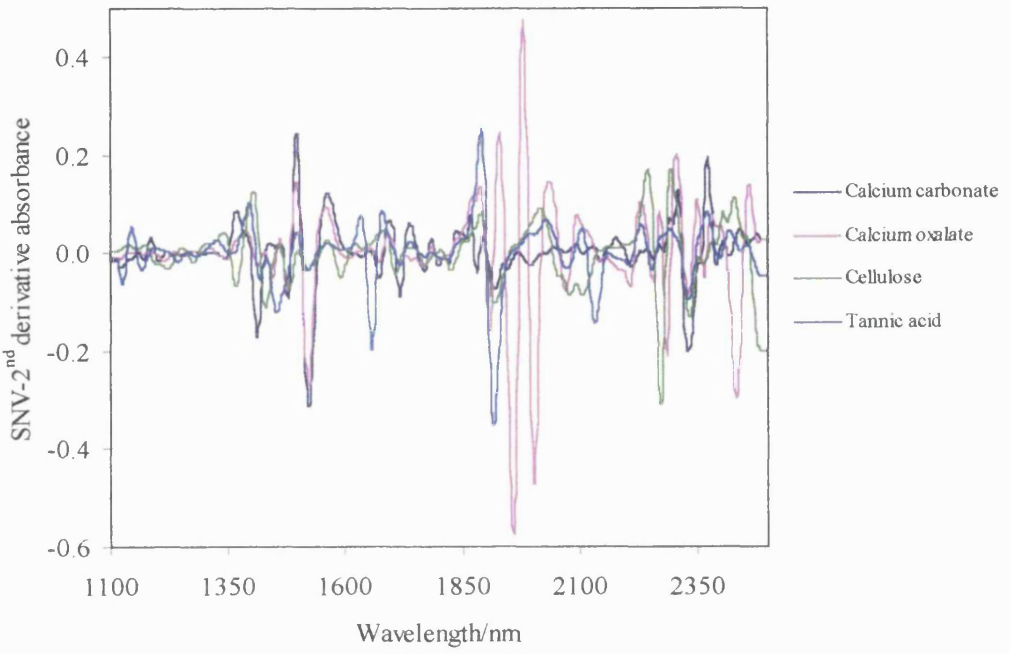


Figure 7.4. SNV-transformed, 2nd derivative spectra of four leaf components

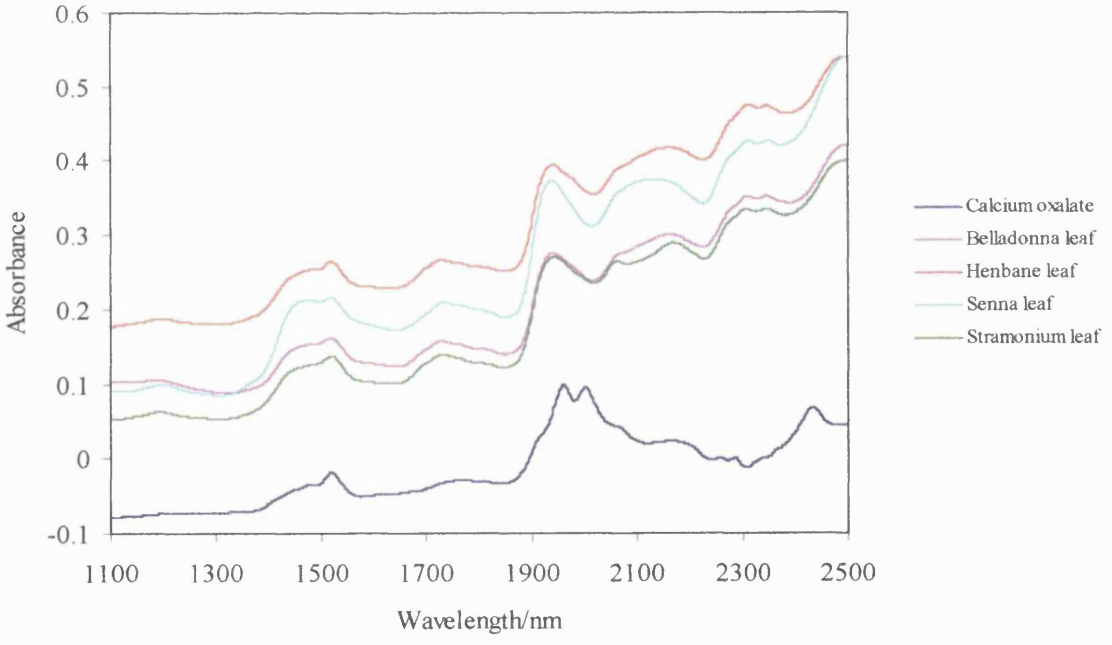


Figure 7.5. Spectra of four leaf materials and calcium oxalate

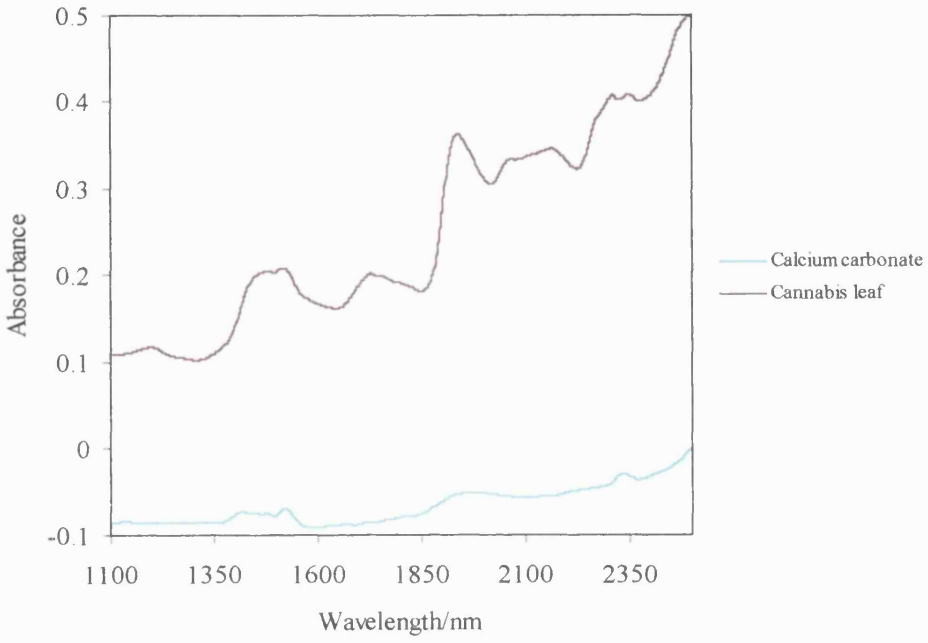


Figure 7.6. Spectra of calcium carbonate and cannabis leaf

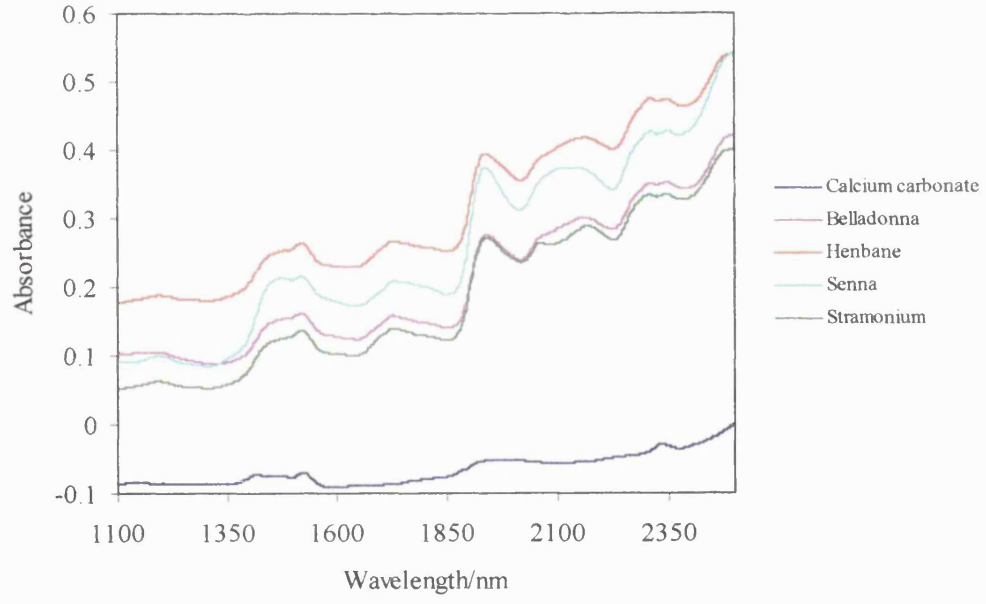


Figure 7.7. Spectra of 4 leaves and calcium carbonate

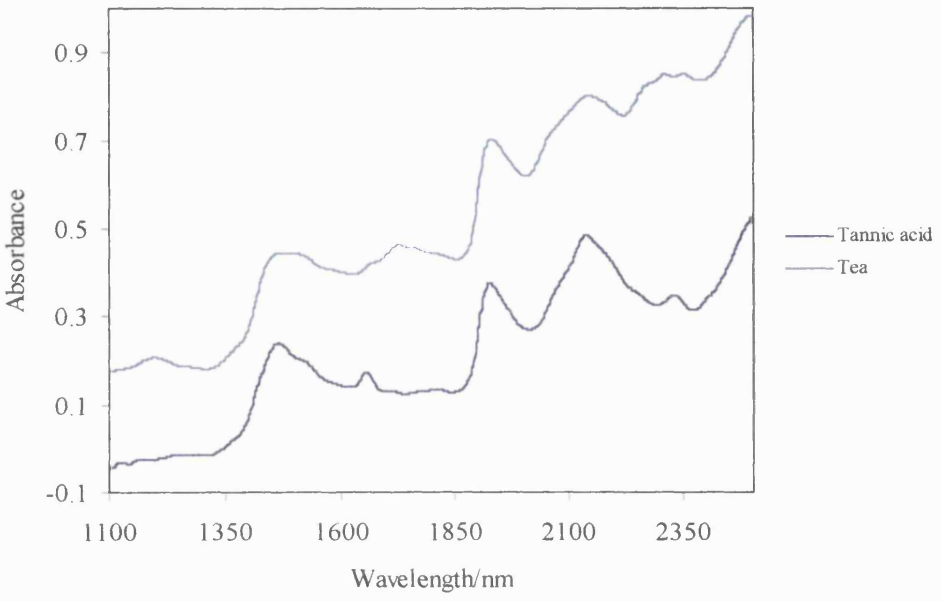


Figure 7.8. Spectra of tannic acid and tea

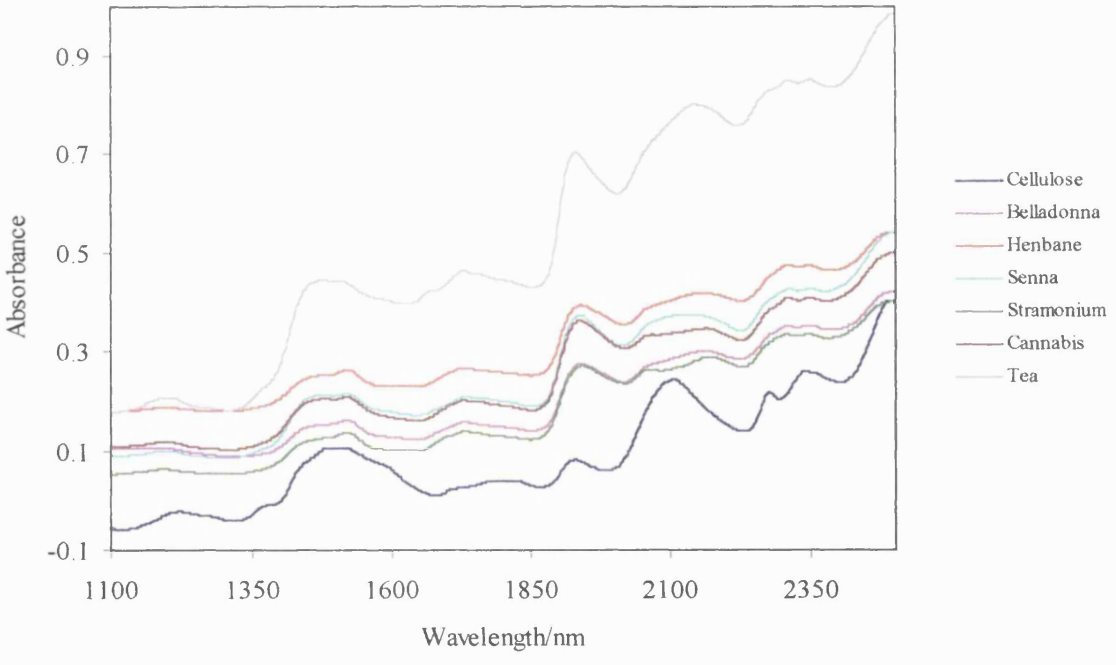


Figure 7.9. Spectra of cellulose and 6 leaf materials

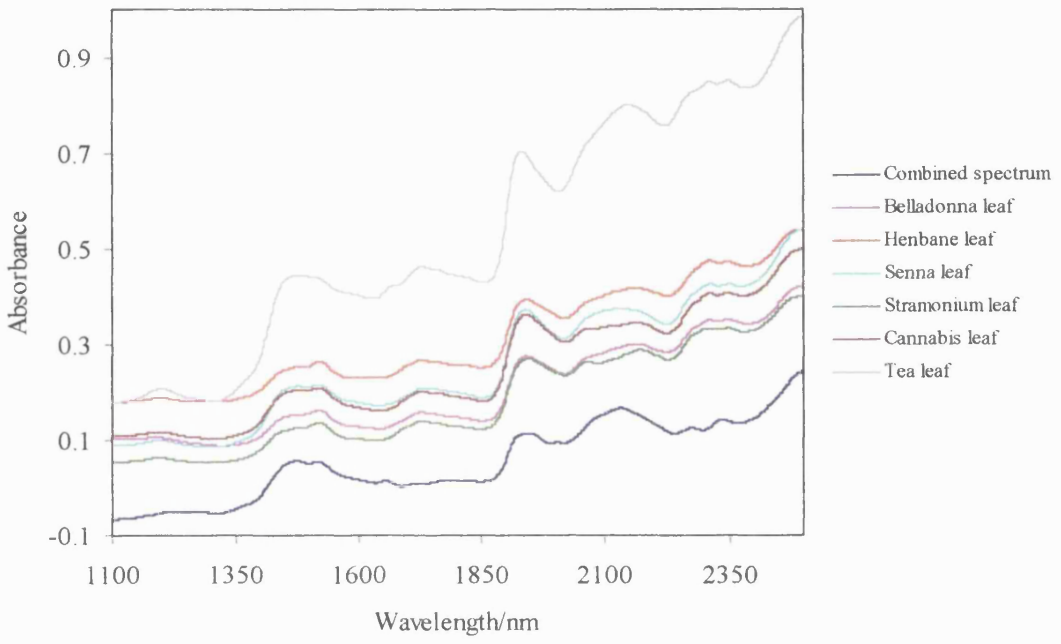


Figure 7.10. Spectra of leaf materials and combined spectrum of leaf components

7.3.2 Vision® analysis

7.3.2.1 Maximum Distance in Wavelength Space

Table 7.1 shows match values obtained when using the Maximum Distance in Wavelength Space method on SNV-corrected, 2nd derivative transformed spectra between the leaf chemicals and the leaf samples. From this, it can be seen that all results were extremely high, showing no successful identification of any of the materials in the leaves. However, a general pattern can be seen even in these results. The best (i.e. lowest) results were obtained when the leaves were tested against calcium carbonate, with values ranging from 174 for henbane leaves to 266 for stramonium leaves. The next best results were seen in the test against calcium oxalate, ranging from 292 for cannabis to 393 for belladonna. With the exception

of tea, cellulose produced results ranging from 548 for henbane to 648 for cannabis. Again with the exception of tea, tannic acid proved to be so unsuccessful in this method, that it produced only an error message when a comparison was attempted with it and the leaf samples. In tea, it did work and produced a high value of 846 that was in fact lower than that produced between tea and cellulose. This may be explained due to the fact that tea is known to be rich in tannins (20% in black tea), and thus tannic acid was detectable in the sample of tea. As was seen in Figure 7.8, the spectrum of tea and tannic acid did produce some similarities. However, although a pattern was observed in the values, results were not conclusive. For instance, when looking at calcium oxalate, calcium carbonate, and cellulose, it was seen that cellulose, which is abundant in all leaf materials, produced the worst results in all the samples when it would have been expected to produce the best match values. In addition, those plants that were known to be abundant in calcium oxalate (belladonna, henbane, senna and stramonium) produced worse results than those that were known to be not as abundant. In the same way, cannabis did not produce the expected lowest result when compared with calcium carbonate, although it was the second lowest. The "combined" spectrum could not be used in this method, as Vision® software does not allow the combination of individual spectra in analysis.

7.3.2.2 Correlation in Wavelength Space

Again, a basic pattern was observed when using the Correlation in Wavelength Space method (Table 7.2). Calcium oxalate produced the worst correlation values overall, ranging from 0.035 in tea to 0.188 in stramonium. The next lowest results

were observed in the case of calcium carbonate, ranging from 0.209 in cannabis to 0.319 in henbane. However, tea produced the value 0.116 that was in fact lower than that between stramonium and calcium oxalate. The second highest results were obtained in cellulose, ranging from 0.485 in tea to 0.530 in senna leaf. Surprisingly, tannic acid, which proved so difficult in Maximum Distance in Wavelength Space, produced the best results, ranging from 0.614 in belladonna leaf to 0.791 in tea. Again, tea, which is rich in tannins, gave the highest results in this category. However, as in the previous method, although a basic pattern was observed, results were inconclusive. Although the four leaves that are rich in calcium oxalate did produce the four best results in that category, the same was not the case in calcium carbonate and cannabis. Tannic acid, which produced a high number of errors in Maximum Distance in Wavelength Space, produced the best results overall in this method. The fairly similar results between cellulose and tannic acid in this case might also be supported by the fact that the spectra of these two were quite alike (Figure 7.3).

Table 7.1. Maximum Distance in Wavelength Space match values for four leaf components and 6 leaf materials

	Calcium oxalate	Calcium Carbonate	Cellulose	Tannic acid
Belladonna	392.9	219.5	641.4	Error
Henbane	399.0	174.0	548.2	Error
Senna	329.9	237.5	624.8	Error
Stramonium	325.1	266.1	561.1	Error
Cannabis	291.6	205.0	648.5	Error
Tea	304.9	248.8	873.2	845.5

Figures are average values

Table 7.2. Correlation in Wavelength Space values for four leaf components and 6 leaf materials

	Calcium oxalate	Calcium Carbonate	Cellulose	Tannic acid
Belladonna	0.073	0.254	0.523	0.614
Henbane	0.065	0.319	0.507	0.642
Senna	0.068	0.222	0.530	0.673
Stramonium	0.188	0.278	0.492	0.632
Cannabis	0.057	0.209	0.519	0.642
Tea	0.035	0.116	0.485	0.791

Figures are average values

7.4 Cellulose

Taking into account that much of the plant cell (40%) consists of cellulose (Kokaly and Clark, 1999), further investigations were carried out to ascertain it could be more successfully detected in NIR spectra using other means.

In one investigation, the spectrum for cellulose was converted to a de-trended SNV spectrum along with spectra for other plant materials of pharmaceutical interest. De-trended SNV absorbances at 1698 nm and 2336 nm were then plotted against each other. The resulting plot showed that the datapoint for cellulose was located within the cluster formed by the datapoints for leaf materials (Figure 7.11). Points that were not located in or near the cluster were those for non-leaf materials, including the 3 other chemicals (tannic acid, calcium carbonate and calcium oxalate) and some resins which would be expected to be significantly different to leaves in their chemical composition.

In another investigation, spectra for the same materials were SNV-corrected, 2^{nd} derivative transformed. The average distance of the spectra of these materials to the cellulose spectrum was calculated and used as data for the first axis. Data for the second axis was provided by the average value provided by the absorbances of

each of the materials. That is, the absorbances for each of the materials were added together and divided by 700 (number of values for each spectrum). The resulting plot showed that the datapoints for leaf materials were arranged in a cluster (Figure 7.12) while the datapoints for the other materials were arranged further away from this cluster.

Both these investigations show that cellulose, which forms the bulk component of leaf materials, can be detected to some extent in leaf spectra.

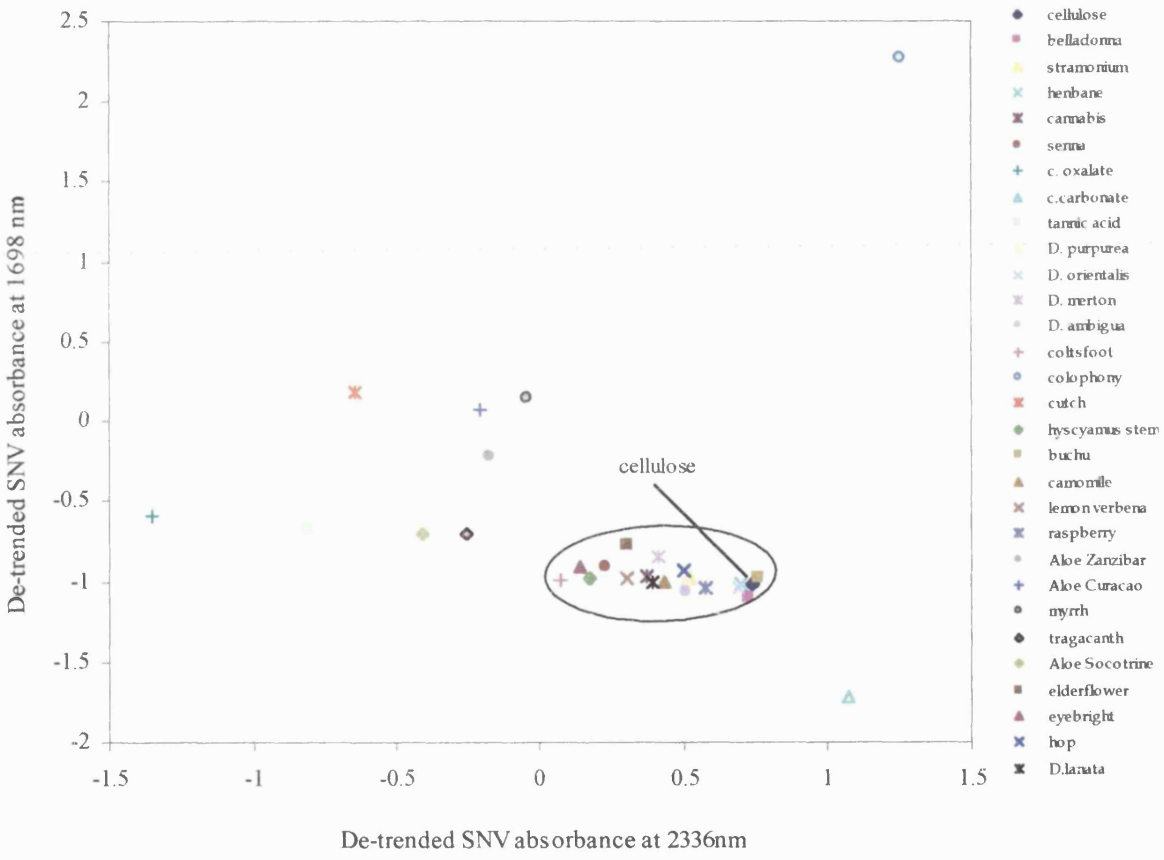


Figure 7.11. De-trended SNV absorbance at 1698 nm against de-trended SNV absorbance at 2336 nm for cellulose, calcium carbonate, calcium oxalate, tannic acid and plant materials of pharmaceutical interest. The ellipse indicates the general cluster formed by materials of leaf origin.

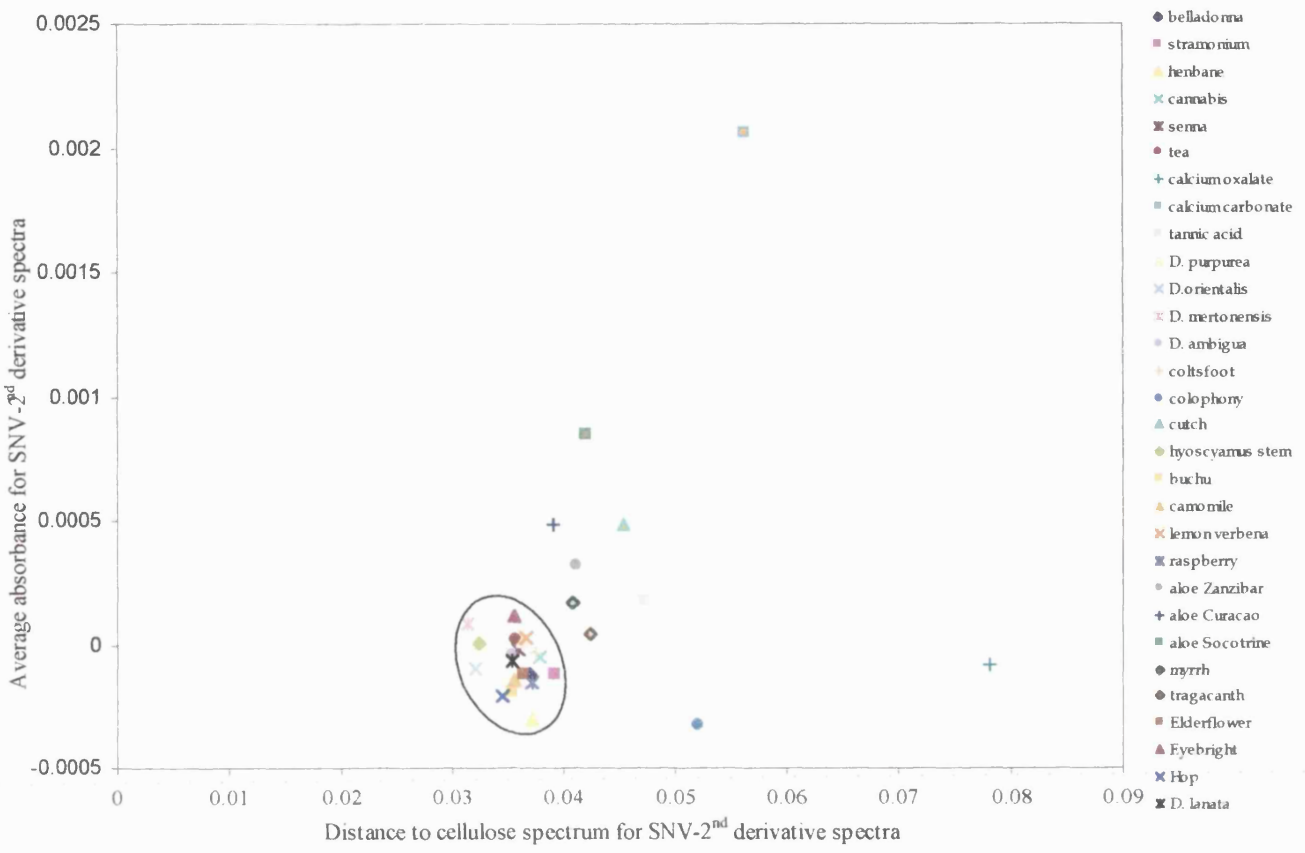


Figure 7.12. Distance to cellulose spectrum of SNV-2nd derivative spectra against average absorbance for SNV-2nd derivative spectra for calcium carbonate, calcium oxalate, tannic acid and plant materials of pharmaceutical interest. The ellipse indicates the general cluster formed by materials of leaf origin.

7.5 Quantification of cellulose and tannic acid

Figure 7.13 shows SNV-corrected, 2nd derivative spectra of cellulose and some leaf materials, while Figure 7.14 shows SNV-2nd derivative spectra of tannic acid and tea. These spectra were used to attempt quantification of cellulose in the leaf materials and of tannic acid in tea. For cellulose, its absorbance at 2406 nm was used and compared against respective absorbances of the leaf materials. That is, the strength of absorbance of the leaf materials at that wavelength in relation to the

strength of absorbance for cellulose (treated as 100%) was investigated. The wavelength 2406 nm was used in this case as all spectra showed a peak in that area (Figure 7.15). The same procedure was used for tannic acid and tea, but in this case, the wavelength 2372nm was used (Figure 7.16). Results are shown in Table 7.3. It is clear from these that the percentage values obtained are highly comparable to expected values (40% for cellulose and 24% for tannic acid, see earlier). However, although some degree of quantification is possible, results obtained from this type of analysis should be approached with caution, as it is quite likely that other materials present in the leaf samples may absorb at the same wavelengths as well. Calcium oxalate was not quantified using this method, as the percentage of expected calcium oxalate is small (1%), making this type of analysis difficult. Calciumcarbonatevas also not quantified using this technique, as the exact expected percentages were not known.

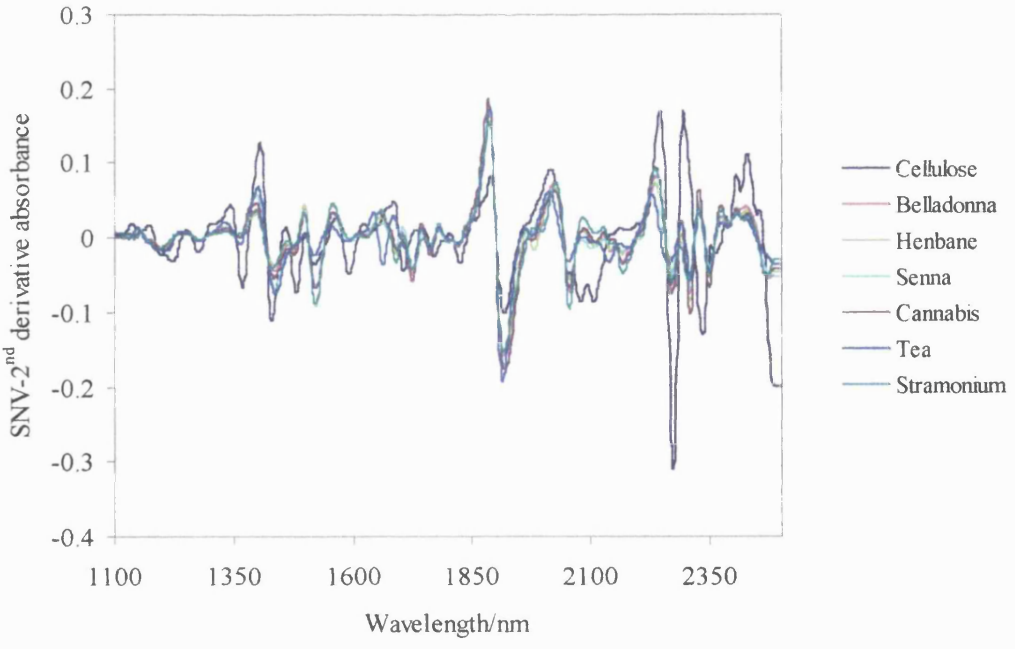


Figure 7.13. SNV-2nd derivative spectra of cellulose and 6 leaf materials

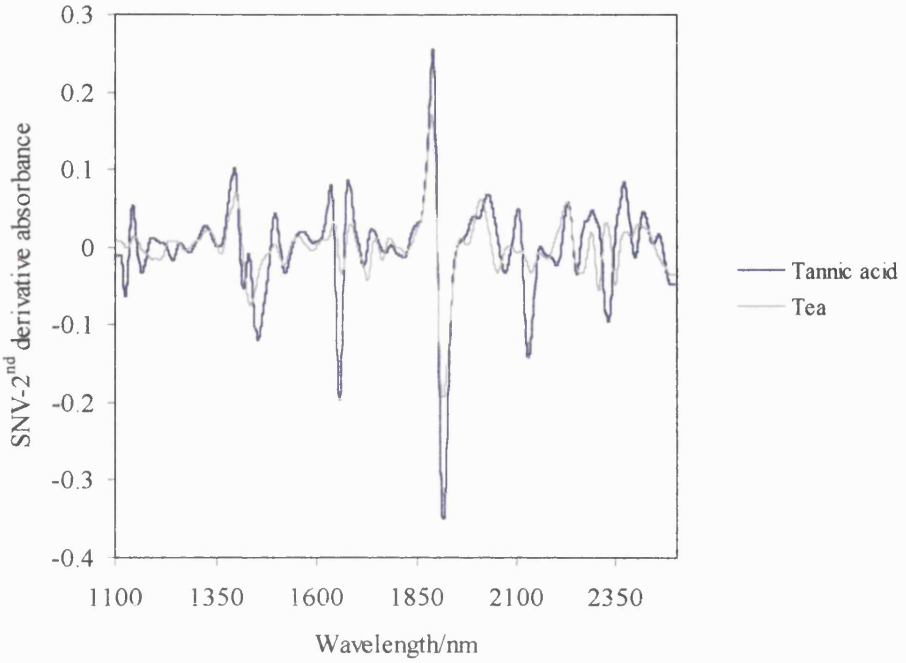


Figure 7.14. SNV-2nd derivative spectra of tannic acid and tea

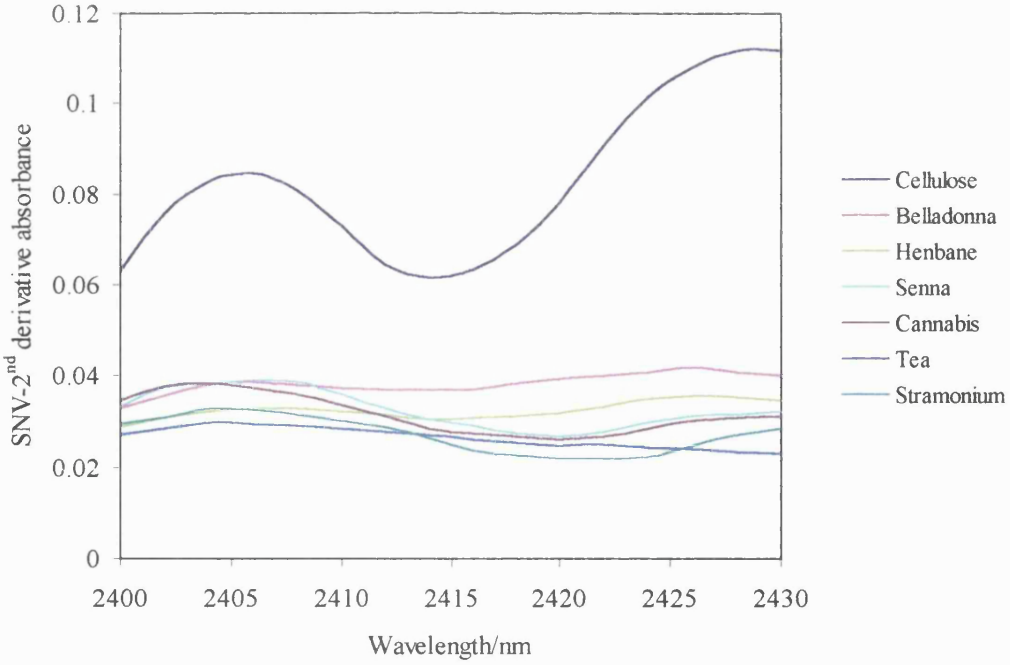


Figure 7.15. Close-up SNV-2nd derivative spectra of cellulose and 6 leaf materials (2400-2430 nm)

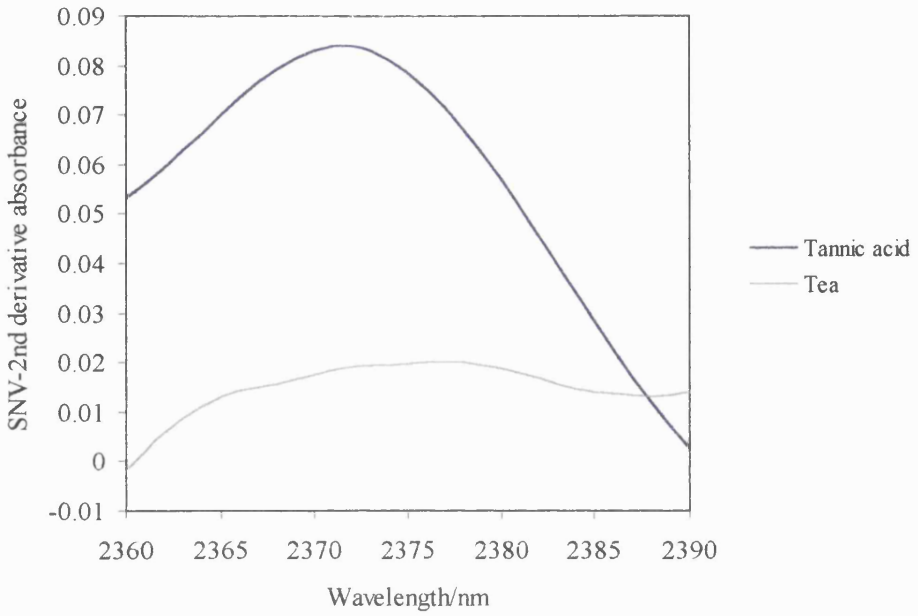


Figure 7.16. Close-up SNV-2nd derivative spectra of tannic acid and tea (2360-2390 nm)

Table 7.3. Percentages obtained for cellulose and tannic acid using spectral peaks

	Cellulose (%)	Tannic acid (%)
Belladonna	45	-
Henbane	39	-
Senna	46	-
Cannabis	44	-
Stramonium	35	-
Tea	38	23

7.6 Conclusion

When looking at the spectra of the leaf components with those of leaf samples, it is clear that these chemicals do, to some extent, contribute to the spectral characteristics of plant materials. Some clear similarities were observed between the plant spectra and the spectra of the chemical components, in particular between certain chemicals and those plants that were known to contain these in abundance. In addition, combining the spectra of the chemicals produced a spectrum that was very similar to those produced by real leaves. This would suggest that leaf spectra are made up of a combination of all sorts of chemicals, including the four under investigation. This would also suggest that variations observed between spectra of different natural products are to some extent due to different amounts and proportions of these leaf chemicals.

Using data analysis proved inconclusive, although some patterns were observed, with groups producing characteristic results. However, not all samples that were rich in a particular component produced the best results in that category, although it did appear to work to some extent, for example with tea and tannic acid. The two methods (Maximum Distance in Wavelength Space and Correlation in Wavelength Space) did not support each other in that some categories performed less well in one method as compared to the other. For example, tannic acid produced errors in the majority of samples in Maximum Distance in Wavelength Space, but produced the best results in Correlation in Wavelength Space. A similar pattern was seen in cellulose. Thus, results from data analysis were unconvincing, but it is possible to say that these chemicals do play some part in contributing to the spectra of plant

materials. That is, the methods did at least produce some results (with the exception of tannic acid and five of the leaf materials), with a pattern, indicating that a degree, however small, of similarity does exist.

Focusing on cellulose and carrying out a graphical analysis did produce some results, with leaves producing datapoints similar to that of cellulose (Figure 7.9). In addition, leaf spectra were seen to be closer to the spectrum for cellulose than spectra for non-leaf materials (Figure 7.10). This indicates that NIR can, to some extent be used to search for chemical components in leaves.

Quantification was attempted using absorbance data for cellulose and tannic acid. Results obtained were highly comparable to expected percentages. However, this method should not be interpreted as a completely accurate one, as other leaf components may also absorb at the chosen wavelengths.

A possible extension of this investigation would be to investigate further the quantification of various leaf chemicals using spectral data calibrations. In addition, it would be of interest to look at some other leaf components, such as chlorophyll and sugars.

Chapter 8: The Discrimination of Commercial Teas Using Near-Infrared Reflectance Spectroscopy

8.1 Introduction

For many years, tea factories have been producing tea that almost always finds a market, and thus monitoring for optimum quality has not necessarily been built into the production process (Melican, 2000). However, packing companies are increasingly buying tea directly from the manufacturers and specifying exactly the type of teas they require, and thus it is becoming more important for the manufacturers to be able to meet these specifications. In addition, there are international regulations controlling acceptable levels of pesticides, residues and hygiene, and thus measuring and monitoring are an important part of factory operations (Melican, 2000). For example, monitoring the temperature of the tea leaves after picking is an important part of the production process, as above 43°C, cell walls begin to break down, resulting in the release of enzymes and the oxidation of the catechins in the leaf (Melican, 2000).

Withering is a 16 to 20 hour process carried out after the leaves have been picked and transported. It involves the biochemical process in which the starch in the leaf is converted to sugar and the proteins into amino acids. During the withering process, water is also removed, reducing the moisture to about 50% to 75% (Melican, 2000). This can vary among different tea types; for example, Assams are traditionally "soft-withered" (65-75% moisture) and Ceylons are "hard withered" (50-60% moisture) (Melican, 2000). One problem encountered in the withering process is that variations in the moisture content after the process are often greater than those found in the fresh leaf coming into the factory from the fields. NIRS has been recommended as the best way to tackle the problem, as measurements can be

obtained instantaneously and on-line (Melican, 2000). It also has an advantage in that during the drying process, NIR moisture monitoring allows dryer feed rate to be automatically varied to ensure that there is a constant moisture level entering the dryer. This is important because a dryer will perform more efficiently if the tea to be dried enters the dryer at a fixed rate with a fixed moisture content (Melican, 2000).

In the tea industry, the quality of the finished tea is graded upon several criteria, including the appearance of the dry product as well as various characteristics of the final brew. This grading is done by professional tea tasters, and can be slow, labour intensive and subjective (Osborne et al 1993). Thus, in this aspect, the use of NIRS is also ideal as a more rapid and objective test for tea (Osborne et al 1993). Earlier studies compared various teas against sensory profiles and theaflavine content (Hall et al 1988). For sensory profiles, a standard deviation of 7.8 over a range of 25-74 was obtained, which was favourably comparable to a taster, and theaflavine was estimated with a standard deviation of 3.3μmoles per gram in the range 1.5-26.5 μmoles per gram (Osborne et al 1993). However, commercial blended teas would represent a much narrower range in sensory profile and theaflavin content, yet an experienced taster and the consumer can detect the presence of imbalances in the blends (Osborne et al 1993).

In this investigation, NIRS was used as a potential technique for the discrimination of five of the popular commercial blends of tea. In addition, these teas were compared to a well-known component of tea, tannic acid.

8.2 Materials and methods

8.2.1 Samples

12 brands each of the most popular blends of tea were obtained from various shops and supermarkets. The blends were Assam, Earl Grey, Traditional Afternoon, English Breakfast, and Ceylon. A sample of tannic acid was also obtained from Fluka (Dorset, UK). All samples were powdered and placed in 10 mm diameter glass vials. To verify reproducibility, each sample was scanned 12 times and shaken and tapped between scans, before the spectra were averaged.

8.2.2 Data analysis

Spectral data were transferred into FOSS Vision® software and transformed into SNV-corrected, second derivative spectra. SNV-second derivative spectra were chosen as they produced the best results for various algorithms compared to other data pretreatments (first derivative, SNV – first derivative, second derivative, third derivative, SNV-third derivative, fourth derivative, SNV-fourth derivative, de-trend, de-trend-SNV). Discrimination was then attempted using the Maximum Distance in Wavelength Space and Correlation in Wavelength Space methods. Data were also transferred to Microsoft Excel 1997 and correlation coefficients between the blends were calculated. Also in Excel 1997, two-wavelength analysis was carried out on various pretreatments and wavelength combinations. In addition, PQS was attempted on second derivative spectra. Finally, FOSS Vision® software was again used to compare the five types of tea to tannic acid using

Maximum Distance in Wavelength Space on SNV-corrected, second derivative spectra.

8.3 Results and discussion

8.3.1 Spectral characteristics

Figure 8.1 shows average sample spectra of the types of tea (average of twelve brands), while Figure 8.2 shows the same set of spectra after they have been SNV-corrected, second derivative transformed. It is clear from these spectra that although the teas are of differently named blends, their NIR spectra are extremely similar. Hence, discrimination of these blends through mere visual inspection of the spectra is an impossibility. Thus, statistical methods are needed to carry out identification.

8.3.2 Vision® Analysis

8.3.2.1 Maximum Distance in Wavelength Space

Table 8.1 shows match values obtained when using the Maximum Distance in Wavelength Space method on SNV-corrected, 2nd derivative transformed spectra between the various tea blends. Results showed that between the correctly matched blends, match values ranged from 1.65 for Assam to 2.45 for English Breakfast, while between mismatched samples, the lowest value was 9.36 for English Breakfast and Ceylon and the highest 89.5 for Earl Grey and Traditional Afternoon. The match values for the correctly matched samples were well within the range that would be expected for accurate identification of the sample using

the method. As there is a substantial gap of 6.91 between the highest match value for correctly matched samples and the lowest match value for mismatched samples, the threshold valued of 4.0 appears to be a more than generous threshold limit. Thus, it can be said that the Maximum Distance in Wavelength Space method is a successful one for the correct discrimination of the five blends of tea. Results were also reproducible – using different batches taken from the same blend produced match values that were comparable to the ones shown on Table 8.1. That is, match values for correctly matched samples ranged from 1.52 - 1.73 for Assam and 2.33 - 2.61 for English Breakfast. For mismatched samples, values ranged from 9.23 – 9.41 for English Breakfast and Ceylon to 72.4 – 90.1 for Earl Grey and Traditional Afternoon.

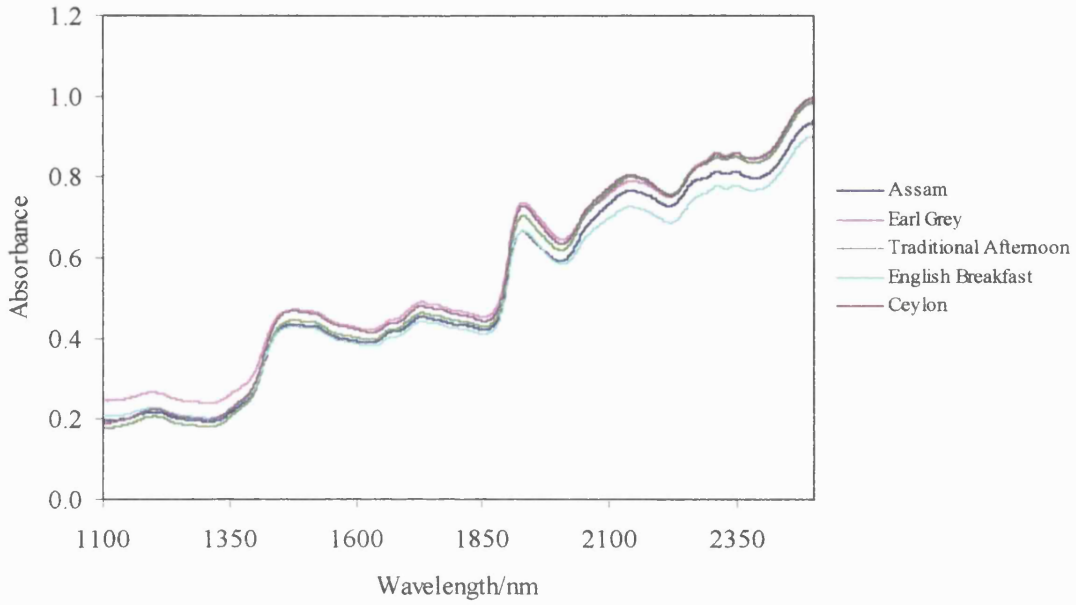


Figure 8.1. NIR spectra of five commercial blends of tea

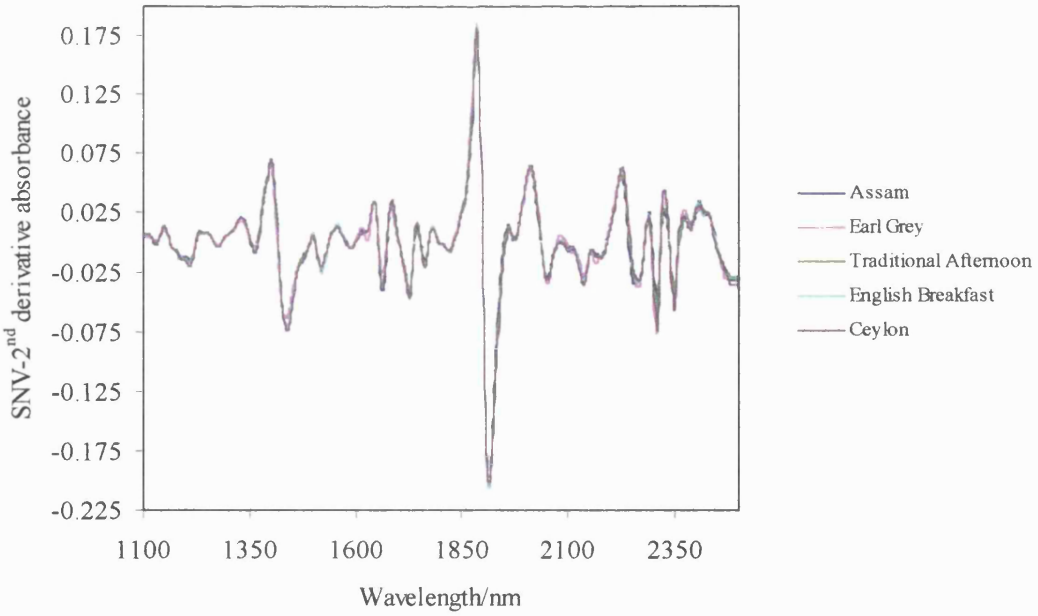


Figure 8.2. SNV-corrected, 2nd derivative transformed NIR spectra of five commercial blends of tea

Table 8.1. Maximum Distance in Wavelength Match Values for five types of tea

	Assam	Earl Grey	Traditional Afternoon	English Breakfast	Ceylon
Assam	1.65-2.01-2.15				
Earl Grey	18.4-20.4-23.7	1.70-1.99-2.41			
Traditional Afternoon	18.7-20.7-24.0	73.9-78.9-89.5	1.77-2.15-2.39		
English Breakfast	18.3-20.8-23.2	43.8-46.8-53.8	12.4-14.5-15.3	1.72-2.03-2.45	
Ceylon	17.7-18.8-19.1	28.2-30.1-34.0	10.5-11.4-16.7	9.36-10.3-12.25	1.86-2.10-2.3

Left hand number = smallest value, right hand number = largest value, underlined number = median value

8.3.2.2 Correlation in Wavelength Space

Although the Correlation in Wavelength Space method was attempted on various data pretreatments of the five tea blends, in all cases, resulting values were extremely high, with all correlation values being above 0.99. Thus, it can be said that this method is not a successful one for the discrimination of the teas. The high values reflect how similar the tea blends are.

8.3.2.3 Correlation Coefficients

Table 8.2 shows correlation coefficients between the various tea blends for SNV-corrected, 2nd derivative transformed spectra. As in the Correlation in Wavelength Space method, results showed extremely high correlations between the blends, with 0.988 being the lowest and 0.999 the highest values for mismatched samples. Thus, as the samples appear so similar, it does not appear practical or feasible to use correlation coefficients for confident discrimination of the samples in question. The two correlation methods may not be as successful compared to Maximum Distance in Wavelength Space, as while they look for similarities (correlations)

between samples, the Maximum Distance in Wavelength Space method works by looking for often minute differences between samples.

Table 8.2. Correlation coefficients between five types of tea

	Assam	Earl Grey	Traditional Afternoon	English Breakfast	Ceylon
Assam	1.0				
Earl Grey	0.988	1.0			
Traditional Afternoon	0.998	0.993	1.0		
English Breakfast	0.995	0.998	0.998	1.0	
Ceylon	0.994	0.995	0.999	0.999	1.0

Numbers are average values

8.3.2.4 Two-wavelength analysis

Figure 8.3 shows a two-wavelength plot for the five blends of tea using SNV-corrected, 2nd derivative transformed spectra and the wavelengths 1130nm and 2252. The plot shows the different teas clustering in specific areas according to their blend. However, more closely occurring clusters and an improved separation were obtained when using the wavelengths 1160nm and 2268nm on SNV-corrected, de-trended spectra (Figure 8.4). These were more easily selected by visual inspection of the spectra that showed more separation between samples in the latter data pretreatment. Although the two-wavelength method requires some degree of manual selection of the wavelengths and a trial and error method of determining ideal data pretreatments, it can be said that it is a good method for the visual separation of the different blends.

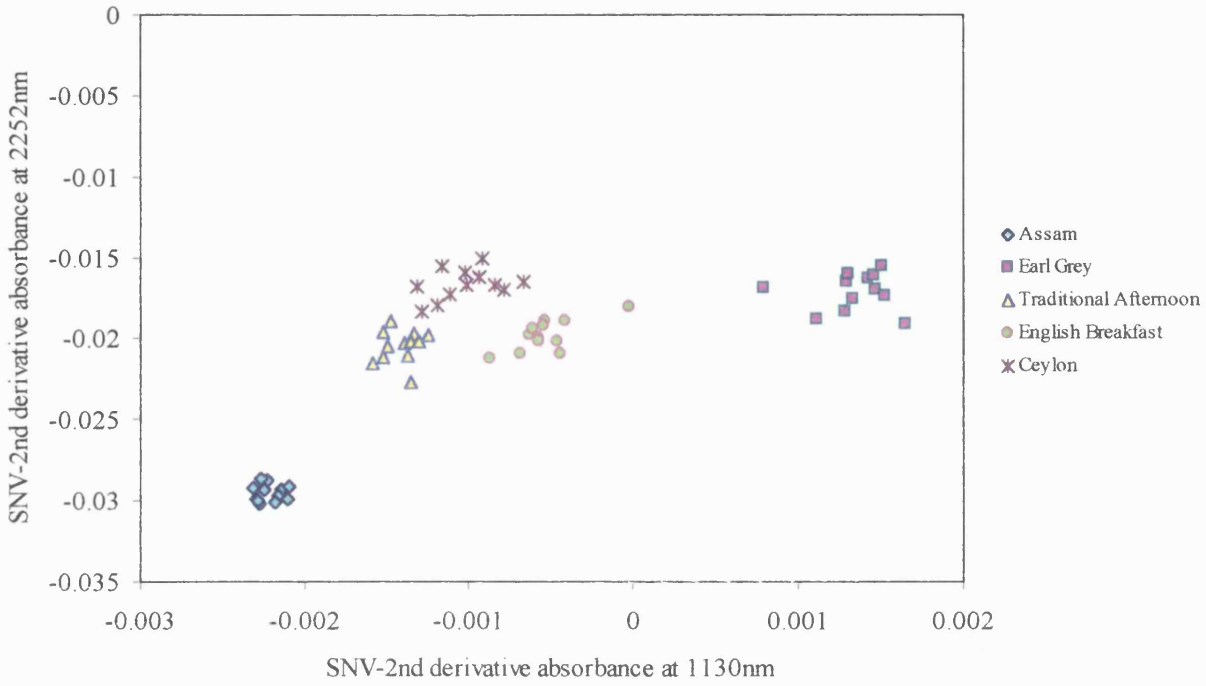


Figure 8.3. A two wavelength plot for five tea blends using SNV-2nd derivative spectra and the wavelengths 1130 and 2252nm

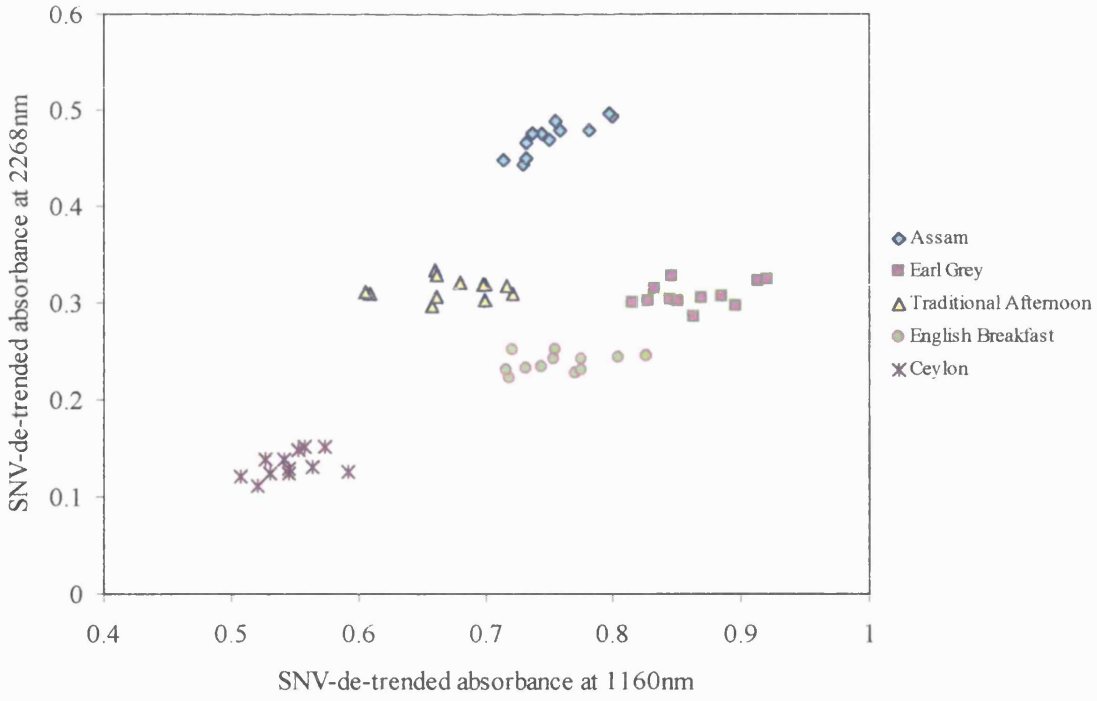


Figure 8.4. A two wavelength plot for five tea blends using SNV-de-trended spectra and the wavelengths 1160 and 2268nm

8.3.2.5 Polar Qualification System

When attempting PQS on 2nd derivative spectra using 100nm increments, the most successful separation was obtained when using the wavelength range 2050-2150 (Figure 8.5). Assam, Earl Grey, and English Breakfast appeared to show good separation on the resulting plot. However, although a clustering behaviour was observed to some extent, separation was less clear for Traditional Afternoon and Ceylon teas. Figure 8.6 shows the resulting plot obtained when carrying out the PQS method on just Traditional Afternoon and Ceylon teas. Here, when using the wavelength range 1200-1300, a clear separation was achieved between these two blends. Thus, it can be said that the PQS system is relatively successful in discriminating between the five tea blends. Using a two-step process, further separation of less well-separated groups can be achieved. This shows an advantage of the PQS method in that data can be reworked (or simultaneously worked) to give improved results.

8.3.3 Tannic acid comparison

Table 8.3 shows resulting average match values obtained when comparing the twelve different brands of five tea blends with one chemical component, tannic acid. In all cases, results were extremely high, with the lowest value being 278 for English Breakfast and the highest being 791 for Assam. Although the match values were so high, they provided interesting information in that patterns were seen according to tea type. That is, for Assam, all values, regardless of brand, were between 786 to 791, while for Earl Grey, values ranged from 278 to 293. For

Traditional Afternoon, they ranged from 572 to 605, for English Breakfast, they were from 332 to 343, and for Ceylon, they ranged from 450 to 475. Thus, these distinct patterns may suggest that tannic acid content may play some part in accounting for the differences between blends.

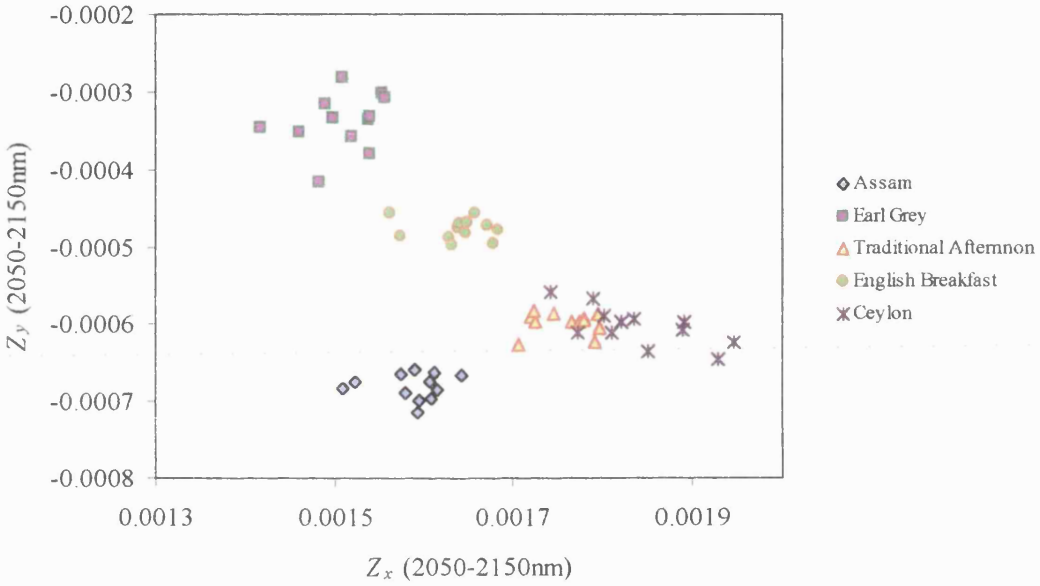


Figure 8.5. A PQS plot of five tea types using the wavelength range 2050-2150nm

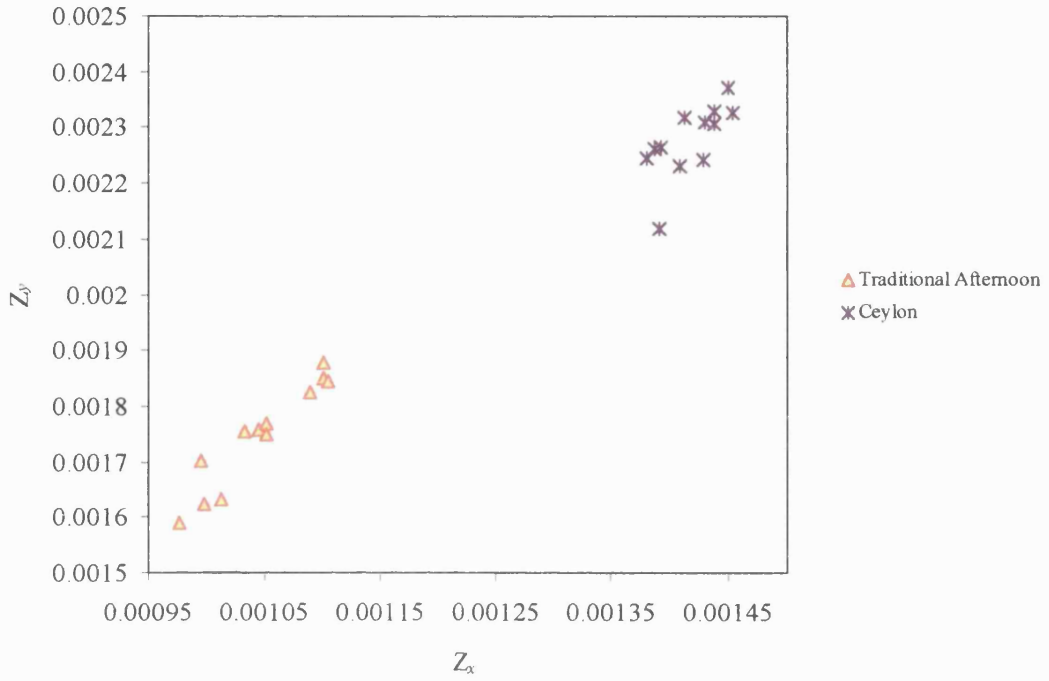


Figure 8.6. A PQS plot of Traditional Afternoon and Ceylon tea using the wavelength range 1200-1300nm

Table 8.3. Maximum Distance in Wavelength Space match values between tannic acid and five types of tea

	Assam	Earl Grey	Traditional Afternoon	English Breakfast	Ceylon
Tannic Acid	789	285	587	339	463
	789	278	572	343	450
	788	282	582	339	458
	787	281	580	338	456
	788	282	580	342	457
	786	287	592	339	466
	791	280	578	340	454
	788	284	585	335	461
	787	280	579	339	455
	788	291	603	333	473
	789	293	605	332	475
	786	288	594	337	467

Numbers are average values obtained from 12 spectra. Each row in the table denotes a different tea brand.

8.4 Conclusion

Overall, in the analysis of popular commercial tea blends, NIRS used in combination with Maximum Distance in Wavelength Space is a successful method for their discrimination. All blends could be identified as distinct from each other using this method, regardless of the brand and batch. That is, match values fell comfortably into the threshold limit for correctly matched blends. The two-wavelength method and PQS were useful in providing a visual representation of the differences between the tea types. The PQS has further advantages in that the method can be performed on spectra more than once to refine the separation of particular samples or areas of spectra.

One limitation of NIRS in the discrimination of teas is that although Maximum Distance in Wavelength Space is a successful method, other statistical comparison methods such as Correlation in Wavelength Space and Correlation Coefficients were not effective. This may be due to the fact that as the samples are so fundamentally similar, using a method which uses similarities (correlation) as a discrimination criteria is not the most suitable one. Maximum Distance in Wavelength Space, on the other hand, works by using what are often small differences between spectra.

The analysis of tannic acid in the various tea blends proved difficult, although it did appear to play some role in causing the slight differences between the samples. That is, certain trends in the Maximum Distance in Wavelength Space match values were seen depending on certain tea types.

Results were reproducible, with tests producing highly comparable values when Maximum Distance in Wavelength Space was performed on different batches of the tea blends.

Using NIRS to discriminate between various teas has its advantages in that it can reduce or eliminate the need to use professional tasters to carry out evaluation, and thus rule out human error for any discrepancies observed. In addition, it will be far more rapid than the traditional methods. In addition, as it is already used to some extent to assess moisture during the manufacturing process, it would be ideal if the same instrument(s) could be used to assess blend quality.

Chapter 9: General Conclusion

This project has shown that overall, the technique of NIRS has powerful potential in the rapid analysis of natural products and therefore of herbal medicines. This included the identification of samples down to their species, as well as identification of the part of the plant a sample came from. That is, all samples could be identified correctly to their species whether they were leaves, roots, barks, flowers, seeds or resins. Samples of different plant parts (e.g. leaves, stems, roots) could first be identified down to the species and then to the particular plant part, if required. While the NIR method was sensitive enough to detect even a minute (2% in the case of *Digitalis*) amount of a different plant part in a sample, it was also capable of identifying a sample correctly as a species even when it consisted of a mixture of different plant parts. This would in some ways be a useful approach, since where identification is the issue, it is not always the main concern as to what specific plant part an unknown sample originated from, as long as its identification is confirmed. Identification of samples was successful within a family (e.g. *Umbelliferae*) as well as within a genus (e.g. *Digitalis*).

In addition to the identification of samples, it was also found that NIRS could be used to successfully categorise samples of the same plant species according to their geographical origin. This was particularly successful in the case of *Cannabis sativa.* This aspect of identification is of major importance especially in forensic science, where the country of origin of a suspect material is often required quickly and accurately.

NIRS was also successful as a tool for measuring the moisture content of a sample. When used against traditional techniques (e.g. Karl Fischer, loss on drying), it

produced highly comparable results with the added advantage of being quick, easy, and non-destructive.

While the chemical basis for the differentiation of samples by NIRS is still rather ambiguous, it was possible to see that various plant chemicals do, to some extent, contribute to the NIR spectra of plant materials. Data analysis in this instance proved difficult, but certain patterns were observed, with groups producing characteristic results. While NIR spectra do not give clear-cut peaks representing specific components, it was useful to see that the minute differences between natural products could be due to variations in their chemical composition. Some clear clustering patterns were observed when leaf materials were compared with cellulose.

The most successful data analysis method for this project was the Maximum Distance in Wavelength Space method on SNV-corrected, 2^{nd} derivative transformed spectra using a threshold match value of 4. Most samples could be characterised correctly using this procedure, and it was possible to set up routine analysis procedures to identify samples. Results were reproducible, with highly comparable results being obtained when experiments were repeated. Other techniques (such as Correlation in Wavelength Space, correlation coefficients, Residual Variance in Principal Components Space) were successful to some extent, although there were limitations. For example, in analyses using correlations, there was the problem in that often, samples, even of varying species, were so similar that thresholds set had to be unfeasibly high.

Visual methods of differentiation were also explored. It was found that one simple method to visually separate samples was to plot absorbances at two wavelengths against each other. Although effective, this method had the disadvantage in that the most ideal data pre-treatment methods and wavelengths varied according to plant type and thus there were no clear-cut parameters to follow. Another similar technique that was explored was the use of the Polar Qualification System. By using relatively simple mathematics, it was possible to separate samples according to type without the need for randomly selecting two wavelengths, as was the case in the previous procedure. Both methods are effective visual techniques that could also be used for identification procedures, using the *k*-NN method. To some extent, PQS has its advantages over the 2-wavelength technique in that there is no need to have to search for two suitable wavelengths or to establish the best data pretreatment methods. Using small sets of spectra, it is possible to draw preliminary conclusions as neither reference spectra or large libraries are required (van der Vlies 1995). In addition, it is also possible to rework data to get optimum results.

A suitable identification procedure for natural products would potentially involve the following:

1) Obtain spectra and pre-treat to SNV-2nd derivative.
2) Use Maximum Distance in Wavelength Space against library spectra.
3) Select the natural product(s) that produced match values less than 4.
4) Allocate possible identities for the unknown sample.
5) Confirm using a visual technique (2-wavelength plot) or PQS.

Overall, NIRS for the analysis of natural products has its advantages over traditional techniques in that it is relatively easy to use, rapid, and requires little or no sample preparation or harmful reagents. While traditional techniques can be time-consuming, NIR measurements can be made in a matter of seconds. With the development of efficient computer programs, NIRS could be a powerful tool for analysis. In particular, it would be of great forensic use, where often, time is of importance when it comes to the identification of suspect samples seized at customs or crime scenes. In addition, although the use of herbal medicines is becoming increasingly popular, there are as yet very few quality control procedures carried out on them, and often, they may be contaminated. In this aspect, NIRS would be useful as a rapid and non-destructive method that would identify a sample as well as test for adulteration.

REFERENCES

Adams, M.J. (1995) Feature selection and extraction. *In* Chemometrics in Analytical Spectroscopy. Royal Society of Chemistry, Cambridge, U.K.

American Society of Testing and Materials. Standard definitions of terms and symbols relating to molecular spectroscopy. **14.01:** Standard E131-190

Banwell, C.N. (1966) Fundamentals of Molecular Spectroscopy. McGraw-Hill Limited, Berkshire, U.K.

Barnes, R. J., Dhanoa, M.S., Lister, S. J. (1989) Standard normal variate transformation and de-trending of near-infrared reflectance spectra. *Applied Spectroscopy* **43**: 772-777

Batten, G.D., Flinn, P.C., Welsh, L.A., Blakeney, A.B. (Eds.), NIR Spectroscopy Group, Melbourne, Australia, pp. 181-185

Benson, I.B. (1996) Not just moisture: a review of some commercially successful near infrared applications. *In* Near-Infrared Spectroscopy: The Future Waves. Davies, A.M.C and Williams, P. (eds.). NIR Publications, Chichester, U.K.

Bisset, N.G. and Wichtl, M. (2001a) Herbal Drugs and Phytopharmaceuticals, 2nd edn. Medpharm GmbH Scientific Publishers, Stuttgart, Germany, p. 159

Bisset, N.G. and Wichtl, M. (2001b) Herbal Drugs and Phytopharmaceuticals, 2nd edn. Medpharm GmbH Scientific Publishers, Stuttgart, Germany, pp.73-75

Blakeney, A.B., Welsh, L.A., Sharman, J.P., Ronalds, J.A., and Reece, J.E. (1994) Rice moisture analysis by NIT. *In* Leaping Ahead with Near-Infrared Spectroscopy.

Blanco, M., Coello, J., Iturriaga, H., Maspoch, S. and de la Pezuela, C. (1994) Control analysis of a pharmaceutical preparation by near-infrared reflectance spectroscopy. A comparative study of a spinning module and fibre optic probe. *Anal Chim. Acta.* **298:** 183-191

Blanco, M., Coello, J., Iturriaga, H., Maspoch, S., de la Pezuela, C. (1998) Near-infrared spectroscopy in the pharmaceutical industry *Analyst* **123** 135R-150R

Blanco, M., Eustaquio, A., González, J.M., Serrano, D. (2000) Identification and quantitation assays for intact tablets of two related pharmaceutical preparations by reflectance near-infrared spectroscopy; validation of the procedure. *J. Pharm. Biomed Anal.* **22**: 139-148

British Pharmacopoeia (1999) H.M. Stationery Office, London, 1999, **Vol 1:** pp505-506

British Pharmacopoeia (2000a) H.M Stationery Office, London, 2000, **Vol. 1:** pp. 541-542

British Pharmacopoeia (2000b) H.M Stationery Office, London, 2000, **Vol. 1:** pp. 1364-1365

British Pharmacopoeia (2000c) H.M Stationery Office, London, 2000, **Vol. 2:** A194

British Pharmacopoeia (2000d) H.M Stationery Office, London, 2000, **Vol. 2:** A192

British Pharmacopoeia (2000e) H.M Stationery Office, London, 2000, **Vol. 1:** p. 1190

British Pharmacopoeia (2000f) H.M Stationery Office, London, 2000, **Vol. 1:** pp. 541-542

Bromba, M.U.A., and Ziegler, H. (1981) *Anal. Chem.* **53**: 1583

Bromba, M.U.A., and Ziegler, H. (1983) *Anal. Chem.* **55**: 1299

Burns, D.A., Ciurczak, E.W. (1992) Handbook of Near-Infrared Analysis, Marcel Dekker, New York, New York, pp. 410-442

Candolfi, A., De Maesschalck, R., Jouan-Rimbaud, D., Hailey, P.A., Massart, D.L. (1999) The influence of data pre-processing in the pattern recognition of excipients near infrared spectra. *J. Pharm. Biomed. Anal.* **21:** 115-132

Chalmers, J.M. eds. (1999) Spectroscopy in Process Analysis. CRC Press, Boca Raton, Florida.

Chalmers, J.M. (2000) Spectroscopy in Process Analysis, Sheffield Academic Press Ltd., Sheffield, England.

Cho, R.K., and Lee, K.H. (1994) Use of near-infrared reflectance spectroscopy for quality evaluation of dried Korean ginseng and its extracts. *In* Leaping Ahead with Near-Infrared Spectroscopy. Batten, G.D., Flinn, P.C., Welsh, L.A., Blakeney, A.B. (Eds.), NIR Spectroscopy Group, Melbourne, Australia, pp.214-217

Corti, P., Dreassi, G.G., Corbini, G., Moggi, A., Gravina, S. (1990) Application of NIR reflectance spectroscopy to the identification of drugs derived from plants. *Int. J. Crude Drug Res.* **28:** 185-92

Cox, P.A., Balick, M.J. (1994) The ethnobotanical approach to drug discovery. *Scientific American.* June: 2-7

Davies, A.M.C. (1990) Subdivisions of the infrared region. *Applied Spectroscopy* **44**: 14A

Delwiche, S.R., Pitt, R.E., Norris, K.H. (1992) Sensitivity of near-infrared absorption to moisture content versus water activity in starch and cellulose. *Cereal Chem.* **69:**(1) 107-9

Dempster, M.A., Jones, J.A., Last, I.R., MacDonald, B.F. and Prebble, K.A. (1993) Near-infrared methods for the identification of tablets in clinical trial supplies. *J. Pharm. Biomed. Anal.* **11:**1087-1092

Dominguez, L.M., Seymour, S.K. (1992) *In* Making Light Work: Advances in Near-Infrared Spectroscopy. Murray, I. and Cowe, I.A. (eds.) Weinheim, p179

Downey, G. (1994) Grain analysis by NIR: is the harvest in? *Proceedings of NIR94.* 136-147

Downey, G. (1996) Authentication of food and food ingredients by near infrared spectroscopy. *J. Near-Infrared Spectrosc.* **4**: 47-61

Downey, G., Boussion, J., Beauchêne, D. (1994) Authentication of whole and ground coffee beans by near-infrared reflectance spectroscopy. *J. Near-Infrared Spectrosc.* **2**: 85-92

Edney, M.J., Morgan, J.E., Williams, P.C., Campbell, L.D. (1994) Analysis of feed barley by near infrared reflectance technology. *J. Near-Infrared Spectrosc.* **2**: 33-41

Evans, W.C. (1989a) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, pp. 205-206

Evans, W.C. (1989b) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, p. 440

Evans, W.C. (1989c) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, p. 437

Evans, W.C. (1989d) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, p. 486

Evans, W.C. (1989e) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, pp. 500-510

Evans, W.C. (1989f) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, pp. 65-67

Evans, W.C. (1989g) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, pp. 413-416

Evans, W.C. (1989h) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, pp. 128-130

Evans, W.C. (1989i) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, pp. 89-90

Evans, W.C. (1989j) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, pp. 23-25

Evans, W.C. (1989k) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, p. 339

Evans, W.C. (1989l) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, p. 33

Evans, W.C. (1989m) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, pp. 53-57

Evans, W.C. (1989n) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, pp. 386-388

Evans, W.C. (1989o) Trease and Evans' Pharmacognosy, 13th edn. Ballière Tindall, London, p. 630

Evans, D.G., Scotter, C.N.G., Day, L.Z., Hall, M.N. (1993) Determination of the authenticity of orange juice by discriminant analysis of near-infrared spectra. *J. Near-Infrared Spectrosc.* **1**: 33-44

Fehrmann, A., Schulz, H., Pank, F. (1998) Non-destructive NIRS measurements in caraway (*Carum carvi* L.) and fennel (*Foeniculum vulgare* Mill.) Fruits. Source unknown pp. 418-421

FOSS Vision® Software manual (1998) FOSS NIRSystems, Silver Spring, MD, U.S.A.

Galante, L.J., Brinkley, M.A., Drennen, Lodder, R.A. (1990) Near-infrared spectrometry of microorganisms in liquid pharmaceuticals. *Anal. Chem.* **62:** 2514-2521

Gemperline, P.J., Boyer, N.R. (1995) Classification of near-infrared spectra using Wavelength Distances: comparision to the Mahalanobis Distance and Residual Variance methods. *Anal. Chem.* **67:** 160-166

Gonzales, F. and Pous, R. (1995) Quality control in manufacturing process by near-infrared spectroscopy. *J. Pharm. and Biomed. Anal.* **13, No. 4/5:** 419-423

Graham, A. (1999) Teach Yourself Statistics, 2nd edn. Hodder and Stoughton, London, pp. 191-197

Hall, M.N., Robertson, A., Scotter, C.N.G. (1988) Near-infrared reflectance prediction of quality, theaflavin content and moisture content of black tea. *Food Chemistry* **27**: 61-75

Hana, M., McClure, W.F., Whitaker, T.B., White, M.W., Bahler, D.R. (1997) Applying artificial neural networks: Part II. Using near infrared data to classify tobacco types and identify native grown tobacco. *J. Near Infrared Spectroscopy* 5: 19-25

Holdcroft, A., Smith, M., Smith, B., Hodgson, H., Evan, F.J. (1997) Clinical trial experience with cannabinoids. *Pharmaceutical Sciences* 3: 546-550

Kaffka, K.J, Gyarmati, L.S. (1990) *In* R. Biston and N. Bartiaux-Thill (eds.) Proceedings of the 3rd International Conference on NIR Spectroscopy, (Agriculture Research Centre, Gembloux, Belgium, 1990) **1:** pp. 135-139

Khan, P.R., Jee, R.D., Watt, R.A., Moffat, A.C. (1997) The identification of active drugs in tablets using near-infrared spectroscopy. *Pharmaceutical Sciences*, **3** 1-7

Kokaly, R.F., Clark, R.N. (1999) Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise linear regression. U.S. Geological Survey, Denver, CO, U.S.A.

Krzanowski, W.J. (1995) The authentication of Basmati rice using near infrared spectroscopy: some further analysis. *J. Near-Infrared Spectrosc.* **3**: 111-117

Kortum, G. (1969) Reflectance Spectroscopy, Principles, Methods, Applications. Springer-verlag, Berlin

Li, W., Foulon, M., Meurens, M. (1996) Quantitative near-infrared analysis of orange juices using partial least squares method. *In* Near-Infrared Spectroscopy: The Future Waves. Davies, A.M.C and Williams, P. (eds.). NIR Publications, Chichester, U.K.

Lo, S.C. (1996) Rapid classification and blends analysis of tobacco mixtures using near-infrared and artificial neural networks. *In* Near-Infrared Spectroscopy: The Future Waves. Davies, A.M.C and Williams, P. (eds.). NIR Publications, Chichester, U.K.

Mark, H. (1992) Qualitative Discriminant Analysis *in* Handbook of Near-Infrared Analysis, Burns, D.A., and Ciurczak, E.W. eds., Marcel Dekker, New York, New York, U.S.A.

Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y., Kauffman, L. (1988) Chemometrics: A Textbook. Elsevier Science Publishers B.V., Amsterdam. pp. 395-397

McElhinney, J., Downey, G. (1999) Chemometric processing of visible and near infrared reflectance spectra for species identification in selected raw homogenised meats. *J. Near-Infrared Spectrosc.* **7**: 145-154

Martens, H., Naes, T. (1989) Multivariate Calibration, John Wiley and Sons, New York, New York, p134-135

Melican, N. (2000) Monitoring in tea production. *Tea and Coffee Trade Journal.* December 172:11

MM710 Users Manual (2000a) NDC Infrared Engineering, Malden, Essex, UK, pp. 1:3-1:11

MM710 Users Manual (2000b) NDC Infrared Engineering, Malden, Essex, UK, pp. 4:3-4:15

Moffat, A.C., Jee, R.D., Watt, R.A. (1997) New near infrared centre of excellence for Europe. *Eur. Pharm. Rev.* 2: 37-40

Morriseau, K.T., Rhodes, C.T. (1995) Pharmaceutical uses of near-infrared spectroscopy. *Drug Development and Industrial Pharmacy* **21:** 1071-1090

Newell, C.A., Anderson, L.A., Phillipson, J.D. (1996a) Herbal Medicines- A Guide for Health-care Professionals, Pharmaceutical Press, London, pp. 30-31

Newell, C.A., Anderson, L.A., Phillipson, J.D. (1996b) Herbal Medicines- A Guide for Health-care Professionals, Pharmaceutical Press, London, pp. 28-29

Newell, C.A., Anderson, L.A., Phillipson, J.D. (1996c) Herbal Medicines- A Guide for Health-care Professionals, Pharmaceutical Press, London, pp. 87-89

Notcutt W., Price, M., Chapman, G. (1997) Clinical experience with nabilone for chronic pain. *Pharmaceutical Sciences* 3: 551-555

Ollinger, J.M. and Griffiths, P.R. (1993) Effects of sample dilution and particle size/morphology on diffuse reflection spectra of carbohydrate systems in the near-infrared and mid-infrared. Part I – Single Analytes. *Appl. Spectrosc.*, **47**: 687-694

O'Neil, A.J. (2000) Multivariate Statistical Quality Control of a Pharmaceutical Manufacturing Process Using Near Infrared Spectroscopy And Imaging Microscopy. PhD Thesis, School of Pharmacy, University of London

Osborne, B.G., Fearn, T., Hindle, P.H. (1993) Practical NIR Spectroscopy with Applications in Food and Beverage Analysis, 2nd edn. Longman, Harlow

Osborne, B.G., Mertens, B., Thompson, M., Fearn, T. (1993) The authentication of Basmati rice using near-infrared spectroscopy. *J. Near-Infrared Spectrosc.* **1**: 77-83

Osborne, B.G., Mertens, B., Thompson, M., Fearn, T. (1994) Authentication of Basmati rice by near infrared transmittance spectra of individual grains. *Proceedings of NIR94*. pp. 161-163

Pertwee, R.G. (1997) Cannabis and cannabinoids: pharmacology and rationale for clinical use. *Pharmaceutical Sciences* 3: 539-545

Polunin, M. and Robbins, C. (1999a) The Natural Pharmacy, Dorling Kindersley, London p. 103

Polunin, M. and Robbins, C. (1999b) The Natural Pharmacy, Dorling Kindersley, London p. 97

Polunin, M. and Robbins, C. (1999c) The Natural Pharmacy, Dorling Kindersley, London p. 85

Rannou, H., Downey, G. (1997) Discrimination of raw pork, chicken, and turkey meat by spectroscopy in the visible, near- and mid-infrared ranges. *Analytical Communications*. **34** 401-404

Ridgway, C., Chambers, J. (1998) Detection of insects inside wheat kernels by NIR imaging. *J. Near-Infrared Spectrosc.* 6: 115-119

Ridgway, C., Chambers, J. (1999) Detection of grain weevils inside single wheat kernels by a very near-infrared two-wavelength model. *J. Near-Infrared Spectrosc.* **7**: 213-221

Samuelsson, G. (1999) Drugs of Natural Origin: A Textbook of Pharmacognosy. Swedish Pharmaceutical Press, Stockholm, Sweden. pp. 36-37

Savitsky, A., Golay, M.J.E. (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**: 1627

Simard, C., Buijs, H. (1996) An alternative to handheld fibre optic probes. *In* Near-Infrared Spectroscopy: The Future Waves. Davies, A.M.C and Williams, P. (eds.). NIR Publications, Chichester, U.K.

Sirieix, A., and Downey, G. (1993) Commercial wheatflour authentication by discriminant analysis of near infrared reflectance spectra. *J. Near-Infrared Spectrosc.* **1**: 187-197

Stuart, B., George, B., McIntyre, P. (1996) Modern Infrared Spectroscopy. John Wiley and Sons Ltd., Chichester, U.K.

Takahashi, M., Hajika, M., Igita, K., Sato, T. (1996) Rapid estimation of protein, oil and moisture content in whole-grain soybean seeds by near-infrared reflectance spectroscopy. *In* Near-Infrared Spectroscopy: The Future Waves. Davies, A.M.C and Williams, P. (eds.). NIR Publications, Chichester, U.K.

Taylor, R.A. (1994) Moisture in cotton and its contribution to fibre strength changes. *In* Leaping Ahead with Near-Infrared Spectroscopy. Batten, G.D., Flinn, P.C., Welsh, L.A., Blakeney, A.B. (Eds.), NIR Spectroscopy Group, Melbourne, Australia, pp. 460-464

United States Pharmacopoeia Chapter draft (1998), Near-Infrared Spectroscopy

Van der Vlies, C. (1996) Near-infrared spectroscopy and the regulatory hurdles. *Eur. Pharm. Rev.* **Feb:** 49-54

Van der Vlies, C., Kaffka, K.J., Plugge, W. (1995) Qualifying Pharmaceutical Substances by Fingerprinting with NIR Spectroscopy and PQS. *Pharmaceutical Technology Europe.* April edition pp.45-49

Wehling, R.L., Pierce, M.M. (1988) Determination of moisture in cheddar cheese by near-infrared spectroscopy. *J. Assoc. Off. Anal. Chem.* 7(3): 571-574

White, P. (1998) Crime Scene to Court: The Essentials of Forensic Science. Royal Society of Chemistry, Cambridge, U.K.

Wilson, N.D., Syed, A.F., Watt, R.A., Moffat, A.C. (2000) Near-infrared spectroscopic quality control of rosemary essential oil. J. *Pharm. Pharmacol.* 52(S) 241

Wilson, N.D., Watt, R.A., Moffat, A.C. (2001) J. Pharm. Pharmacol. 53: 95-102

Wilson, N.D., Ivanova, M., Watt, R.A., Moffat, A.C. (2001) The use of near-infrared spectroscopy for the determination of citral in lemon and lemongrass oils. British Phamaceutical Conference 2001 Abstract Book: 4

Woo, Y.A., Cho, C.H., Kim, H.J., Cho, J.H., Cho, K.K., Chung, S.S., Kim, S.J., Kim, J.H. (1998) Discrimination of herbal medicines according to geographical origin (Korea, China) using near-infrared reflectance spectroscopy. *Yakhak Hoeji.* 42: 359-363

Woo, Y.A, Kim, H.J., Cho, J.H., Chung, H. (1999) Discrimination of herbal medicines according to geographical origin with near infrared reflectance spectroscopy and pattern recognition techniques. *J. Pharm. and Biomed. Anal.* **21**: 407-413

LIST OF PUBLICATIONS

Kudo, M., Moffat, A.C., Watt, R.A. (1998) The rapid characterisation of natural products and herbal medicines by near-infrared spectroscopy. J. *Pharm. Pharmacol.* 50(S): 258

Kenyon, S.H., Waterfield, C.J., Asker, D.S., Kudo, M., Moss, D.W., Bates, T.E., Nicolaou, A., Gibbons, W.A., Timbrell, J.A. (1999) Effect of hydrazine upon vitamin B_{12}-dependent methionine synthase activity and the sulphur amino acid pathway in isolated rat hepatocytes. *Biochemical Pharmacology* 57 pp. 1311-1319

Kudo, M., Moffat, A.C., Watt, R.A. (1999) The use of near-infrared reflectance spectroscopy for the rapid characterisation of *Digitalis* species. *J. Pharm. Pharmacol.* 51(S): 72

Kudo, M., Moffat, A.C., Watt, R.A. (2000) The use of near-infrared spectroscopy for the identification of the geographical origins of *Cannabis sativa. J. Pharm. Pharmacol.* 52(S): 295

Kudo. M., Watt, R.A., Moffat, A.C. (2000) Rapid identification of *Digitalis purpurea* using near-infrared reflectance spectroscopy. *J. Pharm. Pharmacol.* 2000, 52: 1271-1277

Kudo, M., Moffat, A.C., Watt, R.A. (2001) Controlling the drying process of peppermint leaves using near-infrared spectroscopy. *British Pharmaceutical Conference Science Proceedings 2001.* Pharmaceutical Press, London, 2001. p.36