TRANSFERABILITY OF NEAR-INFRARED SPECTRA FOR THE IDENTIFICATION OF PHARMACEUTICALS

By

Weng Li Yoon

Thesis submitted for the degree of

Doctor of Philosophy

in the Faculty of Medicine

of the University of London

2000

The School of Pharmacy, University of London

29/39 Brunswick Square, London WC1N 1AX

With love and thanks to God and my parents

ACKNOWLEDGEMENTS

The completion of this thesis had been made possible with the inspiration and help of family and friends which can be expressed in these very few words of thank you.

First of all, I thank my two supervisors, Professor Anthony Moffat and Dr Roger Jee. Professor Moffat had offered me not only the opportunity to take upon this project but the motivation to be immersed in the 'whole experience of research', with his encouragement for involvement in various conferences and also publications. I am also most indebted to Dr Roger Jee for his enthusiasm when listening to my ideas, his time in writing the computer software which had been crucial to this work and his feedback on both the practical and written work. My gratitude also goes to David McCarthy for the scanning electron micrographs, Colin James for the computer software support, Fahron Fariba for listening to my lamentations when the going was getting tough and other colleagues who have cheered me up along the way.

I am also grateful to SmithKline Beecham Pharmaceuticals for the financial support of this work and the two industrial liasons, namely, Mr Nigel North and Dr David C Lee. They have supported this work in terms of materials and feedback (which had often offered a refreshing perspective to this work). I am also grateful to Sheelagh Halsey from FOSS NIRSystems, Keith Longdon from Buhler Anatec and Stuart Smith from Bran+Luebbe for their technical assistance during the use of their instruments.

Oliver and Juliet Manyemba for their love and warm friendship, who were more of a family of mine in this country. Kim Yu Ang for his love , care and patience particularly during the last few months in the process of completing this work.

To my parents who are still waiting for the day of my return, who have given me more than I can ask for, I can only say thanks for everything.

Above all, I thank God for the blessings, joy and friendship of the people above.

ABSTRACT

The application of near-infrared spectroscopy (NIRS) in the pharmaceutical industry is facing continued growth due to its speed and ease of use, non-destructive nature and minimal need for sample preparation. The Medicines Control Agency and pharmaceutical industry, however, have expressed their concern about the transferability of the technique. The construction of spectral libraries can be a costly and time-consuming process and therefore the ability to transfer libraries between instruments would offer tremendous savings. However, it is generally recognized that calibrations for quantification purposes are not always directly transferable between instruments. As yet, it is still to be established if libraries of spectra for identification purposes are transferable.

Initially, the transferability of spectra for solvents was examined on eight different instrumental setups which included both grating and Fourier-Transform based spectrophotometers. Careful selection of wavelength range and mathematical pre-treatment were found to be essential parameters for the ability to transfer spectra. The success of this part of the project was particularly marked by the outcome of an inter-laboratory trial conducted based on the above spectral library. An identification rate of almost 100% was achieved for spectra measured from standard modules.

The investigation was then extended to solid pharmaceutical excipients. Effects of sample presentation were first examined. Four important parameters were identified: cup diameter, sample thickness, cup material and packing method, which can significantly affect identification algorithms i.e. Correlation in Wavelength Space (< 0.95) and Maximum Wavelength Distance (> 3.0). Stray light and variation in the reference standards were important sources of errors to be considered. However, their effects can be corrected by measuring the spectra to a common reference standard.

This work has demonstrated that libraries of spectra for purposes can be transferred between different instruments/laboratories providing factors such as spectral wavelength range, mathematical pre-treatment and sample presentation are carefully controlled.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A U	Absorbance units
ASTM	American Society for Testing and Materials
DCA	Direct Content Analyser
FT	Fourier Transform
NIR	Near-infrared
NIST	National Institute of Standards and Technology
PCA	Principal components analysis
PQS	Polar Qualification System
PTFE	Polytetrafluoroethylene
PVC	Polyvinyl chloride
RCA	Rapid Content Analyser
SNV	Standard normal variate
SRM	Standard reference material

CHAPTER 1: GENERAL INTRODUCTION

General Introduction

The near-infrared (NIR) region, comprises the area between 800 and 2500 nm of the electromagnetic spectrum (14300 to 4000 cm^{-1}). NIR absorptions are due to overtones and combinations of molecular vibrations originating in the mid-infrared region. Its development as an analytical technique was initially slow as there was skepticism if useful chemical information could be extracted from these complex and overlapping bands. However, with the introduction of modern computerised instrumentation, NIR spectroscopy has emerged as a powerful analytical technique which affords rapid, precise, convenient (no or little sample preparation), non-destructive and flexible methods for analysis. Within the pharmaceutical industry, applications include quantitative, qualitative and process analysis. An area of particular interest is the identification and qualification of raw materials, resulting from the recent European Commission Directive requiring the testing of every single container of incoming materials for the use in the pharmaceutical industry (European Commission, 1997). With the aid of simple pattern recognition methods to compare sample and reference spectra, NIR spectroscopy has allowed this requirement to become economically viable.

Despite the wide use of the technique within the pharmaceutical industry, it is still viewed by the regulators only as an alternative to the currently accepted official methods (Plugge and Van der Vlies, 1993). One of the main reasons is the apparent non-transferability of the spectral data/applications between different instruments and laboratories, (Van der Vlies, 1996; Moffat *et al.*, 1997). One of the concerns for transferability arises from the high cost incurred in the construction of spectral libraries for both quantitative and qualitative applications. To construct a library, a large number of samples need to be scanned and reference values measured by conventional procedures. The ability to use the same library on different instruments would offer savings in terms of materials, time and human resources.

Two spectra can be considered to be the same if all the elements in the data matrix (i.e. absorbance *vs* wavelength) are identical. In practice, this is unlikely to be attained even for repeat scans on one instrument. There is the ever present random noise due to

fluctuations in ambient conditions, operator variability *etc.* which gives rise to differences. With different instruments, even more sources of errors can be expected.

At present, there is a gap in our knowledge regarding the transferability of spectral libraries for the identification of drugs and medicines. No reports have appeared in the literature. It is likely that the impact of instrumental differences will vary with the type of pattern recognition method used. This thesis investigates the factors affecting the transferability of spectra for identification purposes. Three commonly used pattern recognition methods have been studied: Correlation Coefficient in Wavelength Space, Maximum Wavelength Distance and Polar Qualification System. These methods were chosen because of their mathematical simplicity and in the case of the first two their wide use.

1.1 Theory

The fundamentals of NIR spectroscopy, spectrophotometry and diffuse reflectance theory are covered in this section.

1.1.1 Fundamentals of NIR spectroscopy

Spectra arise from the emission or absorption of definite quanta of radiation when transitions occur between certain energy levels (Banwell, 1966). The energy of a molecule can be considered to be made up from the sum of the electronic, vibrational and rotational *etc.* energies (Born-Oppenheimer approximation).

$$E_{total} = E_{electronic} + E_{vibrational} + E_{rotational} + \ldots \qquad (1.1)$$

When a molecule undergoes a transition from a lower energy state E_2 to a higher energy state E_1 absorption of energy $h\nu$ occurs.

Where v is the frequency of the radiation absorbed and h Planck's constant. The position of a transition is commonly expressed in terms of the wavelength (λ) or wavenumber ($\tilde{\upsilon}$) of the radiation.

$$c = \lambda v \quad \textbf{(1.2)}$$

$$\lambda = c/v \quad \textbf{(1.3)}$$

$$\tilde{\upsilon} = 1/\lambda \quad \textbf{(1.4)}$$

Where c is the speed of light.

The intensity of a transition depends upon the population of the energy levels and the strength of the interaction of molecules with the electromagnetic field. For a sample of N molecules at a temperature T, the number with energy E is given by the Boltzmann distribution:

$$N_E \propto N g e^{-E/kT} \quad \textbf{(1.5)}$$

Where g is the degeneracy of the level (i.e. the number of states corresponding to that energy)

k is the Boltzmann constant.

For the absorption of radiation to occur the lower energy level (E_2) must be populated. The greater the population the more intense the absorption. The absorption of NIR radiation is associated with changes in the vibrational plus rotational energy of molecules, hence for the molecule to interact with the radiation the electric dipole moment of the molecule must change when the atoms are displaced. Vibrations in which there is no dipole moment change will be NIR inactive.

Molecular vibrations

A typical potential energy curve (solid line) for a diatomic molecule is shown **in Fig. 1.1**. There is a repulsion between the positively charge nuclei of both atoms, and between the negative electron "clouds"; on the other hand there is an attraction between the nucleus of one atom and the electrons of the other, and *vice versa*. The two atoms settle at a mean inter-nuclear distance such that these forces are just

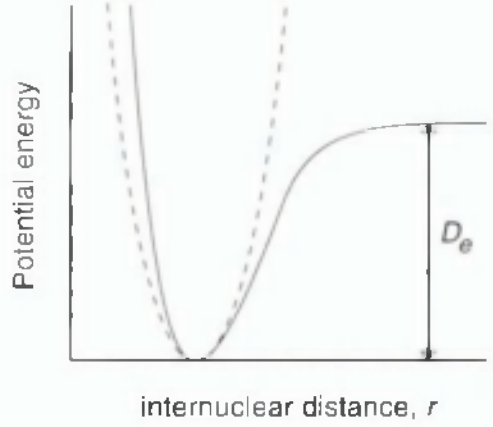


Fig. 1.1
Potential energy curve for a diatomic molecule. Solid line - Morse curve. Dashed line - parabolic curve.

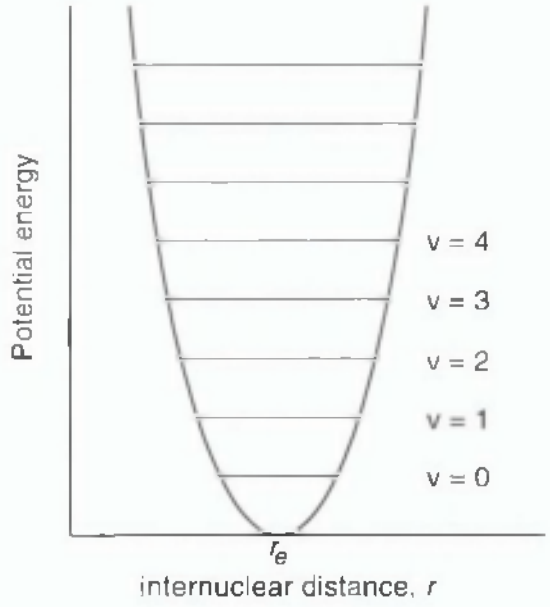


Fig. 1.2
Harmonic oscillator.

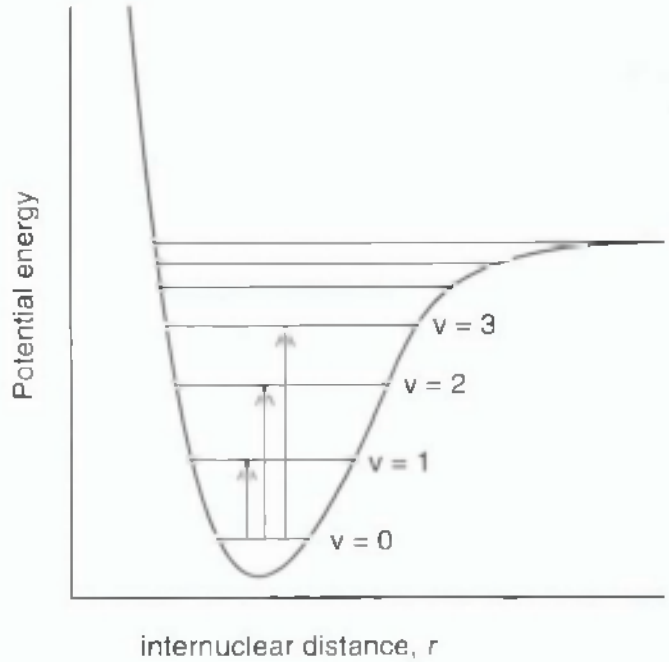


Fig. 1.3
Fundamental, first and second overtones for an anharmonic oscillator.

balanced and the energy of the whole system is at a minimum. If the bond length is stretched too far the bond will break.

The simplest model for the vibration of a diatomic molecule is to assume that the potential energy curve can be approximated to a parabola (**Fig. 1.1**, dashed line), i.e. Hooke's Law applies – the bond extenstion/compression is proportional to the force applied.

Classically the frequency of vibration, ν_o of a *harmonic oscillator* is given by :

$$\nu_o = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad \textbf{(1.6)}$$

Where k is the force constant

μ is the reduced mass of the vibrating system.

For a diatomic molecule the reduced mass is given by :

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \qquad \textbf{(1.7)}$$

Where m_1 and m_2 are the masses of the atoms.

The potential energy V of the oscillator as a function of the displacement, r, is represented by the parabolic equation:

$$V = \tfrac{1}{2}k(r - r_e)^2 \qquad \textbf{(1.8)}$$

Where r_e is the equilibrium distance. The quantized vibrational energies are given by the following formula from quantum mechanics:

$$E_{vibration} = (\upsilon + \tfrac{1}{2})h\nu_o \qquad \textbf{(1.9)}$$

Where υ is the vibrational quantum number and can take the values 1, 2, 3,.....This represents a series of equally spaced energy levels in the molecule, **Fig. 1.2.**

For a harmonic oscillator the selection rule is $\Delta\upsilon = \pm 1$. The allowed absorption transitions are therefore:

$$E_{1\leftarrow 0} = (1 + \tfrac{1}{2})h\nu_o - \tfrac{1}{2}\nu_o = h\nu_o$$

$$E_{2\leftarrow 1} = (2 + \tfrac{1}{2})h\nu_o - (1 + \tfrac{1}{2})h\nu_o = h\nu_o$$

$$E_{2\leftarrow 1} = (3 + \tfrac{1}{2})h\nu_o - (2 + \tfrac{1}{2})h\nu_o = h\nu_o$$

etc.

The predicted vibrational spectrum will consist of a single line of energy $h\nu_o$. At room temperature the majority of the molecules will be in the $\upsilon = 0$ state and therefore only the transition $\upsilon_{1\leftarrow 0}$ will be important. For a typical molecule the values of k and μ are such that the fundamental frequency of vibration will occur in the mid-infrared region of the electromagnetic spectrum. No transitions would be seen in the NIR region.

Real molecules do not behave exactly like simple harmonic oscillators, as r increases the chemical bond becomes weaker and eventually breaks. A more realistic potential energy curve is that shown by the solid line in **Fig. 1.1**. A good mathematical approximation to this *anharmonic oscillator* is the Morse potential energy curve:

$$V = D_e\{1 - e^{-a(r - r_e)}\}^2 \qquad \textbf{(1.10)}$$

Where D_e is the depth of the potential minimum (**Fig. 1.1**)

a is a constant related to various molecular parameters.

When the Schrödinger wave equation is solved for the Morse curve the permitted energy levels are given by :

$$E_{\text{vibration}} = (\upsilon + \tfrac{1}{2})h\nu_e - (\upsilon + \tfrac{1}{2})^2 x_e h\nu_e \qquad \textbf{(1.11)}$$

Where x_e is called the *anharmonicity constant*. The energies of the transitions $\upsilon_{1\leftarrow 0}$, $\upsilon_{2\leftarrow 1}$, $\upsilon_{3\leftarrow 2}$, *etc.* will now be a slightly different from one another, however, as noted above only the $\upsilon_{1\leftarrow 0}$ transition will be important at room temperature. The selection rule for the anharmonic oscillator is $\Delta\upsilon = \pm 1, \pm 2, \pm 3, \ldots$ and consequently transitions such as $\upsilon_{2\leftarrow 0}$, $\upsilon_{3\leftarrow 0}$ are now possible, **Fig. 1.3**.

For $\Delta\upsilon = +1$,

$$E_{1\leftarrow 0} = \{(1 + \tfrac{1}{2})h\nu_e - (1 + \tfrac{1}{2})^2 x_e h\nu_e\} - \{\tfrac{1}{2}h\nu_e - (\tfrac{1}{2})^2 x_e h\nu_e\}$$
$$= (1 - 2x_e)\, h\nu_e$$

For $\Delta\upsilon = +2$,

$$E_{2\leftarrow 0} = \{(2 + \tfrac{1}{2})h\nu_e - (2 + \tfrac{1}{2})^2 x_e h\nu_e\} - \{\tfrac{1}{2}h\nu_e - (\tfrac{1}{2})^2 x_e h\nu_e\}$$
$$= (1 - 3x_e)2h\nu_e$$

For $\Delta\upsilon = +3$,

$$E_{3\leftarrow 0} = \{(3 + \tfrac{1}{2})h\nu_e - (3 + \tfrac{1}{2})^2 x_e h\nu_e\} - \{\tfrac{1}{2}h\nu_e - (\tfrac{1}{2})^2 x_e h\nu_e\}$$
$$= (1 - 4x_e)3h\nu_e$$

The anharmonicity constant is typically small ($x_e \approx 0.01$), hence the frequency of these transitions lie close to ν_e, $2\nu_e$ and $3\nu_e$. The line near ν_e is called the fundamental absorption, while those near $2\nu_e$ and $3\nu_e$ are called the *first* and *second* overtones, respectively.

The energies of these overtones are such that for many molecules they lie in the NIR region of the electromagnetic spectrum. The probability of these spectroscopic transitions decreases rapidly with increasing $\Delta\upsilon$ consequently these absorptions are weak. Typically, the first overtone is about $1/10^{th}$ and the second overtone $1/100^{th}$ the intensity of the fundamental absorption. For this reason, NIR spectra can commonly be recorded directly on undiluted samples.

There is only one mode of vibration for a diatomic molecule, the bond stretch. In polyatomic molecules there are several modes because bonds may stretch and angles may bend. For a non-linear molecule that consists of N atoms, there are $3N - 6$ fundamental modes of vibration ($3N - 5$, for a linear molecule). Some vibrations/bends will be associated with small groups of atoms, e.g. just one bond. Absorptions due to these vibrations in the mid-infrared region are very useful for identifying functional groups. Most of these vibrations occur at high energies because the reduced mass of the vibrating system will be small. This gives rise to what is

known as the *functional group* region in the mid-infrared region (4000 to 1300 cm^{-1}). Many vibrations will be associated with the whole structure and will be characteristic of that particular molecule only – this gives rise to what is known as the *fingerprint* region in the mid-infrared region (1300 to 600 cm^{-1}). Polyatomic molecules may exhibit simultaneous changes in the energies of two or more vibrational modes: the frequency observed will be the sum of ($\nu_1 + \nu_2$, $2\nu_1 + \nu_2$, *etc.*) or the difference between ($\nu_1 - \nu_2$, $2\nu_1 - \nu_2$, *etc.*) the individual frequencies (note: $2\nu_1$ represents the first overtone). This results in very weak bands that are called *combination* and *subtraction* bands – the latter are possible but rarely observed in room temperature NIR spectra. The vibrational transitions observed will also be accompanied by changes in rotational energy of the molecule making the absorption lines broad and appear as bands. For even quite small molecules the NIR spectrum will therefore be expected to be complex, although unlike the mid-infrared spectrum, it is more difficult to assign bands.

1.1.2 Fundamentals of NIR spectrophotometry

The transmittance, T, of a sample is defined as I_t/I_0 where I_t is the intensity of the transmitted radiation and I_0 the intensity of the incident radiation (Willard *et al.*, 1988). Absorbance, A, is defined as $\log_{10}(I_0/I_t)$ or $-\log(1/T)$. The dependence of absorbance on sample path-length and concentration is described by the Beer-Lambert Law:

$$A = \varepsilon c l \qquad (1.12)$$

Where ε is the absorption coefficient

c is the concentration

l is the path-length.

The Beer-Lambert Law is based upon a number of assumptions: (1) parallel radiation beam, (2) monochromatic radiation, and (3) no interaction between absorbing centres. Deviations from the law commonly arises from dissociation/association of absorbing centres, changes in the refractive index of the sample, finite bandpass (i.e. non-monochromatic radiation), stray radiation, *etc.*

Stray radiation can be a major problem. Usually stray radiation is not absorbed by the sample and causes negative deviations from the Beer-Lambert Law.

$$A_{observed} = \log_{10}[(I_o + I_s)/(I_t + I_s)] \quad \textbf{(1.13)}$$

Where I_s is the intensity of stray light.

For transmission spectroscopy, this effect can be minimised by avoiding the use of strongly absorbing peaks/high concentrations.

1.1.3 Theory of diffuse reflectance

Radiation reflected from a material is made up of two components: specular and diffuse (see **Fig. 1. 4**). Specular reflectance (also known as surface, regular or Fresnel reflectance) is the name given for light reflected from the surface of the sample. Its intensity has a Gaussian distribution with regards to the direction of the exciting light source and does not tend to contain useful information for quantitative and qualititative chemical analysis. Diffuse reflectance (also known as body reflectance) results from radiation which has penetrated and interacted with the molecules within the sample. It emerges through 180° at random and forms the basis for NIR reflectance measurements. As it is dependent on both the scattering and absorption properties of the sample, it provides 'physical' and 'chemical' information. Diffuse reflectance can be affected by particle size (Devaux *et al.*, 1995), particle shape (Olinger and Griffiths, 1993), particle size distribution (O'Neil *et al.*, 1998), sample porosity (Mark and Tunnel, 1985) and crystallization state (Norris *et al.*, 1997) of a sample and renders NIR reflectance spectroscopy a powerful tool for measuring physical properties. An exact mathematical description of diffuse reflectance has not yet been modelled due to the difficulty in describing the path for the propagation of radiation (Osborne *et al.*, 1993). The best model at present is the Kubelka Munk Theory which attempts to account for scatter by describing the reflection from an infinitely thick sample as a function of absorption and scatter coefficients.

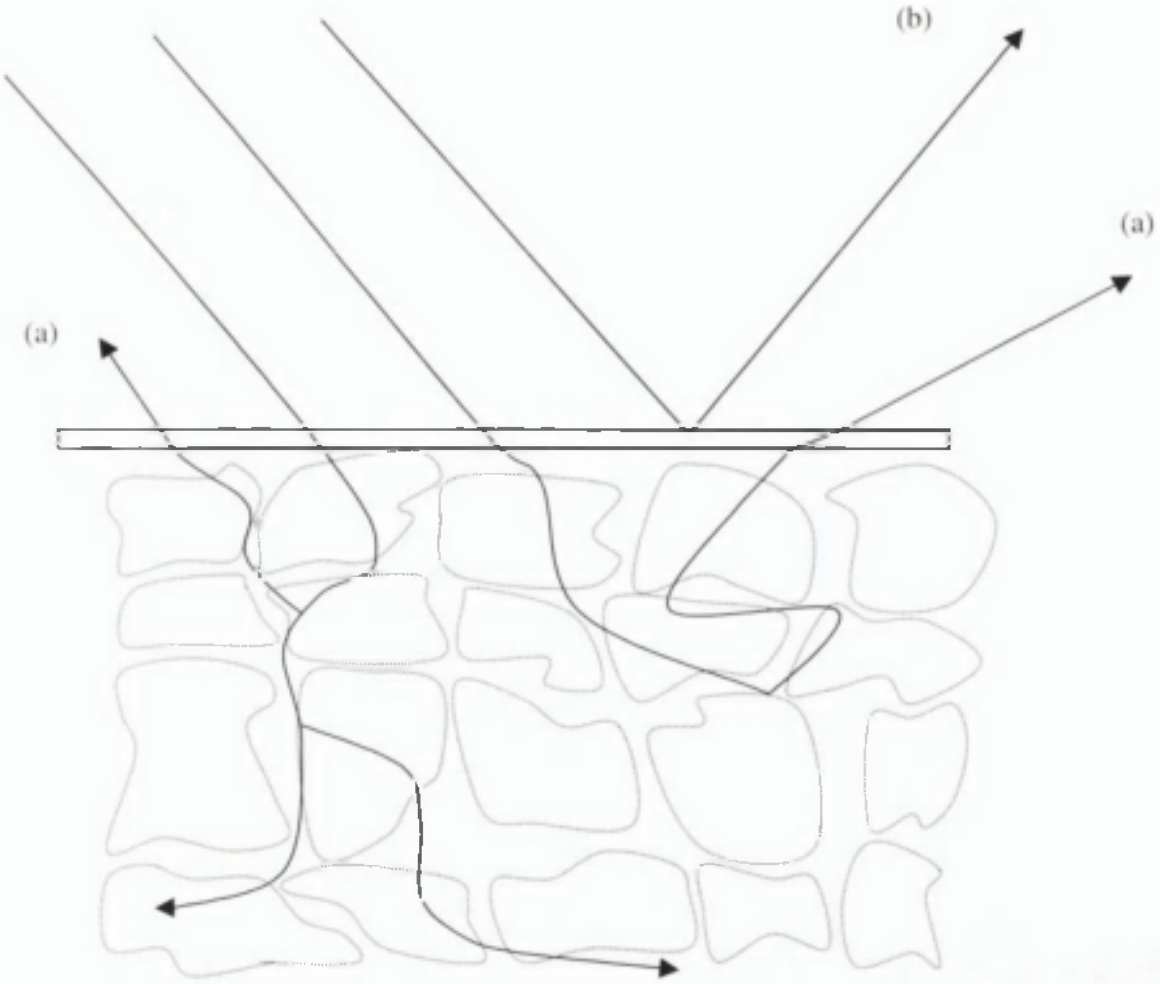


Fig. 1.4 (a) Diffuse (body) and (b) specular reflectance

$$F(R) = \frac{(1 - R_\infty)^2}{2R_\infty} = \frac{k}{s} = \frac{\varepsilon ln10}{s} = \frac{c}{a} \qquad \textbf{(1.14)}$$

Where $F(R)$ is the Kubelka Munk function

R_∞ is the absolute reflectance of an infinitely thick sample

k is the absorption coefficient

s is the scattering coefficient

According to this theory, light scatter is multiplicative. Therefore, by multiplying each wavelength of the sample by the coefficients, the non-linearity can be modelled. However, Næs and Isaksson (1994) noted that there is also an additive component to the model arising from the construction of equipment where only part of the reflected light is detected.

In practice, the relative reflectance (R) which is the ratio of the intensity of light reflected by the sample to that by the standard, is preferred to the absolute measurement. For a further discussion on the reflectance standards, please refer to section 1.3.6.

Relative reflectance, $R = I_r / I_{st}$ **(1.15)**

Where I_r is the intensity of light reflected from the sample.

I_{st} is the intensity of light reflected from a standard.

Many systems do not follow the Kubelka-Munk equation and a commonly used practical alternative is based on a relationship between concentration and relative reflectance similar to the Beer-Lambert Law:

$$A = -\log_{10} R = a'c \qquad \textbf{(1.16)}$$

Where A is the apparent absorbance

c is the concentration

a' a proportionality constant.

Although this relationship has no theoretical basis, it provides satisfactory results in practice (Blanco *et al.*, 1998).

The effects of stray radiation on reflectance measurements will be considered in chapter 6.

1.2 NIR spectroscopy as an analytical technique

NIR spectroscopy started as an analytical technique in the agricultural industry when Karl Norris (1962) constructed the first 'grain moisture meter'. As noted in the general introduction, development of the technique was slow as there was skepticism if the complex and overlapping bands of the NIR region could be interpreted to give useful chemical information. However, with the advent of powerful and sophisticated computers, growth in the use of multivariate analysis and also the vast improvements achieved on the capability of the instrumentation since the 80's (McClure, 1994), its usage has expanded to other industries such as the petrochemical (Buchanan, 1992) polymer (Kradjel & McDermott, 1992), textile (Ghosh & Rodgers, 1992) and pharmaceutical industry.

The past few years has seen the growth in the use of NIR spectroscopy in the pharmaceutical industry which can be attributed to the following advantages of the technique:

- Speed of measurement. The time required to record a spectrum is typically less than 1 minute. High performance liquid chromatography, a method that is presently dominant within pharmaceutical analysis, typically requires at least half an hour.

- No need for sample preparation. The small absorption coefficients for overtones and combination bands allows measurements *in situ*. As such, it lends itself as a method that is quick, convenient, non-destructive and also environmentally friendly as compared to the 'wet chemical' methods requiring the use of hazardous chemicals. Without the need for sample preparation, operator error is much decreased.

- Quality of measurement. The technique is marked by a high signal to noise ratio (10000 : 1) and good reproducibility.

- Richness of information of the near-infrared spectrum. As the NIR spectrum is affected by both physical and chemical parameters, it enables a 'holistic' analysis of a sample, particularly for qualitative analysis.

- Flexibility of measurement. Many commercial NIR instruments have a wide range of sampling accessories to measure samples such as tablets, powders and liquids.

Examples of the applications of NIR spectroscopy are given in the following sections. With the above advantages, the application of the technique has increased significantly within the pharmaceutical industry (Blanco *et al.*, 1998). Such applications include assay of actives, moisture determination, identification of raw materials, physical testing (e.g. particle size determination) *etc.* Also, it has enabled the development of new areas of applications not afforded by more conventional methods. One example is process chemical analysis. The speed of measurement has allowed real time monitoring of processes such as blending (Hailey *et al.*, 1996) and polymorphic conversion (Norris *et al.*, 1997).

1.2.1 Quantitative analysis

Quantitative analysis formed the initial focus for the use of NIR spectroscopy within the pharmaceutical industry, with assays of actives and moisture determination being the mainstream applications (Blanco *et al.*, 1998).

Assay methods for actives, intermediates and finished pharmaceutical preparations have been successfully developed. The nature of the samples include syrups (Ciurczak, and Torlini, 1987) suspensions and powders (Zappala and Post, 1977; Last *et al.*, 1993), gel preparations (Corti *et al.*, 1989) whole tablets (Trafford *et al.*, 1998), crushed tablets (Corti *et al.*, 1990), antibiotic powders (Lonardi *et al.*, 1989), creams, granules and injections. Transdermal patches can also be assayed directly using a sampling accessory developed by FOSS NIRSystems called RapiDerm™.

Quantification of multi-component preparation was also reported to be possible (Ciurczack and Maldacker, 1986).

As NIR spectroscopy is not considered a method of direct measurement due to its lack of distinct chemical bands that can be directly correlated to mass, volume or concentration, the spectral data will need to be referenced to measurements of a primary technique. Even though this remains a point of contention (Moffat, 1998), most quantitative methods would be based on regressing the NIR measurements of a representative set of samples to a reference method. Regression techniques include multiple linear regression (Mark, 1992), principal components regression (Mark, 1992a) and partial least squares regression (Bjorsvik and Martens, 1992). To extend the range of concentration beyond the typical nominal value for assay determination, out-of-range samples should be prepared at the pilot plant (Jouan-Rimbaud *et al.*, 1995) where possible. Otherwise, synthetic samples that have been prepared at concentrations over the manufacturers specified range (Blanco *et al.*, 1996) or production batches that have been spiked with small amounts of actives or excipients (Blanco *et al.*,1994) can be used and have reportedly produced comparable calibration results (Blanco *et al.*, 1997a).

1.2.2 Physical determination

NIR diffuse reflectance spectroscopy is also an important tool for physical determination as the spectrum is sensitive to the changes in the physical properties of the sample as described in section 1.1.3. A previous study had shown that reflectance varies non-linearly with particle size (Ciurczak *et al.*, 1986a). Evidently, data pre-treatment plays a significant role in extracting physical information. O'Neil *et al.* (1998) found that the NIR reflectance data were best correlated to log median particle size as determined using forward angle laser light diffraction. They then extended their measurements to the 5 – 95% quantiles and were able to determine the particle size distribution (O'Neil, 1999).

Morriseau and Rhodes (1997) found an increase in tablet hardness produced an upward shift (increase in absorbance) in the NIR spectra and found that hardness predicted using NIR spectroscopy was as least as precise as the conventional methods.

1.2.3 Process analysis

The concept of process analytical chemistry (PAC) stresses that measurements should be made on processes rather than end-products for reasons of greater efficiency in energy, cost and time (Kowalski *et al.*, 1987). NIR spectroscopy is ideally suited to this purpose. One such area that has been explored is blend analysis. Ciurczak (1991) showed that the technique was capable of determining blend homogeneity off-line, while Sekulic *et al.* (1996) developed on-line measurements by fitting a fibre optic probe through the rotation axis of the blender. Homogeneity was monitored by calculating the standard deviation of blocks of spectral data as a function of time. The convergence of spectral data to a point of constant variance was taken to indicate that the blend had achieved a 'homogenous' mix. Automation of the above process using a compilation of software was also possible (Hailey *et al.*, 1996). Another potential area is reaction monitoring. Using the example of an esterification process, Hammond *et al.* (1999) showed that it was possible to perform real-time reaction monitoring in solid phase synthesis by NIR spectroscopy. Norris *et al.* (1997) have used principal components to model the free energy profile of the inter-conversion of polymorphic forms.

1.2.4 Identification and qualitative analysis

The British Pharmacopoeia (1999) describes an identity test as a verification of the material according to the label on the container. In essence, it focuses on the determination of the chemical entity. Qualification analysis, focuses on the fitness of the sample for use, and tests such as particle size analysis and detection of impurities.

To determine the quality of a sample, the underlying pharmacopoeial philosophy requires several different tests to be carried out to determine the holistic characteristic of the sample i.e. mid-infrared spectroscopic fingerprinting for identification, sulphated ash for non volatile inorganic impurities in organic substances *etc.* However, Van der Vlies *et al.* (1995) have reported that the outcome of these tests, that are sample destructive and time-consuming, do not always represent the functionality of the sample. Instead, these workers and Lieper *et al.* (1998) have pointed out that quality determination should provide a means of discriminating

between samples representing the normal chemical and physical distribution from a validated manufacturing process. As such, NIR spectroscopy with its high power of discrimination and simplicity of measurement, has also become an increasingly popular analytical tool for identification and qualification. It has been used to identify and qualify raw materials (Gemperline *et al.*, 1989; Brunner *et al.*, 1999) identify end products (Khan *et al.*, 1997; Ciurczack and Maldecker, 1986), directly classify tablets for clinical supplies (Han and Faulkner, 1996) and also for the detection of microbiological contamination (Galante *et al.*, 1990). One significant contribution of the technique is to have made the requirement by a recent EC Directive (European Commission, 1997) to test every single container for raw materials, economically viable. Brunner *et al.* (1999) reported a cost reduction of 86% when using near-infrared spectroscopy as compared to traditional wet chemistry and mid-infrared methods for this purpose.

1.3 Instrumentation

The basic instrumental configuration for NIR spectroscopy is similar to that used in the visible and ultra-violet region, **Fig. 1.5.**

1.3.1 Source

The most commonly used source for NIR radiation is the quartz halogen lamp in the power range of 10 to 200 W. Its incandescent filament produces continuous radiation with a running temperature of 2600 to 3200 K and typically have a life of 2000 – 3000 hours. The spectral output changes with age and will affect consistency of

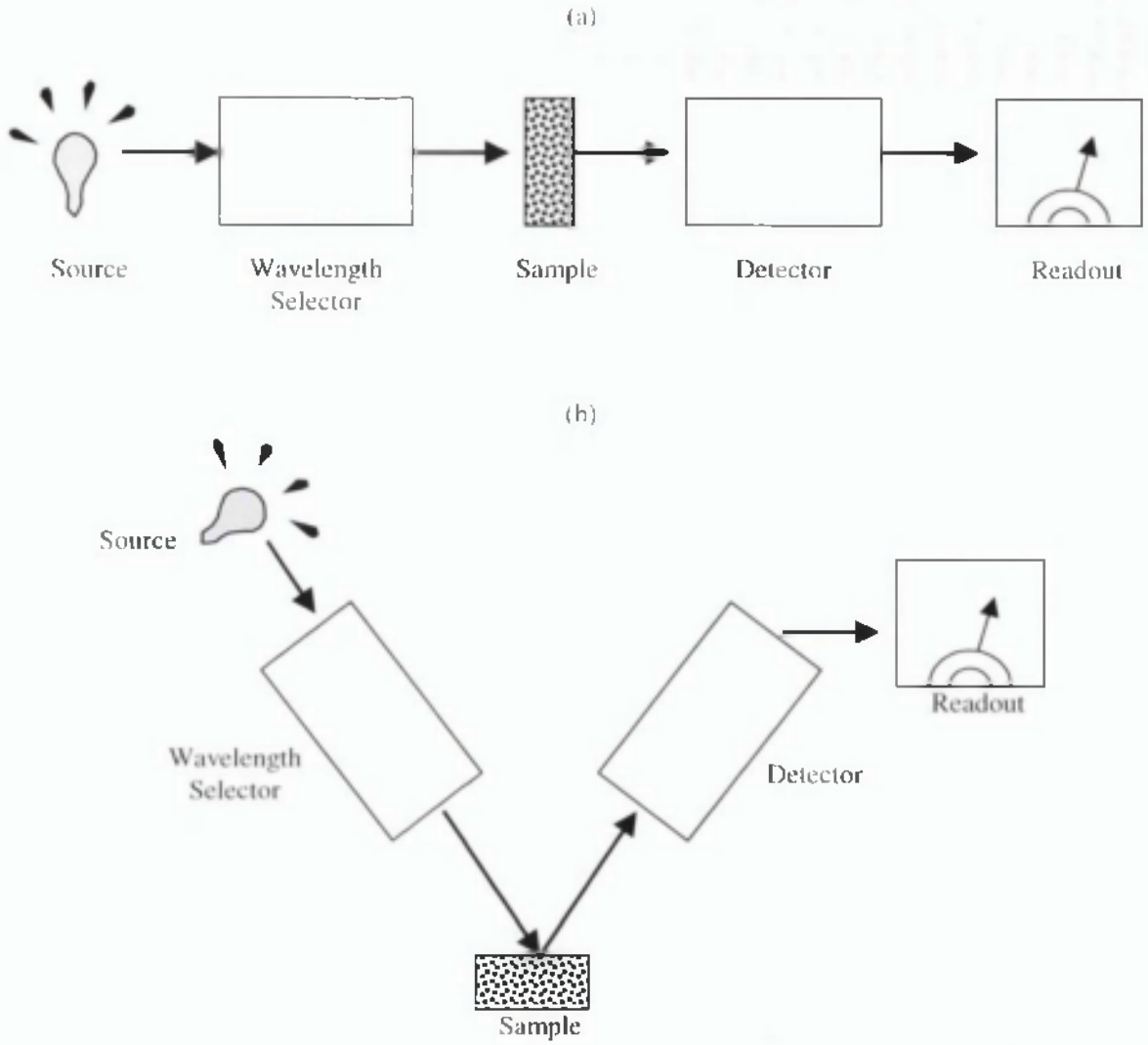


Fig. 1.5 Basic instrument configurations for (a) transmittance and (b) reflectance measurements.

measurement. This is overcome by internal referencing in many NIR instruments. Relative reflectance measurements should not be affected by such long term changes.

Non thermal sources with lower power requirements have also been used. These include light emitting diodes (LED), laser diodes and lasers (Osborne *et al.*, 1993). They offer advantages of emitting radiation over a selected and narrow range of wavelength, hence, pre-filtering may not be necessary. These non-thermal systems are often more compact than thermal sources and are more portable.

1.3.2 Wavelength selection

Instruments using continuous radiation sources use a number of different systems for wavelength selection. Early NIR instruments commonly used discrete filters. Grating monochromators are thought to be most commonly used at the present time (McClure, 1994). Holographic gratings, that are of excellent quality and cheaper than conventional mechanically ruled surfaces, can be economically produced by lasers and photoetching. Recorded spectra have a linear wavelength scale. Increasingly, Fourier Transform (FT) and acousto optic tuneable filter (AOTF) instruments are becoming increasingly popular due to fewer moving parts in such instruments. As spectra recorded from FT are linear to wavenumber, difficulty in directly transferring spectra to systems measured on dispersive instruments can be expected.

1.3.3 Sampling accessories

The design of sampling accessories are dictated by the type of measurement used; reflectance, transmittance or transflectance.

Reflectance

Reflectance measurement systems are designed in such a way as to maximise the collection of the diffuse component and to minimise the specular component.

This can be achieved using '0 - 45°' detector geometry, **Fig. 1.6**. Usually, 2 – 6 detectors are arranged concentrically, or in parallel rows on either side of the source. Alternatively, an integrating sphere can be used as a greater proportion of the

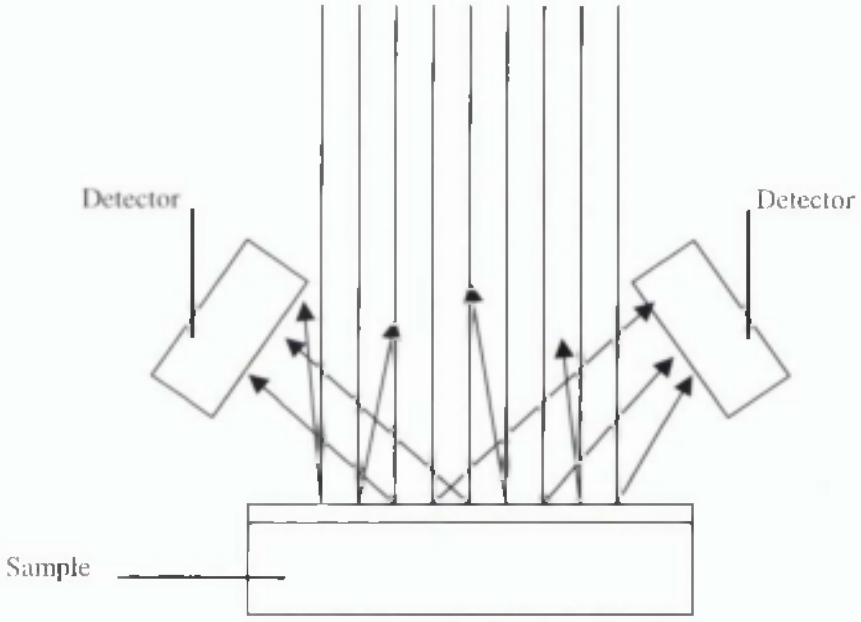


Fig. 1.6 '0-45°' detector geometry

radiation can be collected. Yet another option is the fibre optic probe which allows for remote sampling.

For measurement of spectra, samples can be placed in vials. Ideally, quartz vials that have minimal absorbances in the NIR region should be used. However, for economic reasons, glass vials are often used although the effects of their background absorption are yet to be ascertained. Fibre optics offer tremendous convenience, particularly in a warehouse situation, where the probe can be placed directly into the drums of samples.

The wide variety of sampling options available also presents a challenge to the transferability of spectra. For the 0 - 45° geometry, variations in the arrangement of detectors and sizes and shapes of sources are known to affect the spectra. The use of fibre optics results in attenuation of signal in the higher wavelength region that limits the usable range of the spectra. Therefore, when transferring spectra between instruments of different working ranges, wavelength regions will have to be considered.

Transmittance

Transmittance can be used for the measurement of liquids and solids. For liquids, path-lengths of approximately 1 to 1.5 mm in the 1000 to 2500 nm region tend to give the best spectra. Unless the liquid is clear, deviation from the Lambert Law can be expected and data pre-treatment needed. For diffuse transmittance measurements of solids, absorbances up to 3 - 6 A.U. can be expected (FOSS NIRSystems, 1996a).

Transflectance

Transflectance can be used for the measurement of liquids and combines reflectance and transmission, **Fig. 1.7.** Measurements can be made using a sample cup and reflector on a horizontal setup or a fibre optic probe, fitted with a reflector tip. The radiation traverses the sample and is then reflected back through the sample. The optical path-length is equivalent to twice the physical path-distance (between source and reflector).

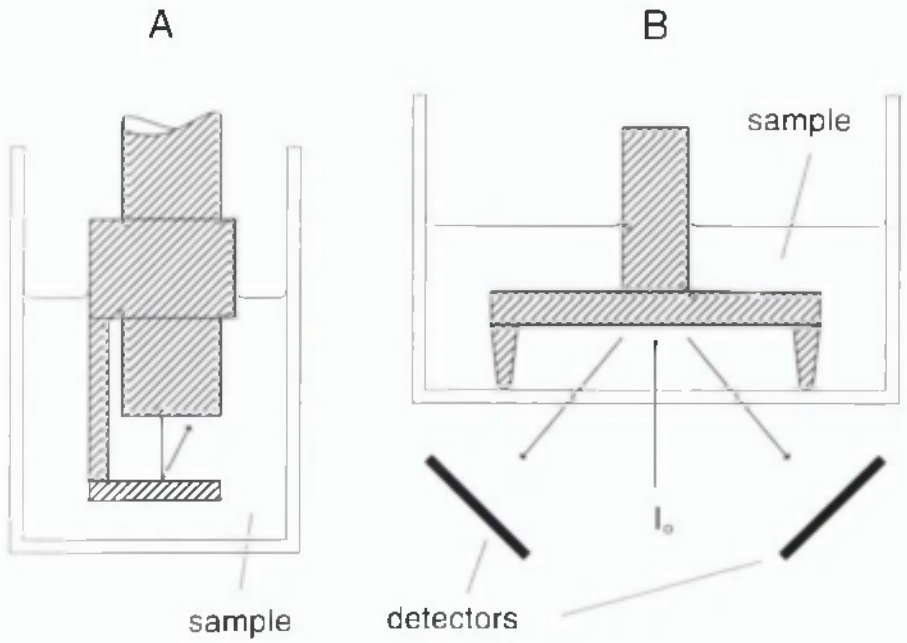


Fig. 1.7 Schematic diagram for transflectance measurements using (A) a fibre optic probe, and (B) a sample cup and reflector.

1.3.4 Detectors

The most commonly used detectors in the NIR region are: lead sulphide, PbS (usable over the range of 1000 to 2500 nm), lead selenide, PbSe (usable over the range of 2500 to 3500 nm) and indium gallium arsenide, InGaAs (usable over the range of 1000 to 1800 nm.) The lead salt detectors are typically squares 1 mm ×1 mm to 10 mm × 10 mm. InGaAS are generally smaller (1 mm × 1 mm to 6 mm× 6 mm) and they are more sensitive than PbS, size for size.

Detectors are affected by changes in temperature and are often temperature stabilised. Lead salt detectors may be cooled to increase their sensitivity to the longer wavelength range and also improve their signal to noise ratio (Osborne *et al.*, 1993).

1.3.5 Fibre optics

Fibre optics, which provides a means for transferring near-infrared energy, allows for remote sampling from the instrument. The fibres can be made from silica glass (which operates up to 2500 nm) or infrared grade (water free) quartz (which extends the range to 3500 nm). Light transmission occurs via the phenomenon of total reflection (Osborne *et al.*, 1993). Attenuation of the signal often limits the workable range to below 2200 nm (FOSS NIRSystems, 1996). Fibre optics are available as single fibres (monofilaments) or as bundles. Care is needed during use as bending can cause microfractures, which can seriously alter their transmission characteristics. A 'minimum bending radius' is commonly specified for each type of fibre optic. For example, for the Bran+Luebbe Diffuse Reflectance Probe, it is 200 mm (Bran+Luebbe, 1996).

1.3.6 Reference standards

Reference standards form an integral part of any instrumentation as they serve to ensure the stability of measurements made. Taylor (1983) defined a reference standard as *"a substance for which one or more properties are established sufficiently well for use to calibrate a chemical analyzer or to validate a measurement process"*.

Therefore, it is imperative that the standards themselves are of high quality and reproducibility. The National Institute of Standards and Technology (NIST) produce certified standard reference materials (SRM) and these often form the 'gold standard' to which other standards can be traced. As SRMs are expensive and scarce, alternative materials are necessary. As such, the European Pharmacopoeia (1997) allows considerable flexibility for the standards used to verify the wavelength and photometric absorbance scale of NIR instruments.

Reflectance standards

In NIR spectroscopy, reflectance standards serve two purposes: to set the 100% reflectance and to ensure the stability of the photometric scale. The Bureau Central Internationale de l'Eclairage (1970) has described the perfect diffuse reflectance standard to be 100% diffusely reflecting, transportable, stable, homogenous with a smooth surface, non-translucent, nonfluorescent and easy to handle. In a roughly chronological order, materials that have been employed to make the proverbial 'white standard' are as follows: magnesium carbonate, calcium carbonate, magnesium sulphate, barium sulphate powders, magnesium oxide powders, polytetrafluoroethylene (PTFE) and ceramic tiles.

The early standards i.e. magnesium carbonate, chalk and magnesium oxide fell out of favour due to reasons of instability and also poor reflectance despite being white materials. Barium sulphate and magnesium oxide powders offered great improvement but the purity of component materials was still questionable (Grum, 1976). Weidner *et al.* (1981 & 1985) then reported that pressed PTFE powder offered remarkably high diffuse reflectance (> 99 % between 350 to 1800 nm) and that the material could be reproducibly manufactured between different laboratories. Certified grades of PTFE i.e. Spectralons® from Labsphere are now available commercially. Spectralons® are used by many instrument manufacturers such as Buhler Anatec and Bran+Luebbe as reflectance standards.

Despite having near-ideal properties for a reflectance standard, PTFE suffers the disadvantage of being quite soft and easily damaged if too much pressure is applied during measurement. It has also been reported that PTFE readily attracts dust due to

the formation of static charges. Hence, care of use and cleaning are necessary (Springsteen *et al.*, 1996) to obtain reproducible results.

Ceramic tiles containing a white pigment (usually titanium oxide) and backed by a metallic substrate, on the other hand, offer the practical advantage of being durable. FOSS NIRSystems use such standards.

The use of different reflectance standards give rise to discrepancies in measurements between different instruments. In a study to compare diffuse reflectance measurements between different laboratories, Knee and Deadman (1998) reported that differences of measurements were significant (as high as 88.28% in the most extreme case) and the reference was identified as an important contributing factor. In fact, the authors also recognised that even if the same reflectance material was used, differences would still occur as materials do change with time and usage. This presents a very significant problem for transferability.

NIST traceable-portable standards to standardise the linearity of instruments are available, e.g. a set of 8 diffuse reflectance standards, which have been calibrated to a 'master' instrument is available from Bran+Luebbe (Griffin, 1993). FOSS NIRSystems provides a set of 9 grey standards made from Spectralon® (P/N: AP-0100).

Wavelength standards

SRMs for checking the wavelength accuracy within the near-infrared region are available for the reflectance, and more recently the transmittance, mode of measurement. They are the SRM-1920 and SRM-2035 respectively.

The SRM-1920 is a mixture of dysprosium, erbium and holmium oxide covering the region between 700 to 2000 nm with 37 reflectance minima. An uncertainty level of no greater than ± 1 nm with respect to the location of the wavelength reflectance minima is reported (Weidner *et al.*,1986). SRM-2035 has seven certified peaks spanning the range from 10 500 to 5100 cm^{-1} (952 to 1960 nm), with 95% confidence limits of less than 0.5 cm^{-1} for each peak location (Choquette *et al.*, 1999).

Many instruments have in-built standards for checking wavelength accuracy. The FOSS NIRSystems and Bran+Luebbe spectrophotometers use polystyrene and dydinium paddles which have peaks at 1143, 1681, 2164 and 2303 nm and 683, 805 and 878 nm respectively. For an external check, the National Physical Laboratory recommends the use of rare earth oxides embedded in a PTFE matrix for reflectance measurements and glass or metal on silica filters for the transmittance instruments (Verrill, 1994).

1.4 Data Pre-treatment

Adams (1995) described the aims of data pre-treatment as follows :

(a) to reduce the amount of data and eliminate data that are irrelevant to the study being undertaken

(b) to preserve or enhance sufficient information within the data in order to achieve the desired goal

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

Considering the wealth of information and the lack of distinct chemical bands in NIR spectroscopy, the need for data pre-treatment is indisputable. The type of data pre-treatment used should match the need of the particular application. For the identification of powdered substances, data pre-treatment serves the purpose of enhancing the separation of overlapping and broad bands. This can be done using derivatives to remove any baseline shifts arising from the changes in packing density. Conversely, for particle size determination, it would be desirable to preserve the baseline where the physical information resides. In this case, the Kubelka Munk functions would be more applicable.

Facilities to carry out the commonly used data pre-treatments i.e. derivatives, normalization, standard normal variate, multiple scatter correction and also multivariate procedures such principal components analysis *etc.* are generally

available in the software packages of instrument manufacturers. Proprietary packages can offer more sophisticated options such as artificial neural networks.

Data pre-treatment procedures can often be used in combination and there is no definitive criteria for the choice and order used. Often, they are selected on an *a priori* basis and this is dependent on the preference of the user, type of instrument and data set used. Blanco *et al.* (1997) found taking the second derivative most effective in reducing scatter as compared to normalisation, multiplicative scatter correction, standard normal variate and detrending for the quantification of active measured on a dispersive instrument. Using a set of synthetic spectra, Hindle and Smith (1996) reported that multiple scatter correction, log $1/R$ and log $1/R$ in combination with multiple scatter correction were almost equally effective in producing a robust calibration. Some example of typical pre-treatments are illustrated in **Fig. 1.8.**

1.4.1 Derivatives

Derivatisation serves the primary purpose of enhancing the resolution of overlapping bands. In theory, any order of derivative may be used, however, in practice, derivatives greater than the second are often avoided as the signal/noise ratio decreases by a factor of 2 with each derivative calculated. Peaks and valleys on a first derivative spectrum would correspond to points of inflexion on the original spectrum, while the second derivative spectrum gives an inverse peak at the position of a peak in the original spectrum and may render greater ease for visualing band positions.

Derivatives can be calculated using various mathematical procedures. The commonly used procedures are the boxcar averaging method, the moving point average method and the Savitsky-Golay filter (Savitsky and Golay, 1964).

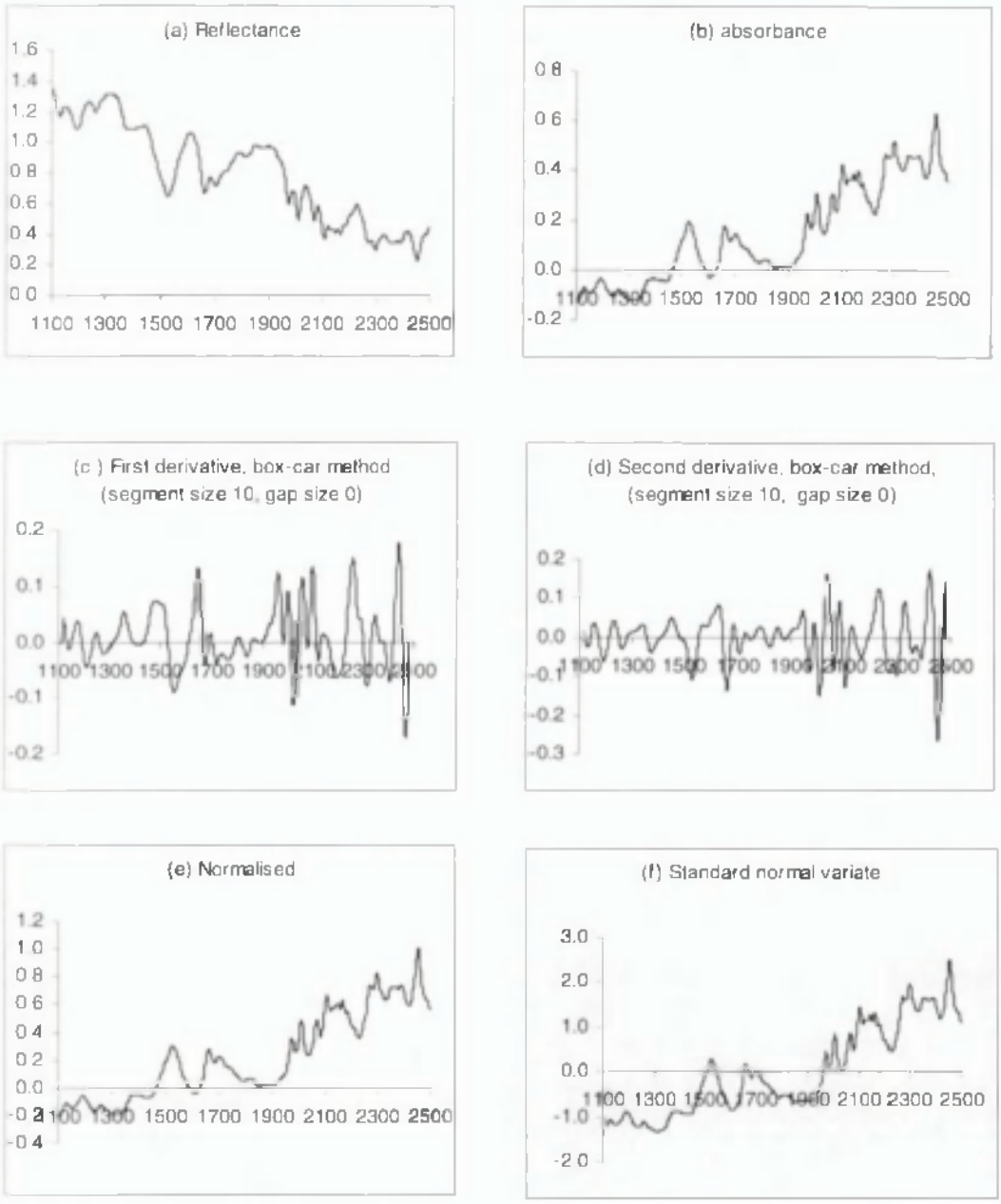


Fig 1.8 Various methods of data pre-treatment for a powdered sample of butyl *para* hydroxybenzoate.
Note : (c) to (f) show the data pre-treatment on the *absorbance* spectra.

The boxcar averaging method divides the spectral data into successive equal-size blocks (or segments) of data-points and the average of each block calculated. The difference between the average values of the respective blocks represents the derivatives. The process can be repeated until the desired order of derivative is obtained. 'Gaps' between segments can also be included to increase the smoothing effect. The 'moving point average' method is effectively based on the same principle of using averages from blocks of spectral data-points, except that the blocks overlap. This method distorts the spectrum to a lesser degree.

The method of Savitsky and Golay, on the other hand, is based on a completely different approach . It uses convolution filters based on polynomial functions, where coefficients are derived from a least squares fit.

1.4.2 Normalization

Many different types of normalisation exists. In this work, normalisation to 1 has been used. The ordinate values at each wavelength in the spectrum (x_i) is replaced by y_i according to equation:

$$y_i = \frac{x_i}{max\ x} \qquad \textbf{(1.17)}$$

Where $max\ x$ is the maximum ordinate value in the original spectrum within the wavelength range being normalised.

1.4.3 Standard normal variate

The aim of applying standard normal variate (Barnes *et al.*, 1989) is to remove the baseline absorbance shifts arising from changes in packing density: It is best described by the equation:

$$y_i = \frac{(x_i - \bar{x})}{s} \qquad \textbf{(1.18)}$$

where x_i is the ordinate values at wavelength i.

s is the standard deviation of the ordinate values over the wavelength range being standardised.

$\bar{x}$ is the mean ordinate value over the wavelength range being standardised.

This transformation first centers the spectral values by subtracting the mean of the individual spectrum ($\bar{x}$) from the values at each wavelength (x_i). These centered values are then scaled by dividing by the standard deviation (s) for that spectrum.

1.4.3 Principal Components Analysis (PCA)

PCA is a multivariate statistical treatment, which has been defined by the American Society for Testing and Materials 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."

Effectively, PCA reduces the data to a small number of variables known as principal components, which are linear combinations of the original variables. The first principal component often models more than 90% of the variation. In the near-infrared spectrum, this often represents information regarding either, particle size or water. These two aspects often overwhelm the NIR spectrum and create a phenomenon known as co-linearity, where measurements at different wavelengths offer the same information. Therefore, PCA helps to reveal the inherent structure of the data which correlates to the analyte in question.

1.5 Identification and qualification in the pharmaceutical industry

To develop a method for identification and qualification requires a representative spectral library to be constructed. The sample in question would be classified with reference to the library using a suitable pattern recognition method. It is important to

validate such libraries with not only materials it contains, but also substances not represented in the library (Plugge and van der Vlies, 1993; Gerhausser and Kovar, 1997). With libraries used for qualification, it is important to include samples representing inter-batch variation and time-variation of a production cycle. To validate the library, out of specification samples should also be used (Gemperline *et al.*, 1989).

A range of pattern recognition methods of varying complexity and discrimination capability are available such as Correlation in Wavelength Space, Maximum Wavelength Distance, Polar Qualification System and also multivariate procedures such as Mahalanobis Distance and Soft Independent Modelling Class Analogy (SIMCA). Applications from the literature based on each of these techniques will be discussed along with the principles of the methods.

1.5.1 Correlation Coefficient in Wavelength Space

The Correlation Coefficient in Wavelength Space, or spectral value or match index, gives a numerical measure of similarity between the two spectra. Two different measures of correlation are in common use, equations **1.19** and **1.20**.

Equation 1.19 gives the *dot product* correlation coefficient and represents the cosine of the angle between the vectors for the two spectra.

Equation 1.20 gives the *product moment* correlation coefficient.

$$r_{jk} = \frac{\sum_i x_{ij} x_{ik}}{\sqrt{\sum_i x_{ij}^{\ 2} \sum_i x_{ik}^{\ 2}}} \qquad \textbf{(1.19)}$$

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

where r_{jk} is the correlation between spectra j and k.

x_{ij} and x_{ik}	are the ordinate values of spectra j and k at wavelength i
p	is number of wavelengths

A perfect match between the two spectra will give a correlation coefficient of 1. In practice, random noise in spectral measurements means r values are always < 1. Some confusion exist between the two types of correlation coefficients in the literature. For example the Training Manual (FOSS NIRSystems, 1996) gives the equation for the product moment correlation coefficient, but the *NSAS* Version 3.52 software actually calculates the dot product correlation coefficient. When comparing second derivative spectra, there is no substantial difference in the results from using either of these equations. However, the product moment correlation coefficient often gives better discrimination with the raw absorbance spectra.

Correlation in Wavelength Space is easy to calculate and easily visualised , **Fig. 1.9**. Hence it renders ease of understanding by users and regulatory authorities. Despite the simplicity of its application, it is very useful for identification purposes as it can discriminate between closely related pure chemical compounds e.g. related beta lactams (Plugge and Van der Vlies, 1993) and benzodiazepine derivatives (Gerhausser and Kovar, 1997). The latter example involved the coupling of correlation with principal components analysis. Correlation coefficient has also found its use in identification of tablets in clinical trial supplies (Dempster *et al.*, 1993; Candolfi *et al.*, 1998), verification of active drugs in tablets (Khan *et al.*, 1997) and detection of the microbiological contamination in injections (Galante *et al.*, 1990).

Experience has shown that Correlation in Wavelength Space, particularly when comparing second derivative spectra, is less sensitive to baseline shifts or the subtle differences in spectral values (Mark, 1992a) and small changes in absorbance values, which reflect changes in the physical properties and the presence of impurities. Therefore, this is less suitable for qualification purposes.

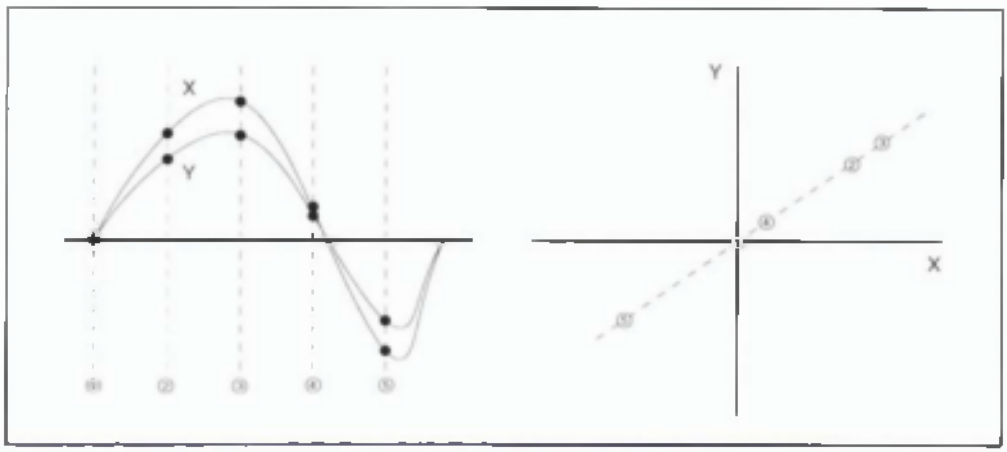


Fig. 1.9 Comparison of spectra using Correlation in Wavelength Space.

1.5.2 Maximum Wavelength Distance

Maximum Wavelength Distance is a widely used pattern recognition method and is available on the FOSS instrument software. For each product/material in the library, a representative number of samples need to be scanned so that the standard deviation at each wavelength in the spectrum can be calculated. At each wavelength the number of standard deviations corresponding to the differences between the unknown spectrum and the mean reference spectrum is calculated. The maximum values over the whole spectrum is recorded, **equation 1.21**.

$$d_{jk} = \max\left[abs\left(\frac{x_{jp} - \bar{x}_{kp}}{s_{kp}} \right) \right]_{over\ all\ p} \qquad \textbf{(1.21)}$$

Where s_{kp} is the inflated standard deviation for the n spectra in set k at wavelength p and given by **equation 1.22**.

$$s_{kp} = \left[1 + \frac{1}{\sqrt{2(n-1)}} \right] \left\{ \frac{\sum_{i=1}^{n} (x_{ikp} - \bar{x}_{kp})^2}{n-1} \right\}^{1/2} \qquad \textbf{(1.22)}$$

Where d_{jk} is the maximum distance between spectra j and k.

x_{jp} is the y-ordinate with maximum distance from between the spectra j and k

p is the number of wavelengths.

$\bar{x}_{kp}$ is the mean value at y-ordinate with largest distance.

The standard deviation tends to vary with wavelength and therefore creates an 'envelope' of acceptable region which also varies in magnitude across the spectrum, see **Fig. 1.10.** This method is more sensitive than the correlation coefficient and affords identification as well as qualification. If measurements at each wavelength have a normal distribution, then providing d does not exceed 3σ there will be a 99.7% probability that an unknown spectrum belongs to the same population as the reference spectrum. This criterion has, however, been found to be too stringent in practice and the critical d values needs to be selected on a case by case basis (Blanco *et al.*, 1998).

Using Maximum Wavelength Distance, Dempster *et al.* (1993) could distinguish between tablets of 5, 10 and 20% w/w of the same active, placebo and the clinical comparator (80% active) even when the tablets remained within the blister pack. Gemperline and Boyer (1995) used what was described as an 'enhanced' wavelength distance method, where the technique was based on a probability threshold rather than a distance threshold as described above. Their preference was on the basis that a probability threshold would be more robust to changes or updates to the training set. With the enhanced method, they were able to identify and qualify cellulose acetate phthalate, cellulose, methyl cellulose, and corn starch with a training set of between 8 – 10 spectra in the library with a probability threshold of 95%.

Plugge and Van der Vlies (1993) have described a slightly modified form of the Wavelength Distance method referred to as conformity index (CI). CI represented the modulus value at the wavelength of a spectrum that presented the greatest distance from the mean. This was used to monitor the process control of ampicillin trihydrate. From a population of 324 samples, CI values below 5 represented batches within normal process conditions. Sources of deviation i.e. below or above the normal range of active content, presence of anhydrous crystal form and residual water could also be traced using a C-PLOT, which displayed the modulus distance values *vs.* wavelengths across the entire spectrum. Gonzalez and Pous (1995) extended the idea of a C-PLOT with a DIS-PLOT. A DIS-PLOT is a C-PLOT which takes into account the direction of the deviation of the sample spectrum from the mean. This was particularly useful for derivative spectra where positive and negative peaks were present. The inclusion of the additional information was found to enhance the sensitivity of the method and could be used to detect particle size changes between different batches.

Careful optimisation of the library and instrumental control is important for the success of this method. Gerhausser & Kovar (1997) and Gonzalez & Pous (1995) warned that the method can be sensitive to noise, water peaks and wavelength shifts, all of which are potentially causes for false negative identifications. The technique

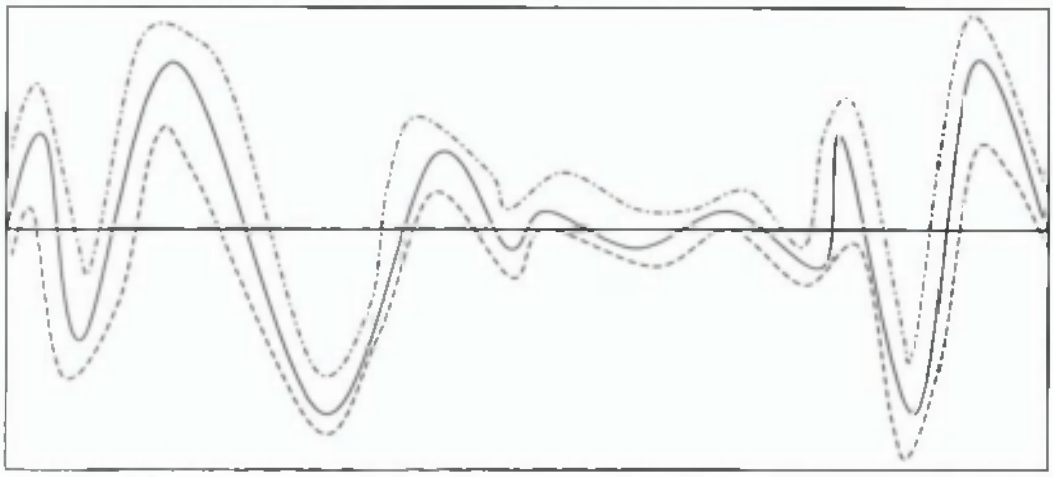


Fig 1.10 Illustration of Maximum Distance Wavelength analysis. Mean spectrum (solid line) with 'envelope' of acceptable region (dashed lines).

suffers from problems when applied to second derivative spectra, because at the cross-over points, ordinate values tend to approach to zero, resulting in standard deviations becoming artificially large. Therefore, this may not be the method of choice if the spectral library is to be transferred to other instruments.

1.5.3 Polar Qualification System

The use of the Polar Qualification System (PQS) in the pharmaceutical industry was first reported by Van der Vlies *et al.* (1995). It offers a simple method for comparing spectra by effectively reducing an entire spectrum to one quality point and also offers a simple visual presentation to the data, **Fig. 1.11**. However, as it retains all the spectral information, it is highly discriminating. In the same publication, it was also reported that PQS could differentiate between two different batches of amoxycillin trihydrate originating from two different sources, which in fact only differed in terms of their particle size.

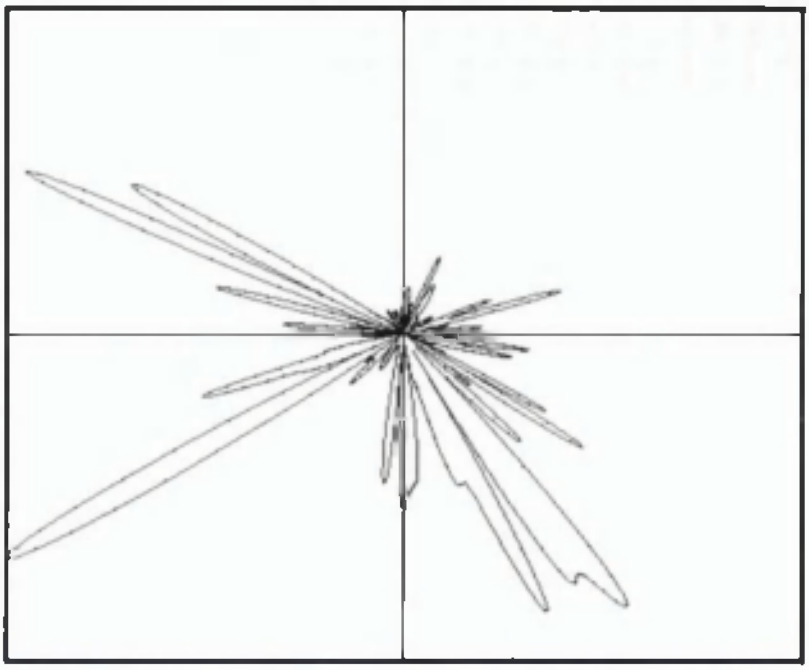
PQS involves two basic steps: (1) transformation of spectra to polar coordinates (equations **1.23** and **1.24**), and (2) calculation of the centre of gravity of the resulting polar plot (equations **1.25** and **1.26**). When repeat measurements or a range of values from different batches of material are available, a confidence ellipse enclosing specified probabilities of the population may be calculated based on a bivariate normal distribution method (Sokal and Rohlf, 1981).

$$X_i = |A_i| \cos\left(2\pi \frac{1}{n}\right) \qquad \textbf{(1.23)}$$

$$Y_i = |A_i| \sin\left(2\pi \frac{1}{n}\right) \qquad \textbf{(1.24)}$$

$$Z_x = \frac{1}{n+1} \sum_{i=0}^{n} X_i \qquad \textbf{(1.25)}$$

$$Z_y = \frac{1}{n+1} \sum_{i=0}^{n} Y_i \qquad \textbf{(1.26)}$$



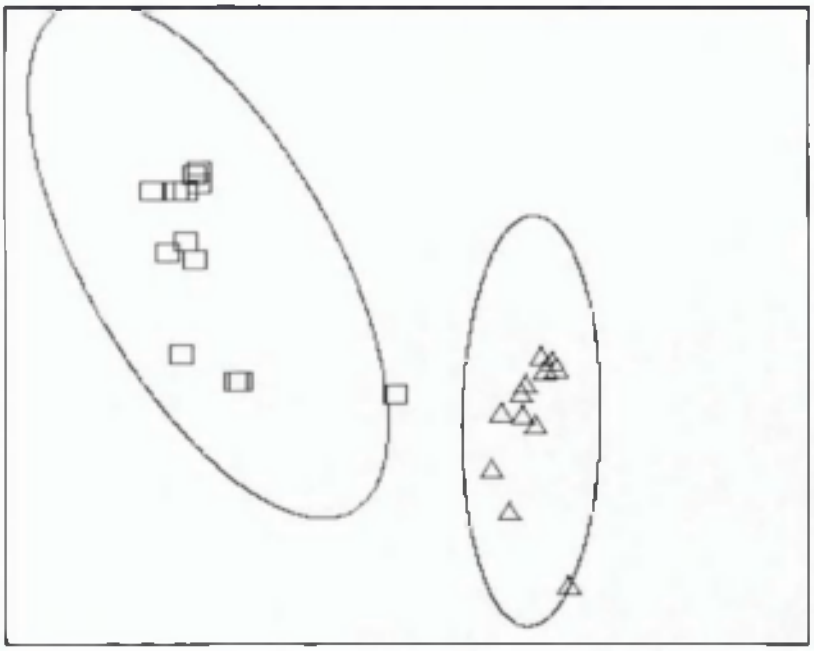


Fig 1.11 (a) Polar coordinate plot of the second derivative spectrum of povidone 30 (b) Ninety-five percent equal probability ellipse for 12 replicated measurements of a sample: (i) povidone 30 and (ii) crospovidone

Where A_i is the ordinate value at wavelength i

Z_x, Z_y are the coordinates of the centre of gravity

With its simplicity and high discriminating power, this new procedure offers enormous potential for qualification purposes.

1.5.4 Other methods

Apart from the 3 methods discussed above, multivariate analysis is also commonly used in identification and qualification processes. Amongst the methods in this category are Mahalanobis Distance and Soft Independent Modelling Class Analogy (SIMCA) which have been widely used and described in the literature. As a full description of the underlying theory is beyond the scope of this work, only the application of these methods will be discussed.

Mahalanobis Distance is a measure that accounts for the different scales of each variable and, in addition, for their correlations.

The use of Mahalanobis Distance in NIR spectroscopy was first published by Mark and Tunnell (1985) for the identification of powdered compounds. A library of 50 different compounds was constructed and challenged with 40 unknowns. The method proved to be accurate, quick and fail-safe in its classification.

Ciurczak and Maldecker (1986) applied the use of Mahalanobis Distance to dosage forms. Three component tablet granulations comprising butalbital, caffeine and aspirin were analysed. Using a 3-wavelength model, the method effectively discriminated between the different types of granulation when each of the components was absent.

SIMCA is a method that primarily discriminates between spectra by their residual variance. Gemperline *et al.* (1989) found SIMCA effective as a method for qualification, being able to detect impurities and degradation products down to a level of 0.5%. Separation of Avicel PH101 and PH102, two grades of microcrystalline cellulose which only differ in terms of particle size (mean particle size of 50 and 100

μm respectively) was found to be possible in 6 dimensions. Using the same data set, Gemperline and Boyer (1995) then compared the use of SIMCA to Mahalanobis Distance and Maximum Wavelength Distance. The study revealed two interesting findings. (1) The Maximum Wavelength Distance demonstrated superior discriminating ability between closely related materials when applied to smaller training sets (sample size < 12). (2) The Maximum Wavelength Distance method was less suitable for detecting low levels of impurities when compared to discriminant analysis (2.0% and 0.5% respectively). The latter result contradicted the outcome of another comparative study carried out by Hammond (1996). He explained that Maximum Wavelength Distance would be a more effective method for detecting impurities as the sensitivity of each wavelength is retained, whereas in SIMCA the effect on each wavelength is diluted out.

1.6 Transferability in NIR spectroscopy

Prior to the recognition of an analytical procedure as a standard method, proof of transferability between different laboratories is recognised as a crucial step in its validation (Thompson *et al.*, 1993). This can be checked through collaborative trials. With NIR spectroscopy, the issue of transferability has been brought to the forefront for one additional reason: cost. As NIR spectroscopy is considered to be a secondary technique, a spectral library that has been referenced to a primary method will have to be first constructed. To build a representative library, a large number of materials comprising different batches will have to be scanned over a period of time. Therefore, the ability to use the same spectral library on different instruments and laboratories would offer tremendous cost savings.

The regulatory concern for the lack of transferability of the technique has been reported by Moffat *et al.* (1997) and Van der Vlies (1996). At present, despite the wide use of the technique within the pharmaceutical industry, only limited NIR-based applications have been accepted as official methods (Blanco *et al.*, 1998). In the other cases, its role remains as an alternative technique.

A literature review carried out to verify the validity of the above claim by the regulatory authorities revealed two interesting points. Firstly, almost all transferability

studies reported in the literature (Bouveresse & Massart, 1996) were for quantitative applications. None has been reported for applications for identification. Secondly, these publications also claimed that, although differences between instruments affect spectral measurements, transferability could still be achieved with the use of standardisation methods. Interestingly, within the agricultural industry, instrument standardisation seems to be accepted as an integral part of a NIR-method. In a collaborative trial to validated a method for the determination of crude protein content, instrument standardisation procedures were included as part of the experimental protocol. The results showed that it was as precise as conventional methods (i.e. the combustion method and Kjeldahl) and has been adopted by the Official Methods Board of AOAC (Delwiche *et al.*, 1998). Therefore, the claim of the apparent non-transferability of the technique by regulatory authorities in the pharmaceutical industry could be a matter of mentality.

The following attempts to look at the sources of errors and correctional procedures in greater detail. Finally the present state of knowledge concerning the transferability of the technique for identification applications will be considered.

1.6.1 Sources of variation

As apparent from the discussions in this chapter, there exist a range of instrumental sources of variations that can contribute to spectral differences. This section intends to provide a recapitulation.

Within instruments from the same manufacturer, spectral variation can arise particularly from differences between reference standards for reflectance measurements. The variety of sampling options also present a challenge to transferability of spectra. For the 0-45° geometry, variations in the arrangement of detectors and sizes and shapes of sources are known to affect the spectra (see **Fig. 1.12**). The use of fibre optics results in an attenuation of signal in the higher wavelength region that typically lies above 2200 nm. Therefore, when transferring spectra

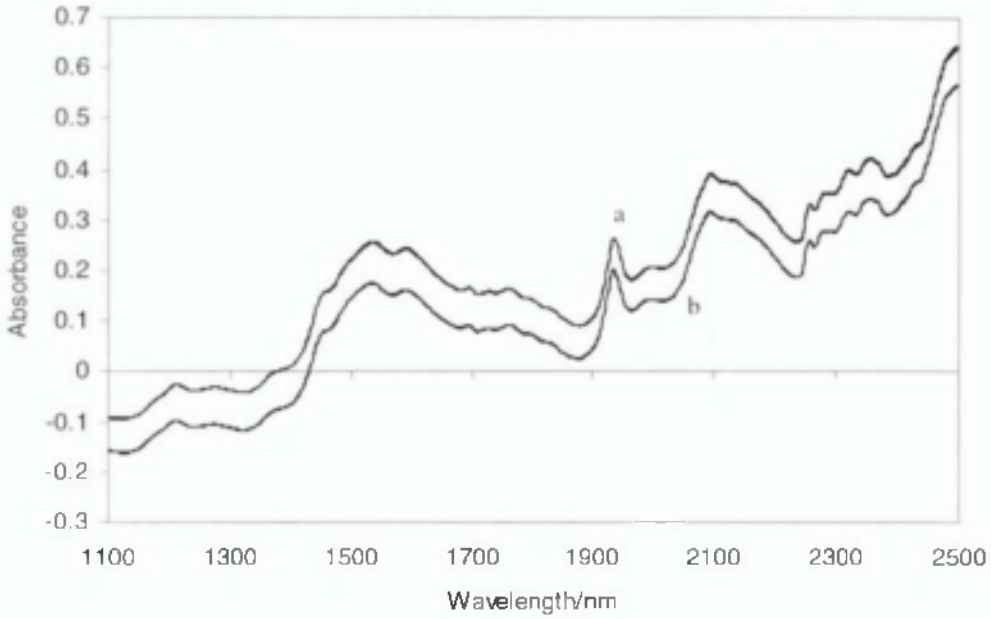


Fig 1.12 Spectrum of lactose monohydrate measured on

(a) FOSS NIRSystems, Direct Content Analyser. Four detector arranged in parallel rows in relation to an oblong radiation source.

(b) FOSS NIRSystems, Rapid Content Analyser. Six detectors arranged concentrically in relation to a circular radiation source.

between instruments of different working ranges, wavelength regions will have to be considered.

When considering instruments originating from different manufacturers, the resulting spectral differences are more complex. Instruments could differ in terms of signal-to-noise ratio or specified resolution, hence affecting spectral quality and bandwidth. When comparing instruments operating on different optical principles (e.g. FT *vs* grating), differences in spectral output could be even more complex. For example, the spectral output from an FT instrument would normally be linear with respect to wavenumber, whereas the grating instruments typically give spectra linear in wavelength. Because of the inverse relationship between wavenumber and wavelength, converting data linear in wavenumber to wavelength will give rise to unevenly spaced data-points. Pattern recognition methods for identification purpose would not be able to directly compare such data-sets.

Time and ambient conditions affect spectral measurements too. Lamp aging and changes of detector responses over time can cause a decrease in the signal-to-noise ratio. Characteristics of samples from a production cycle may change too and will have to be constantly built into the spectral database. Temperature is known to affect inter-molecular forces and hence the vibrational spectra. The influence is most evident in liquids, where an increase in temperature would cause hydroxyl band shifts to lower wavelengths and also give sharper peaks. The S-H and N-H stretch bonds are also affected in a similar manner (Wulfert *et al.*, 1998). The Training Manual from FOSS NIRSystems (1996) also mentions that increases in temperature can affect the diffraction grating, resulting an increase in the signal-to-noise ratio. Davies and Grant (1987) reported that water vapour arising from air conditioning introduced significant noise which exceeded the manufacturer's specified level.

Generally, it is difficult to systematically translate the effects of each source of variation to a particular spectral effect i.e. constant wavelength shifts or a consistent baseline offset. Instead, spectral differences are complex, the extent of wavelength shifts *etc.* varying across the spectrum (Bouveresse *et al.*, 1997). Therefore,

standardisation procedures particularly of a multivariate nature have been commonly used.

1.6.2 Standardisation procedures

Approaches to standardisation procedures vary considerably (Bouveresse and Massart, 1996). They can involve the use of data-pretreatment methods to correct for simple baseline shifts (Swierenga *et al.*, 1998), the selection of wavelength regions that are robust to instrumental differences (Mark and Workman, 1988) or the use of correctional procedures that can be univariate or multivariate (Dreassi *et al.*, 1998; Bouveresse *et al.*, 1996a).

Swierenga *et al.* (1998) reported that the use of pre-treatment methods such as first derivative, second derivative and multiplicative scatter correction can be as effective as correctional procedures. Mark and Workman (1988) developed wavelength search methods to select wavelength regions with high information which remained robust to instrumental differences for the construction of multiple linear regression models. The two methods described above require less calculation and no extra measurements when compared to the use of standardisation samples. However, when spectral differences become more complex, correctional methods may be preferred.

Jones *et al.* (1993) successfully transferred calibrations for moisture determination using a simple slope/bias correction. In this case, two independent data-sets were measured on two different NIR instruments from the same manufacturer. By using a calibration model constructed on the second instrument to predict *the calibration samples measured on the first instrument*, a univariate linear model was computed to correct the predicted values. With the application of this correction, *test samples measured from the first instrument* were correctly predicted with the second instrument. A reversal of the above process for the two instruments gave similarly good predictions.

With instruments where differences are complex (e.g. different optical systems, different signal-to-noise ratio), multivariate correction may be necessary. This usually involves measurement of 'standardisation samples'. These are samples which are to

be measured on primary and secondary instruments and should also resemble the samples in question. From these standardisation samples, a 'transfer matrix' can be generated. A transfer matrix may be generated for the complete spectrum or just a specific wavelength range of interest.

An alternative is the patented method proposed by Shenk *et al.* (1991), that corrects for the wavelength error and the spectral intensity.

1.6.3 Present state of knowledge and aims

It is apparent from the above discussions that there is a gap of knowledge concerning transferability for identification procedures. There is no evidence in the literature to indicate if this is possible or not. Many sources of spectral variation have been identified, although presumably there are still others to be looked into. Therefore, this project intends to take a more systematic approach and try to study the effects of identified sources of variation on a variety of identification algorithms. Three identification algorithms have been chosen for this study: Correlation in Wavelength Space, Maximum Wavelength Distance and Polar Qualification System. The choice of the algorithms are based on their wide range of sensitivities.

In summary, the aims of this work are:

- to develop an understanding on the effects of sources of variation on commonly used identification algorithms
- by minimising sources of variation and optimising wavelength parameters, to attempt to be able to demonstrate the transferability of a fully validated method
- to investigate the feasibility of using novel methods of identification and see if they are transferable.

CHAPTER 2 : GENERAL METHODS

This chapter provides a brief overview of the instruments and software used in all the studies. Specific information for the various investigations is provided in the respective chapters.

2.1 Instrumentation for near-infrared spectroscopy

Instruments from three different manufacturers were used: FOSS NIRSystems, Buhler and Bran+Luebbe. The first was a dispersive grating instrument and the other two were Fourier-Transform instruments. All instruments were loaned from the respective manufacturers. The technical specifications for the instruments are summarised in **Table 2.1**. To extend the range and number of instruments used, those from pharmaceutical laboratories at SmithKline Beecham, Harlow and Tonbridge sites, and GlaxoWellcome, Ware were also used.

Table 2.1 Technical specifications of instruments used

	FOSS	Buhler	Bran+Luebbe
Wavelength range	400/1100 – 2500*	1000 – 2500	1000 – 2500
Usable wavelength range			
Fibre optic probes	400/1100 – 2200*	1000 – 2200	1000 – 2200
Horizontal setups	400/1100 – 2500*	-	1000 – 2500
Noise level (RMS)			
Fibre optic probes	< 0.1 milli A.U.†	1 milli A.U. ‡	0.5 milli A.U.‡
Horizontal setups	< 0.04 milli A.U.†		
Wavelength accuracy	± 0.3 nm	-	-
Wavelength repeatability	± 0.01**	-	-
Photometric range	3 A.U.	2 A. U.	2 A. U.
Spectral bandwidth	10 nm, + 1 nm, - 2 nm	-	-
Resolution (without apodization)	-	At 1000 nm ± 1.2 nm At 2500 nm ± 8.0 nm	At 1000 nm ± 0.6 nm At 2000 nm ± 2.5 nm
Scan time	0.56 second	-	1.1 second
Stray light	< 0.1% at 2300 nm	-	-

* dependent on model of monochromator

† calculated as an average of 32 scans

‡ calculated at 0 A. U. of and for 1 scan

** calculated as thestandard deviation of 10 consecutive scans

RMS root mean square

2.2.1 FOSS NIRSystems (Silver Spring, MD, USA)

FOSS monochromators use holographic gratings. Model 5000 allows measurements over the range 1100 to 2500 nm; while the model 6500 monochromator is useable over the range 400 to 2500 nm. Both monochromators are fitted with a tungsten halogen lamp as radiation source.

A number of accessories can be attached to these monochromators.

Rapid Content Analyser (RCA, AP 1300)

This is a horizontal setup, **Fig. 2.1A** used for the measurement of solids by reflectance or liquids by transflectance. The RCA bolts onto the side of the monochromator and the source radiation transmitted via a short length of fibre optics. Six lead sulphide and silicon detectors are arranged concentrically around the source. Samples are placed in sample cups and placed on the sampling stage. An iris diaphragm is used to centre the sample cup in a reproducible manner. Reference measurements for diffuse reflectance are made using ceramic plate which can be positioned over the sample stage.

Direct Content Analyser (RCA, AP 1200)

This is a horizontal setup mounted directly on top of the monochromator and requiring no fibre-optic connection. The radiation source is oblong shaped and 4 detectors are arranged on either side and parallel to the source. Samples are placed in sample cups on the sample stage. A ceramic plate is used as reference.

Fibre optic probes

These consist of a stainless steel tube fitted with a sapphire window at one end and connected to the monochromator via a length of fibre optics. The probes may be inserted directly into the sample to be measured. To measure liquids by transflectance a reflector tip is fitted. Reference measurements for diffuse reflectance are taken using a ceramic tile fitted inside the holding block of the probe. With the use of fibre optics, attenuation of the signal reduces the useable wavelength range to below 2200 nm.

Interactance Immersion Probe (NR 6770)

This is designed for liquids and uses a bundle of 210 illuminating and collection fibre made from ultra-low moisture silica. The illuminating and collection fibres are arranged concentrically and are in direct contact with the inside of the sapphire window. The outer diameter of the probe is 19 mm.

Interactance Reflectance Probe (NR 6775)

As above for Interactance Immersion Probe but designed for highly scattering liquids. Outer probe diameter 13 mm.

Smartprobe/Optiprobe (NR 6770)*

This type of probe is designed for raw material testing where the measurements can be initiated at different locations within a warehouse, using a trigger located on the probe head, **Fig. 2.1B**. According to the manufacturer, the Smartprobe or Optiprobe have largely replaced the Interactance Immersion Probe in its use, although bearing the same product number.* It consists of a fibre optic bundle made up from 110 illuminating and collection fibres. Probe outer diameter is 127 mm.

Liquid Transmission Module (NR 6509)

This module is designed for measurement of liquids in cuvettes with path-lengths of 0.5, 1, 2, 4, 10, 20 or 30 mm.

2.2.2 Buhler Anatec (Uzwil, Switzerland)

The NIRVIS spectrophotometer is a Fourier Transform NIR instrument (model 100.1). Spectra are generated using a polarization interferometer fitted with a temperature–stabilised lead-sulphide detector. In this work, the NIRVIS was fitted with a Buhler fibre-optic probe (model 1359) and transflectance cap. The optical path-lengths can be adjusted between the range of 1 to 6 mm at intervals of 1 mm using distance rings. Spectralon® was used as the reflector and also for reference measurement. A sapphire window protects the Spectralon® from the possible damaging effects of liquid samples.

2.2.3 Bran+Luebbe (GmbH, Norderstedt, Germany)

The InfraProver II (124-A020-01) is a Fourier Transform Polarisation Interferometer. Two accessories were used: the diffuse reflectance probe (123-b006-01), **Fig. 2.1D** and Sampling Presentation Accessory (a horizontal setup, 124-A020-01), **Fig. 2.1C.** The diffuse reflectance probe can to used to measure liquids by transflectance by fitting a transflectance cap (124-B603-02) to the tip of the probe.

2.3 Software

Both proprietary software and in-house programs were used for data analysis.

2.3.1 Proprietary software

Instruments were controlled by the manufacturers' software. The software used were: *NSAS* Version 3.52 for the FOSS instruments, *NIRCal* Version 3.0 for the Buhler instrument and the *ICAP* Version 5.10 and *Sesame* Version 2.1 software for the Bran+Luebbe instrument. All the software packages provided facilities to perform a variety of data pre-treatment algorithms for qualitative and quantitative analysis. Facilities to import and export spectral files in various formats were also available.

2.3.2 In-house programs

A number of in-house programs were written to assist with data analysis i.e. calculation of data pre-treatments and identification algorithms. Although some of these facilities were available from the commercial software packages, in-house programs were preferred because:

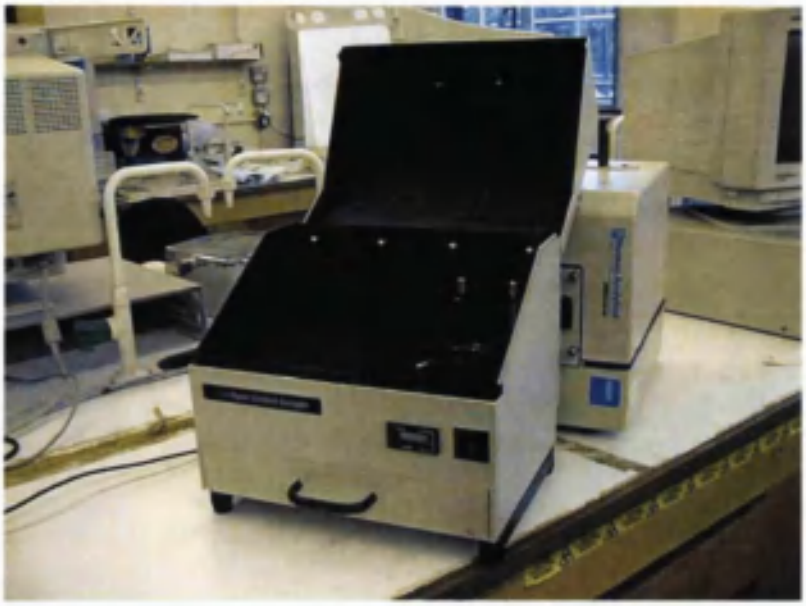
- The exact formula used for data transformation or the identification algorithms were not always specified or available. For example, the *NSAS* (Version 3.52) software calculates the Correlation in Wavelength Space using the dot product correlation coefficient even though the Training Manual (FOSS NIRSystems, 1996) shows the formula for product moment correlation coefficient.
- Variations of calculation methods exist between different instruments and there is a need to ensure uniformity of such calculations when comparing spectra. For example, the methods used to calculate derivative varies between different software packages.

- Variations in spectral formats exists between different manufacturers. As a result, data analysis of spectra from different manufacturers cannot be perform on a single proprietary software. This is discussed in further detail in chapter 4.
- In-house programs can be written to perform specific investigation not afforded by proprietary software i.e. calculate peak positions and perform Polar Coordinates transformation.

A list of in-house programs used and their functions is given in **Table 2.2.** Copies of all these programs are provided on the CD-ROM in directory programs. All programs were written in Microsoft C/C++ Version 7.0. Copies of the source code as well as an executable version are provided. Each program is provided with a short 'ReadMe' file describing how to use the program.

Table 2.2 In-house programs

Name	Function
DNA	To calculate Polar Qualification Coordinates and calculation of Centre of Gravity.
DERIV	To calculate derivatives by simple difference.
CORR	To calculate Correlation in Wavelength Space, r, with the option of using the dot product or product moment formula.
IDENT	To identify unknown samples from a library.
PEAK	To calculate the peak positions in spectra using quadratic interpolation.
DECON	To study the effects of band-pass by applying triangular filters of varying sizes
STRAY	To correct spectra for stray radiation and calculate the reflectance with respect to a given reference material.

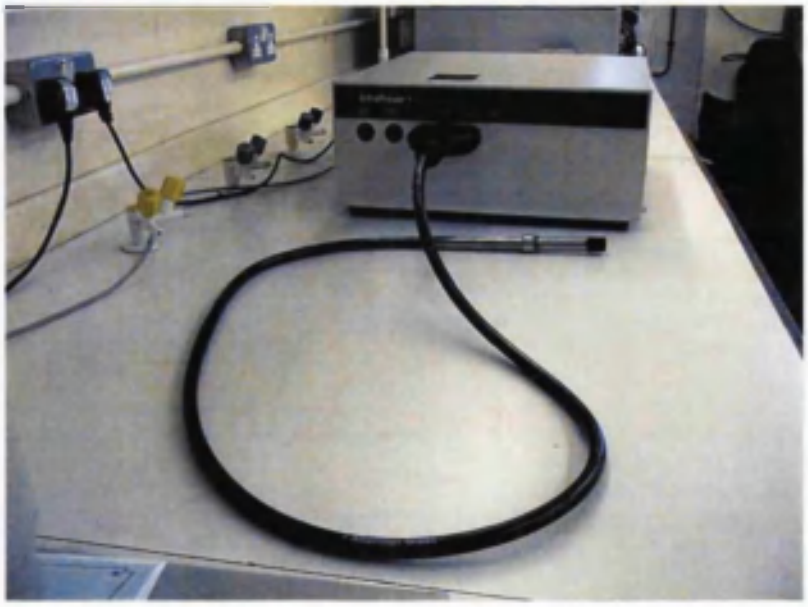


(A) FOSS NIRSystems, 6500 Spectrophotometer, Rapid Content Analyser

(B) FOSS NIRSystems, 6500 Spectrophotometer, Smartprobe

Fig. 2.1 Instruments used in the study.

(C) Bran & Luebbe, InfraProver II Spectrophotometer, with Sample Presentation Accessory.



(D) Bran & Luebbe, InfraProver II Spectrophotometer, with diffuse reflectance probe.

Fig. 2.1 Continued

CHAPTER 3: CONSTRUCTION AND TRANSFERABILITY OF A SPECTRAL LIBRARY FOR THE IDENTIFICATION OF COMMON SOLVENTS BY NEAR-INFRARED SPECTROSCOPY

3.1 Introduction

The use of NIR reflectance spectroscopy for identification purposes requires the setting up of a library of spectra containing all the materials of interest and preferably those of other closely related substances. The setting-up and evaluation of such a library is a time consuming and costly procedure. It is therefore highly desirable that a library once established should be transferable between different instruments. Nevertheless, it is generally recognised that calibrations for quantification purposes are not always transferable between different instruments, even of the same type and from the same manufacturer, as mentioned in chapter 1, section 1.6. Differences in reflectance spectra between instruments can arise from many factors; wavelength accuracy (Mark and Workman, 1988), detector linearity (Miller, 1993), band-pass, stray light, sample presentation (Mark and Tunnel, 1985; Williams, 1992) and most importantly the reflectance standard used (Springsteen and Ricker, 1996; Knee and Deadman, 1998). Varying degrees of success in transferring data have been reported. Jones *et al.* (1993) successfully transferred a calibration for moisture determination between instruments of the same model using a simple slope/bias correction. However, in most other cases reported in the literature, various more complex standardisation approaches have been developed to correct for instrumental differences (see chapter 1, section 1.6.2). It is still to be established if libraries of reflectance spectra for identification purposes are transferable.

The problems of differences of reflectance standards are largely avoided in transflectance measurements on liquid samples as spectra may be referenced to air. Also, the spectra for liquid samples such as solvents are generally less complicated than solid samples, with fewer absorption bands and more consistent baselines (as they are less affected by variation of sample presentation such as packing density for powders or orientation for solids in general). This allows greater ease for a systematic investigation into the effects of sources of variation on the spectrum.

In this work, a library of spectra of fifteen commonly used solvents has been constructed and the optimum wavelength range, mathematical processing *etc.* evaluated for distinguishing between the solvents. Also, the robustness of the library against factors such as wavelength accuracy, band-pass differences, path-length changes, temperature fluctutations and the presence of water were examined. The

transferability of the spectral library was then tested on eight different instrumental setups. These setups differed in terms of models of monochromator, types of sampling accessory (fibre-optic probe *vs* sampling accessory) and optical configuration (scanning monochromator *vs* Fourier Transform).

To compare spectra, Correlation in Wavelength Space has been used. This simple method has the advantage of depending only on the shape of the spectra and not on the absolute magnitude of the response and consequently was expected to be insensitive to path-length changes. The dot product correlation coefficient, r, is given by **equation 1.19** which has been described in Chapter 1, section 1.5.1.

When trying to identify substances there will always be some probability that false positive and false negative identifications will occur (Ellison *et al.*, 1998). By optimising the spectral wavelength range used and the mathematical pre-treatment of the spectra these errors were minimised.

3.2 Experimental

3.2.1 Materials

All solvents were of 99% or greater purity and used as obtained directly from the suppliers. The following solvents were obtained from Fisher Scientific UK (Leicestershire, UK): acetone (Specific Laboratory Reagent), butan-1-ol (Analytical Reagent), butan-2-ol (Analytical Reagent), chloroform, dichloromethane, absolute ethanol, ethanol (96%), ethyl acetate, ethyl methyl ketone, Industrial Methylated Spirit (64 OP and 74 OP), methanol, propan-1-ol (Specific Laboratory Reagent), propan-2-ol, tetrahydrofuran (Specific Laboratory Reagent) and toluene (Analytical Reagent). Additional samples and other solvents were obtained as follows: acetonitrile and propan-2-ol (Romil Ltd., Cambridge, UK, Super Purity grade), chloroform (Spectrosol grade), dichloroethane, n-heptane, iso-pentane and n-pentanol (BDH Merck Ltd., Lutterworth, UK, Spectrosol), benzyl alcohol (Puriss), dichloromethane and dimethylformamide (ACS grade for UV spectroscopy) (Fluka Chemicals, Dorset, UK), absolute ethanol (Hayman Ltd., Essex, UK), octane and nonane (Avocado Research Chemicals, Lanchashire, UK), decane, pentane, n-pentanol and undecane (Aldrich Chemical, UK), and hexane (HPLC grade) (Rathburn Chemicals Ltd., Peebleshire, Scotland, UK).

3.2.2 Instrumentation

Eight different near-infrared instrumental setups were used from 4 different laboratories. Setups 1 to 6 were all based on FOSS NIRSystems (Silver Springs, USA) instrumentation.

Setup 1: 6500 scanning spectrophotometer fitted with an Interactance Immersion Probe (NR 6770). Setup 2: 6500 scanning spectrophotometer fitted with an Interactance Immersion Probe. Setup 3: 6500 scanning spectrophotometer fitted with a Smartprobe (NR 6770). Setup 4: 5000 scanning spectrophotometer fitted with a Smartprobe. Setup 5: 6500 scanning spectrophotometer (same monochromator as used in setup 3) fitted with a Rapid Content Analyser (AP 1300) and liquid sampling kit (NR 6544) comprising of a Pyrex glass cell (40 mm diameter) and gold reflector (2 × 0.5 mm optical path-length). Setup 6: 6500 scanning spectrophotometer fitted with an Interactance Reflectance Probe (NR 6775). Setup 7: Bran+Luebbe (GmbH, Norderstedt, Germany) Infraprover II Fourier Transform polarisation interferometer (124-A020-01) fitted with a diffuse reflectance fibre optic probe with transflectance cap (124-B603-02). Setup 8: Buhler FT-NIR NIRVIS spectrophotometer (model 100.1, Uzwil, Switzerland) fitted with a Buhler fibre-optic probe (model 1359) and transflectance cap. All FOSS probes were fitted with a transflectance probe tip.

3.2.3 Measurement of spectra

All spectra were measured by transflectance with air as reference. For all probes the distance between the probe tip and the reflector was set using a 1 mm gauge. Solvents were poured into small sample bottles and the probe with transflectance probe tip immersed. The probe and reflector were dismantled and cleaned between each measurement. With setups 1 to 6, each recorded spectrum was the average of 32 scans and measured over the range 1100 to 2498 nm (2 nm intervals, 700 data-points). For setup 7 the recorded spectra were the average of 6 scans and measured over the range 4500 to 9996 cm^{-1} (12 cm^{-1} intervals, 459 data-points). Spectra for setup 8 were measured over the range 4000 to 9996 cm^{-1} (12 cm^{-1} intervals, 500 data-points) and each recorded spectrum was the average of 6 scans.

3.2.4 Data treatment

All spectra were exported from the manufacturers' software used to run the instruments as JCAMP files. Dot product correlation coefficients (see section 1.5.1 for equation) were calculated using a simple in-house written computer program, IDENT.

Derivative spectra were calculated by simple difference procedures using an in-house written computer program, DERIV. Equal size blocks of data-points before and after the data point at which the derivative was to be calculated were averaged and the derivative taken as the simple difference in means. The process was repeated across the whole spectrum a number of times until the required derivative was obtained. With derivative spectra some data-points were lost at each end of the spectrum and this was equal to the block size times the derivative used.

Cubic spline interpolation was used to convert spectra measured in wavenumber to equally spaced data-points in terms of wavelength. The wavenumber *vs* transmittance spectra were first converted to absorbance spectra and then absorbance values at 2 nm intervals over the range 1100 to 2498 nm to give 700 calculated values. Out of range values were set to zero. Cubic spline interpolation was performed using an in-house computer written program called CUBIC, which is based on the functions *spline* and *splint* (Press *et al.*, 1992). Simulations of band-pass effects were performed using an in-house computer program, DECON, based on the function *convlv*, (Press *et al.*, 1992).

3.3 Results and discussion

The spectra of all fifteen solvents (**Table 3.3,** page 95) were obtained for all eight setups with the exception of setup 4 for which the ethanol (96%) spectrum was not available. All spectra were measured with respect to air as the reference. In general, different batches of solvents were used with each setup. The correct identity of the solvents used with setup 1 were also confirmed by recording their mid-IR spectra and comparing them to reference spectra (Pouchet *et al.*, 1985).

The instrumental setups varied widely; 1-6 were grating instruments, while 7 and 8 were FT instruments. Setups 1 and 2 used an Interactance Immersion Probe designed for liquids, setup 6 an Interactance Reflectance Probe designed for highly scattering

solutions and slurries, setups 3 and 4 used Smartprobes designed for raw materials (powders), and setup 5 used a sample cup and reflector. **Fig. 3.1** shows the spectrum of dimethylformamide (DMF) as measured on all eight setups. DMF like most of the solvents studied, exhibited strong absorption bands above about 2200 nm. Useful spectral information above this wavelength was not generally obtained when using a fibre-optic probe as compared to the spectra measured using the RCA. Spectra for all the library solvents measured using a fibre-optic probe and using the RCA are shown in **Fig. 3.2** and **Fig. 3.3** respectively. The probe used with the Bran+Luebbe instrument restricted the wavenumber range to above 4500 cm^{-1} (i.e. < 2222 nm). Original spectra from both the FT instruments were in terms of transmittance *vs* wavenumber. After conversion to absorbance, a cubic spline interpolation was used to produce equally spaced wavelength data so as to match the spectra measured on the other instrumental setups. The original wavenumber and converted spectra for DMF are shown in **Fig. 3.1B** and **Fig. 3.1A** respectively. **Fig. 3.4** shows the spectra for all 15 solvents measured on an FT instrument linear to the wavenumber scale. Therefore, any transferable library clearly needs to take into consideration that spectra measured from different instruments can be stored using different scales and number of data-points. It was decided to use the spectra from setup 1 as the reference set and use these to find the optimum conditions for distinguishing between the solvents.

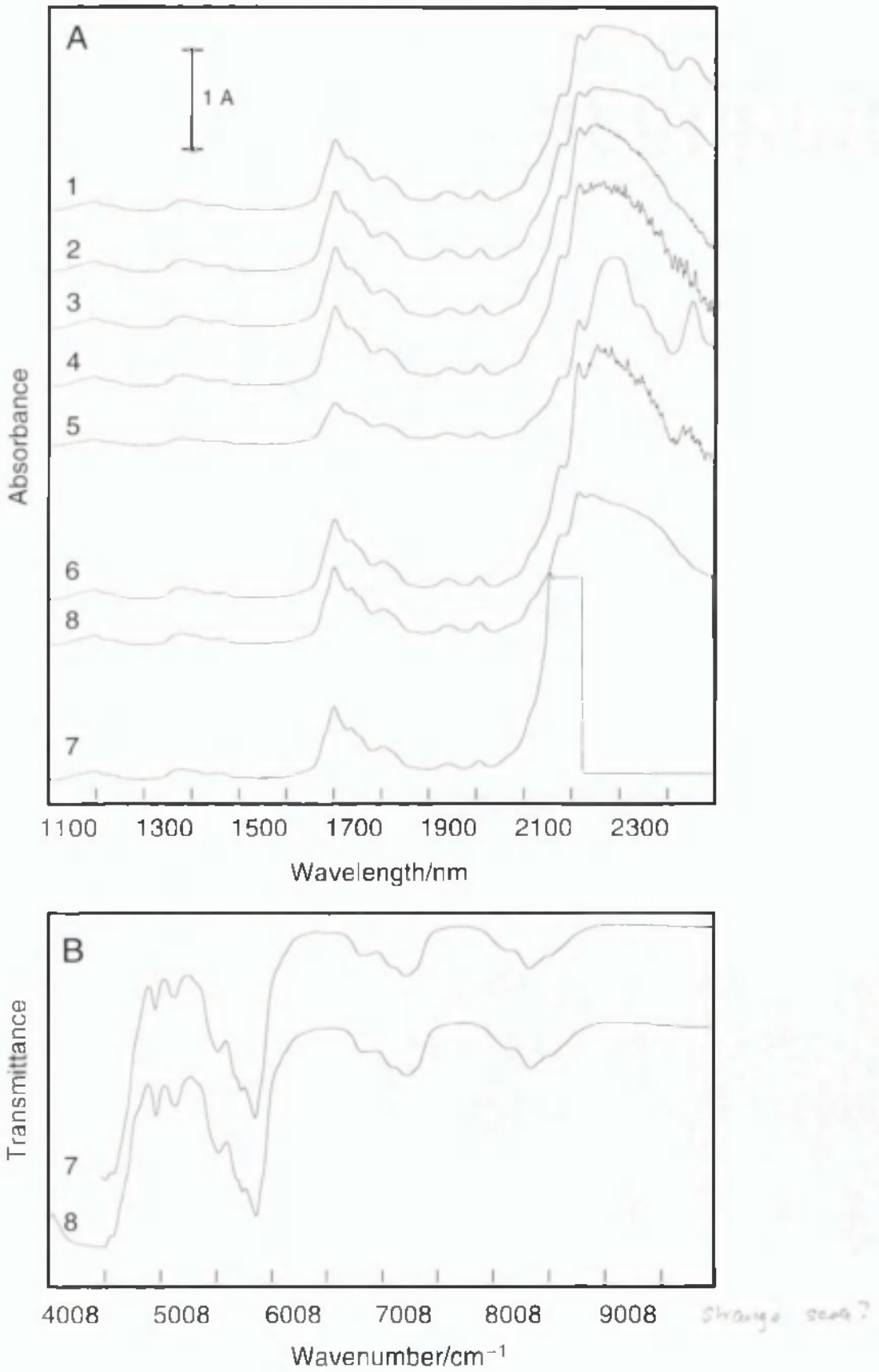


Fig. 3.1 NIR spectra of dimethylformamide as measured on the eight instrumental setups. (A) linear wavelength scale, (B) original wavenumber scale spectra for setups 7 and 8. Number indicates instrumental setup. Spectra have been vertically displaced for clarity.

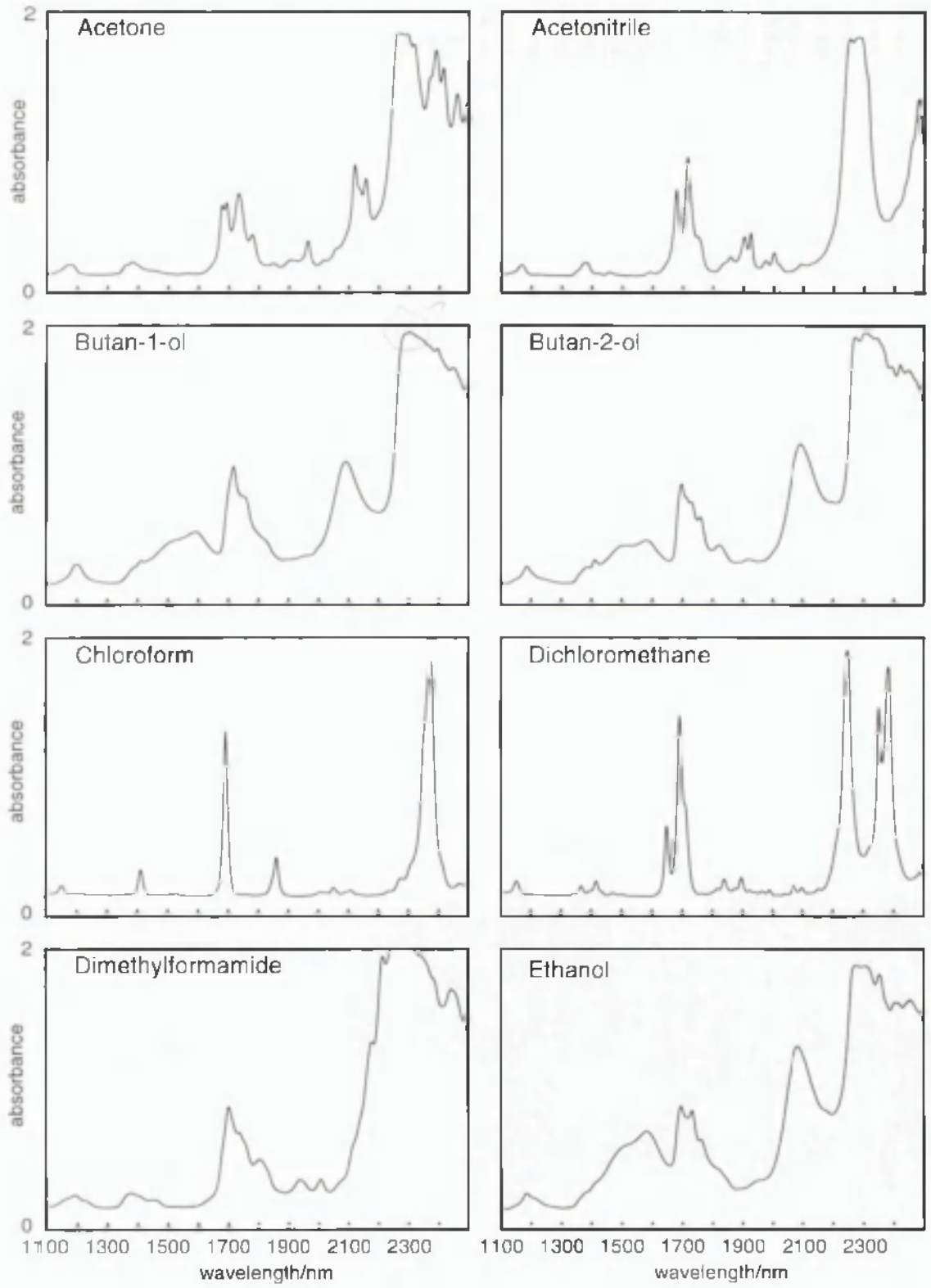


Fig. 3.2 Absorbance spectra of the 15 library solvents using a fibre optic probe (setup 1 - FOSS 6500 diffraction grating spectrophotometer with Interactance immersion probe).

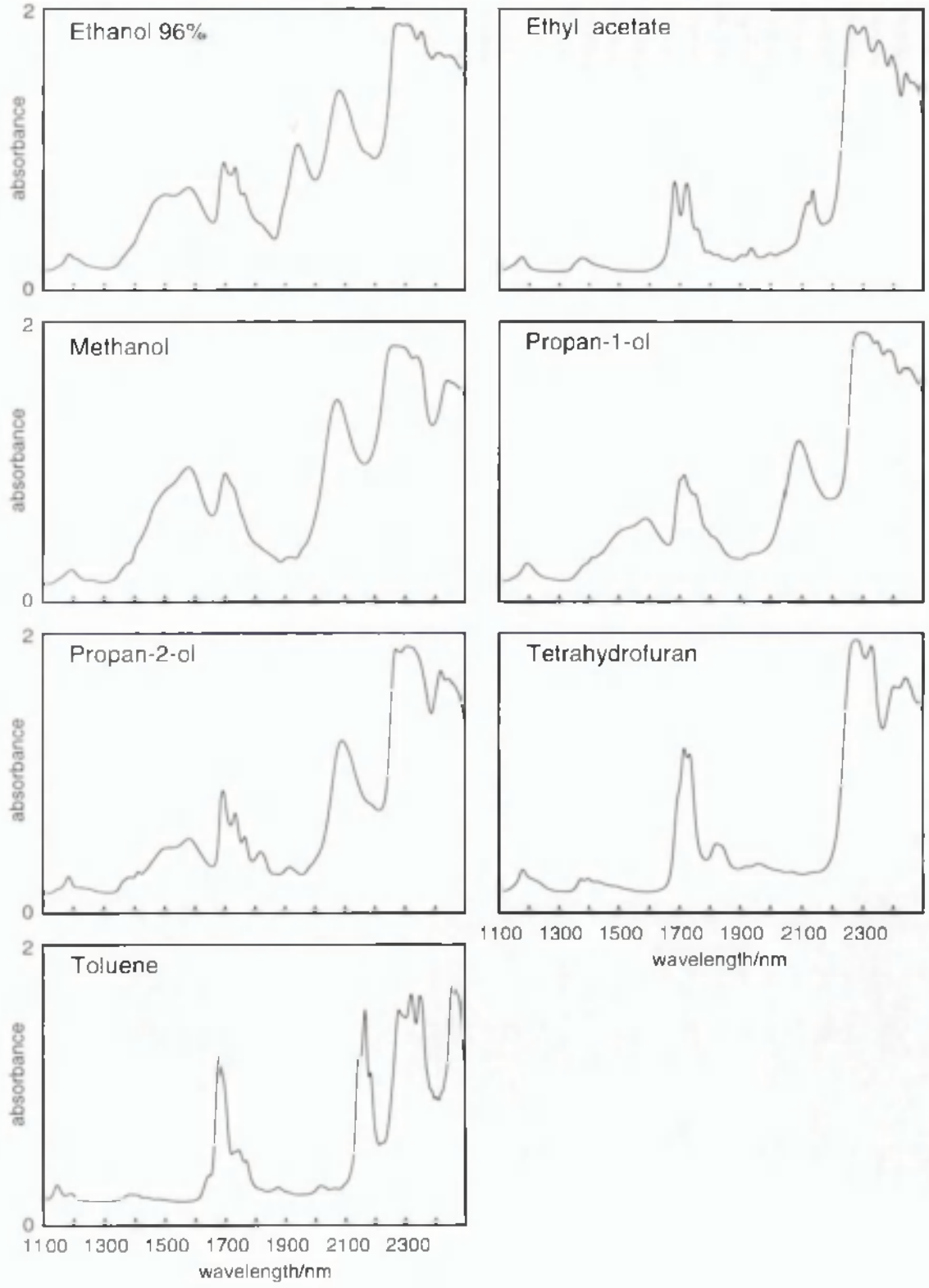


Fig. 3.2 Continued.

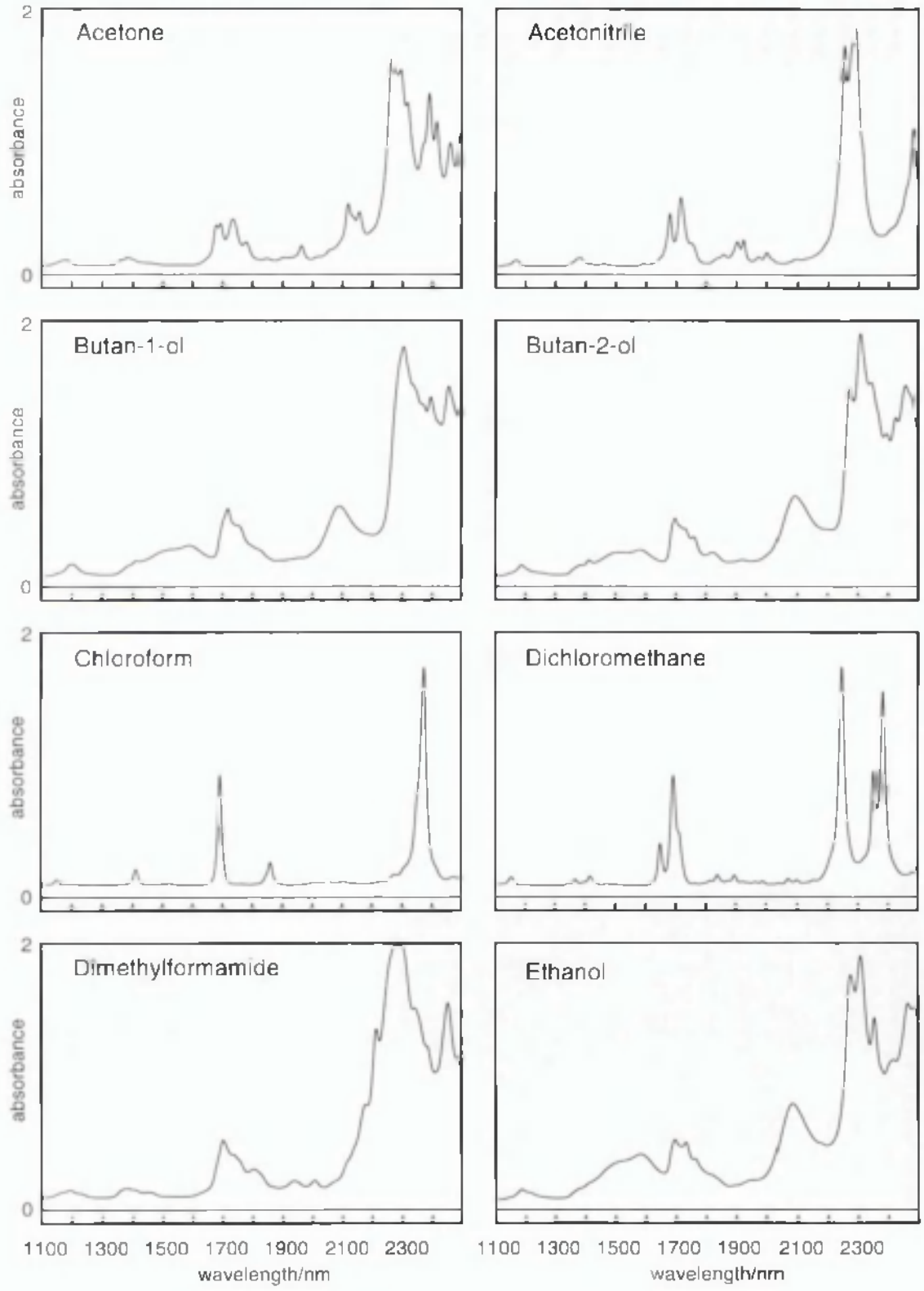


Fig. 3.3 Absorbance spectra of the 15 library solvents measured using a horizontal setup accessory (setup 5 - FOSS 6500 diffraction grating spectrophotometer with a Rapid Content Analyser and liquid sampling kit).

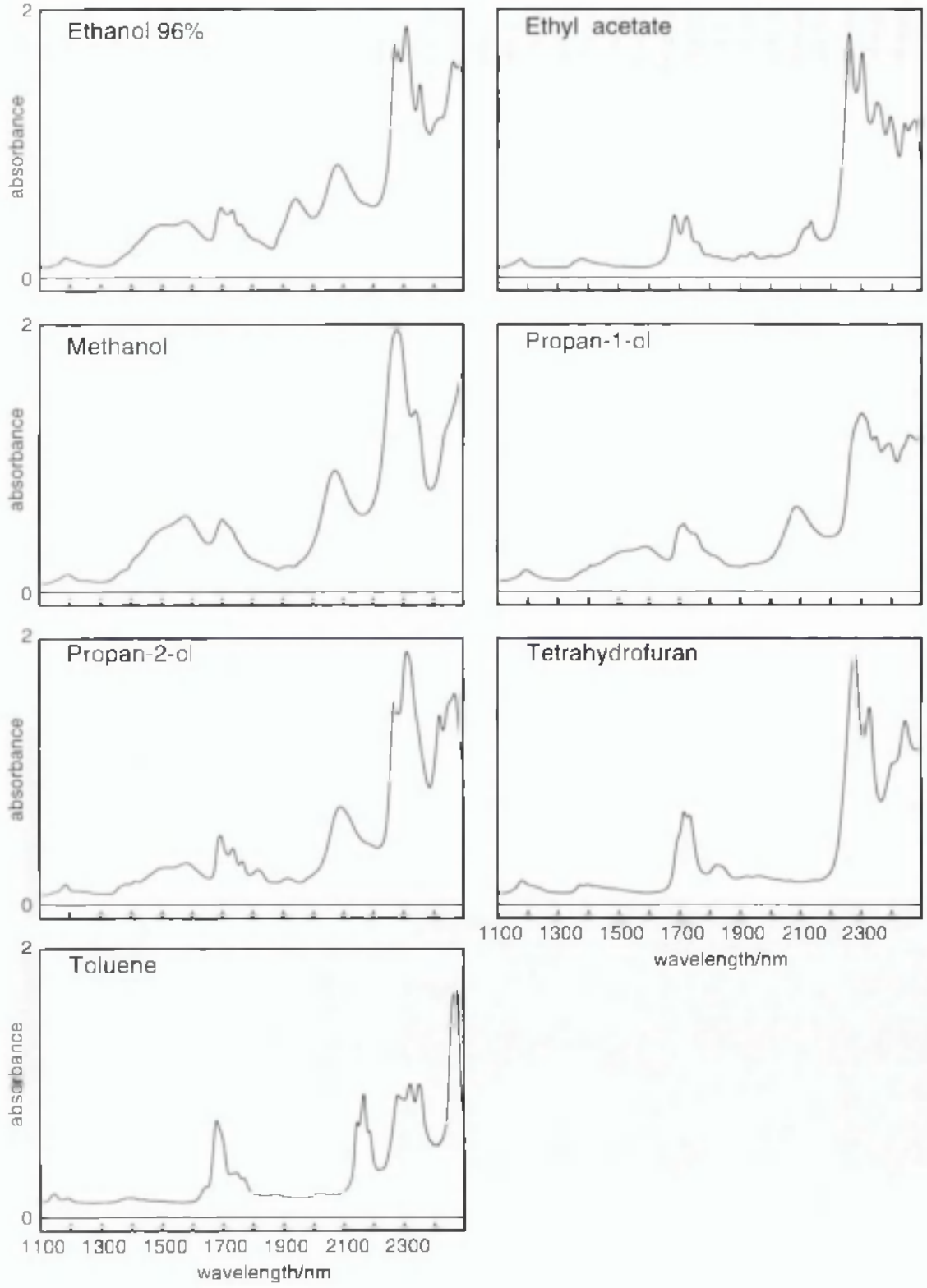


Fig. 3.3 Continued.

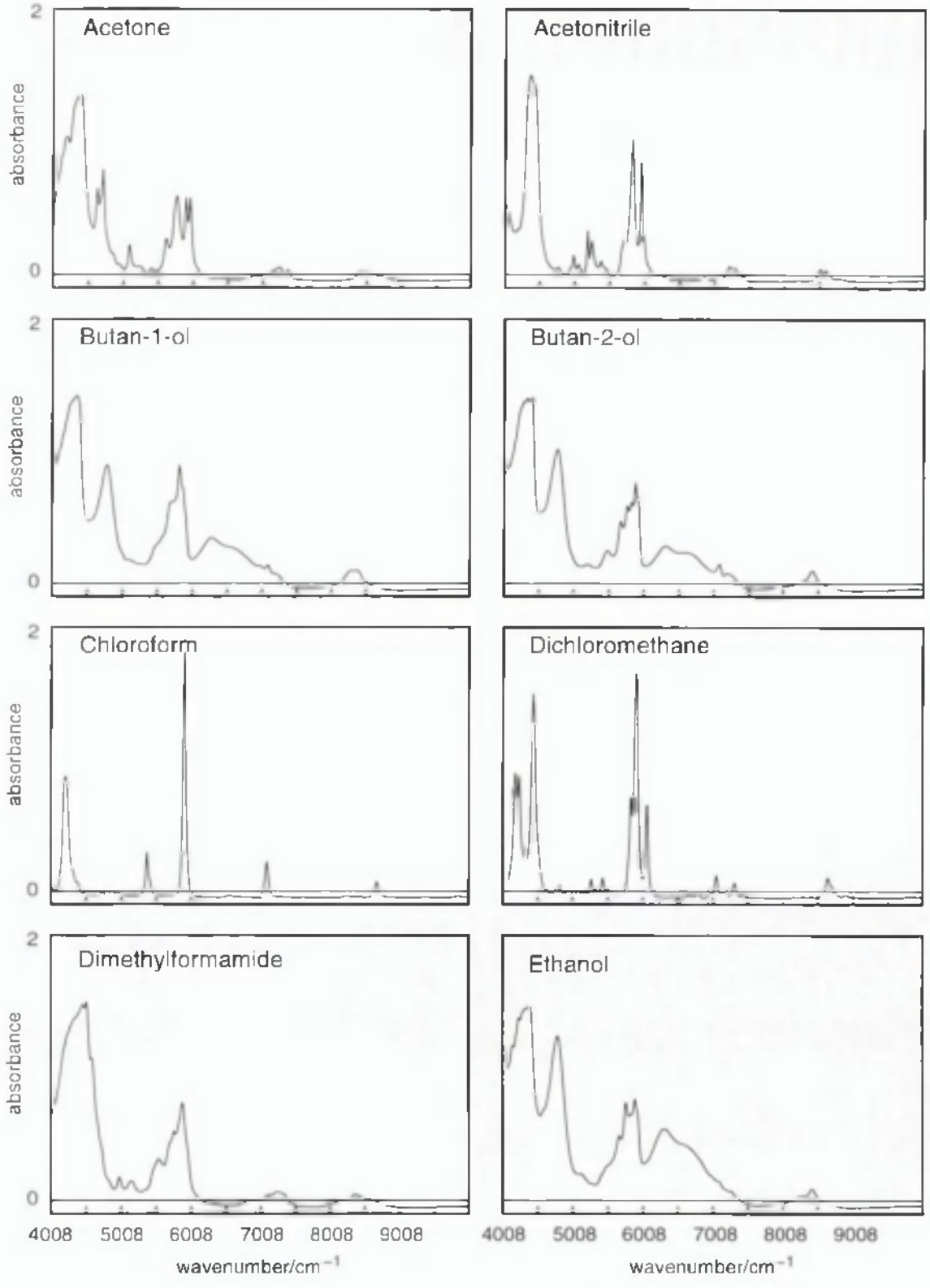


Fig. 3.4 Absorbance spectra of the 15 library solvents measured using a FT spectrophotometer and fibre optic probe accessory (setup 8).

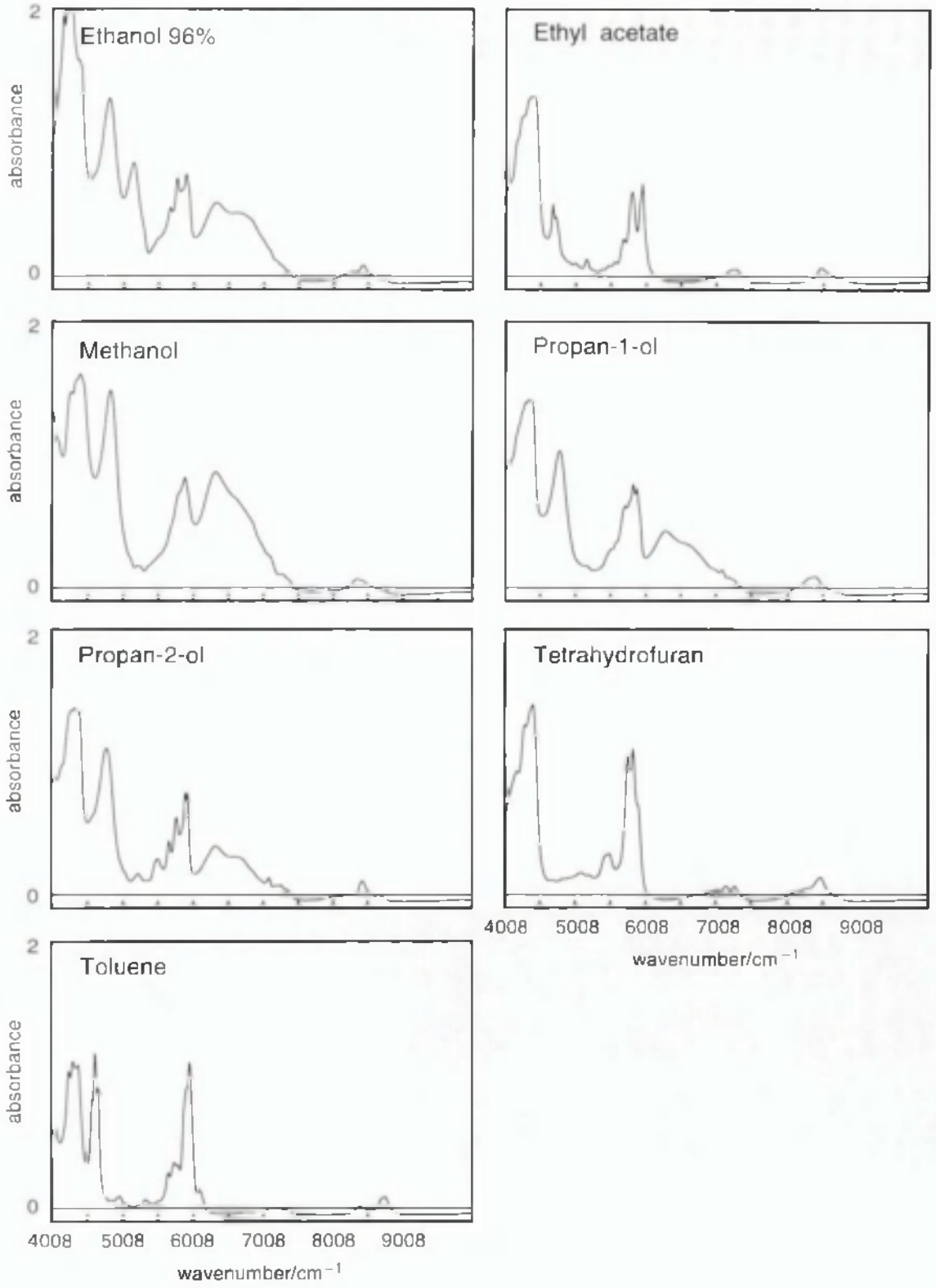


Fig. 3.4 Continued.

Three parameters were investigated to find the optimum conditions: (1) wavelength range, (2) derivative, and (3) data smoothing. Optimum parameters were taken as those that showed the best discrimination between the solvents as measured by the dot product correlation coefficient. Correlation coefficient values were calculated for all the combinations of different pairs of solvents in the library. Data smoothing and taking derivatives while decreasing the information content of a spectrum can, however, increase the ability to distinguish between different spectra. **Fig. 3.5** illustrates the different derivatives for the spectra of acetone and chloroform. The solvents were chosen to cover both broad and sharp spectra. The effects of varying block sizes of data-points on these spectra are also shown (see **Fig. 3.6**). Increasing the block size improves the signal-to-noise ratio, but decreases the resolution.

The correlation coefficients between the original absorbance spectra of different solvents (e.g. 0.997 for butan-1-ol/propan-1-ol) were often very high and not suitable for identification purposes (see **Table 3.1**). A considerable improvement was obtained using first-derivative spectra where the highest r value between different solvents was 0.913 when calculated over the wavelength range of 1100 to 2000 nm (see **Table 3.2a**). Optimal results were obtained using second-derivative spectra calculated with a block size of 9 data-points over the wavelength range 1136 – 2000 nm. Second derivative spectra for these optimum settings are shown for all the library in **Fig. 3.7**. The full correlation matrix is shown in **Table 3.3** from which it can be seen that the highest correlation between spectra of different solvents was 0.903. Third and fourth derivative spectra gave almost equivalent results, with slightly lower r values (see **Table 3.2a**). The same optimum conditions were obtained using the spectra from setup 5 which used a sample cup and reflector and gave usable spectra over the full wavelength range (see **Table 3.2b** and **Fig. 3.8**). Information in the 2000 – 2498 nm range far from enhancing the ability to distinguish between solvents actually decreased it. Many solvents have common strongly absorbing peaks in this region and these peaks tend to dominate the calculation of r.

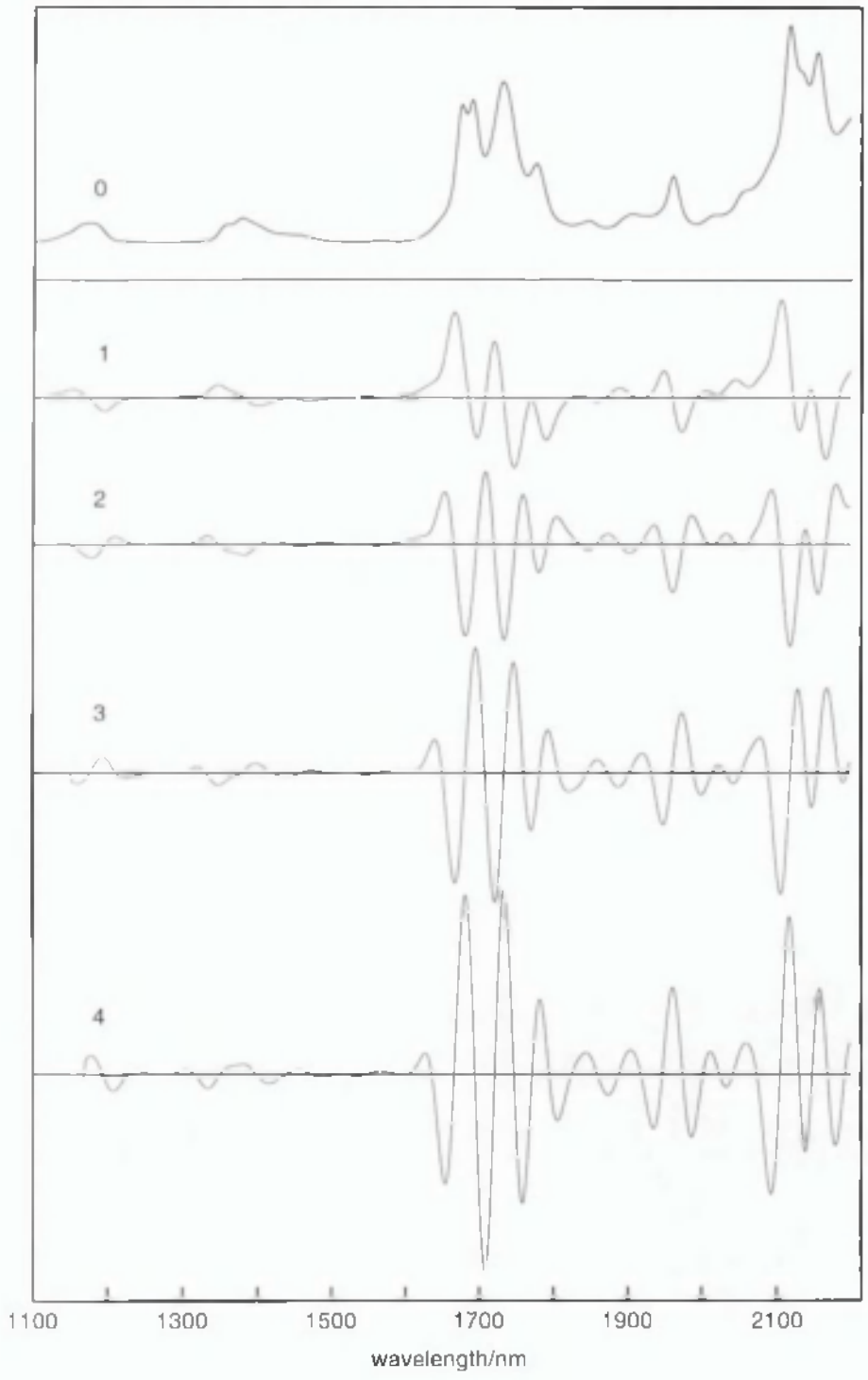


Fig. 3.5A Spectra for acetone calculated using different orders of derivatives. The number on the spectrum indicates the derivative order. All derivatives calculated using nine data-point block size.

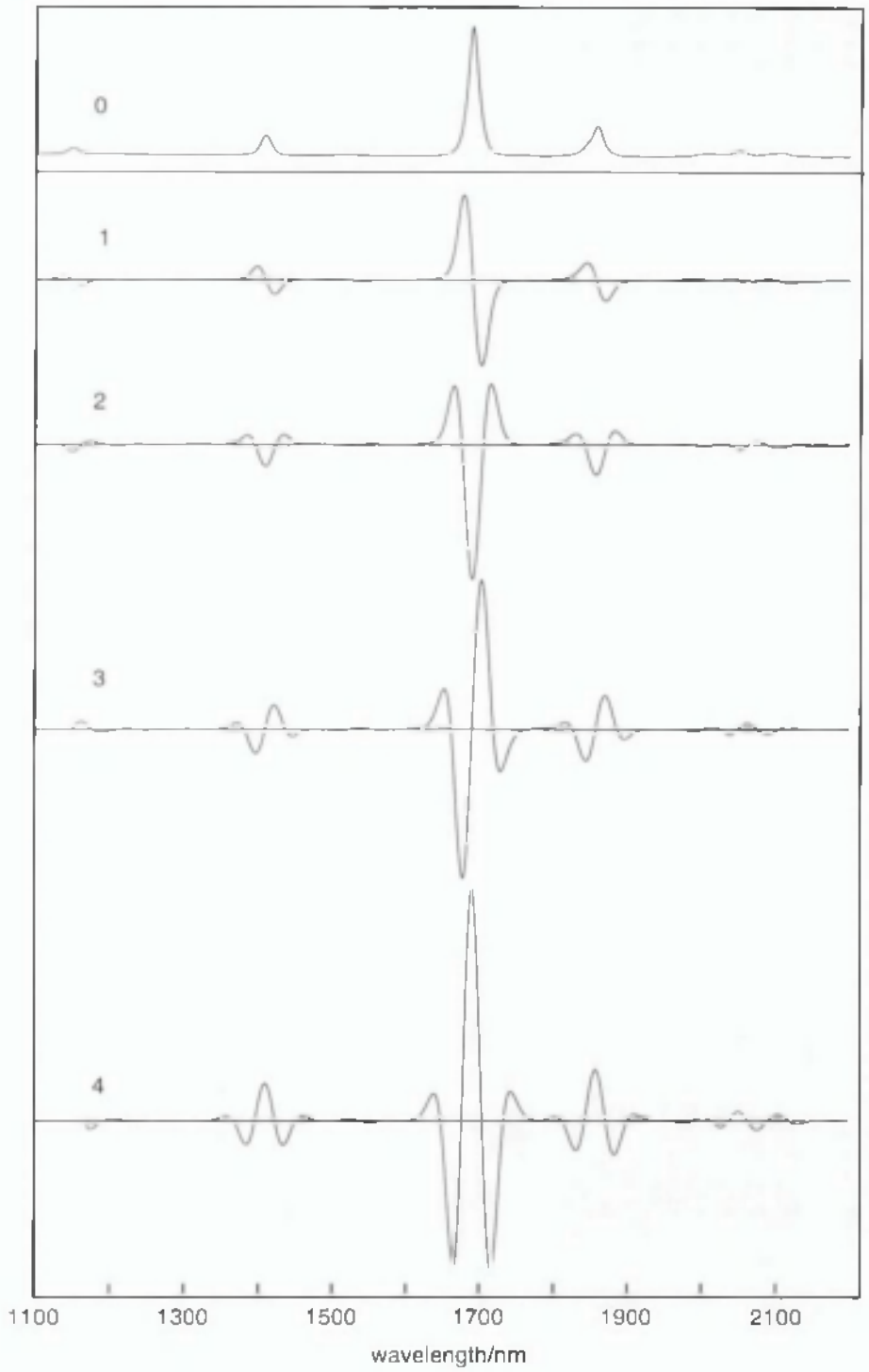


Fig. 3.5B Spectra for chloroform calculated using different orders of derivatives. The number on the spectrum indicates the derivative order. All derivatives calculated using nine data-point block size.

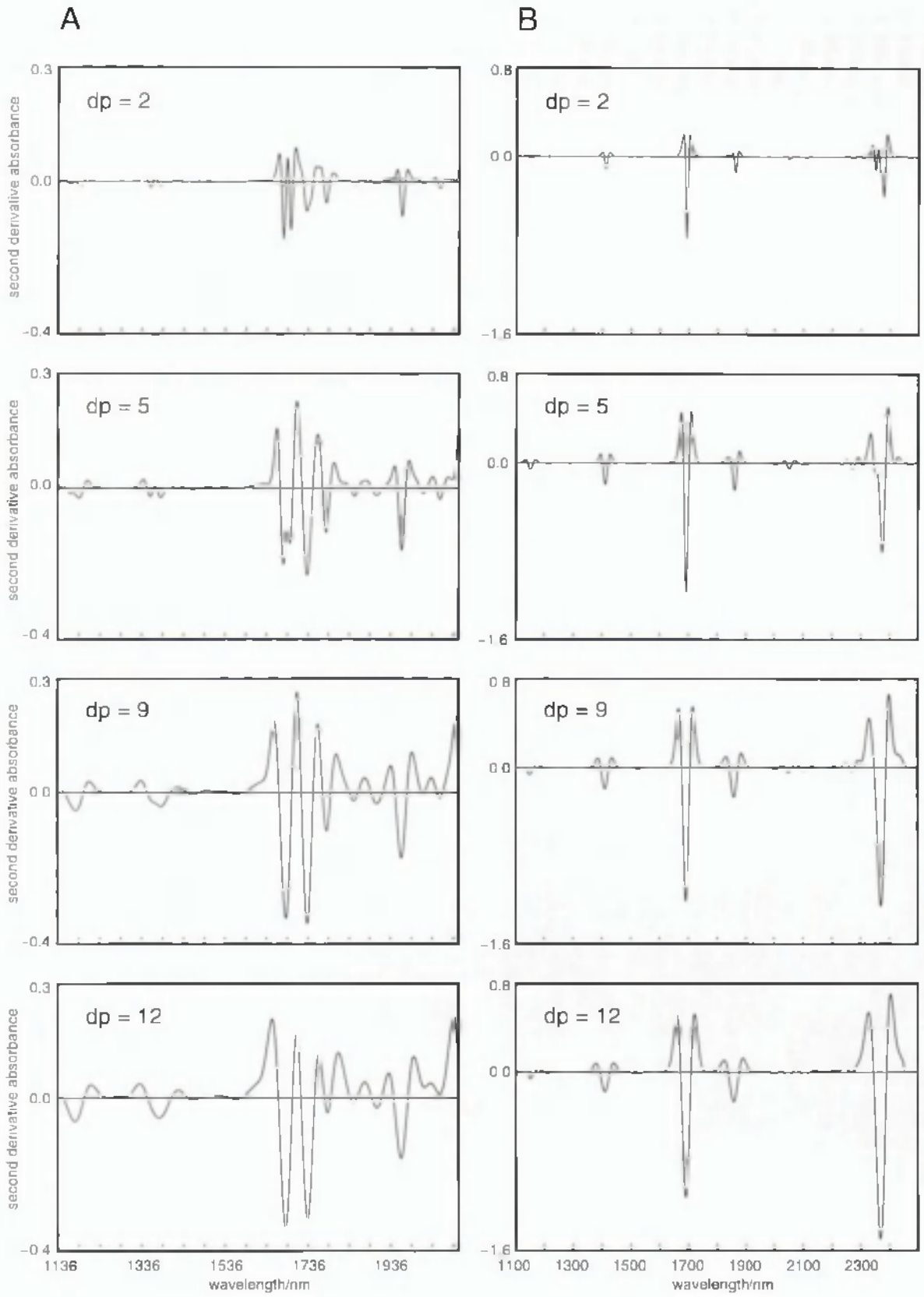


Fig. 3.6 Second derivative spectra calculated using different data-point block sizes (dp) for (A) acetone and (B) chloroform.

Table 3.1 Highest *r* values comparing absorbance spectra of different solvents measured using different wavelength and wavenumber ranges for (a) setup 1 (dispersive instrument) and (b) setup 7 (FT instrument).

(a)

Wavelength/nm	*r* values	Solvent pairs
1100 to 1900	0.9966	Butan-1-ol/propan-1-ol
1100 to 2000	0.9968	
1100 to 2100	0.9969	
1100 to 2200	0.9974	

(b)

Wavenumber/cm^{-1}	*r* values	Solvent pairs
4500 to 9996	0.9934	Butan-1-ol/propan-1-ol
5000 to 9996	0.9896	
5500 to 9996	0.9922	Absolute alcohol/ethanol 96%
6000 to 9996	0.9967	Butan-1-ol/propan-1-ol

Table 3.2 Optimisation of spectral parameters for different instrumental setups, based upon highest r values comparing spectra of different solvents.

(a) Data analysis for spectra measured on setup 1 (grating spectrophotometer with fibre-optic probe) using different wavelength ranges, order of derivatives and data-point block sizes.

	r values			
	1100‡ to 1900 nm	1100‡ to 2000 nm	1100‡ to 2100 nm	1100‡ to 2200 nm
1st derivative				
†2	0.917	0.913*	0.928	0.933
3	0.919	0.918	0.933	0.937
4	0.925	0.924	0.938	0.942
5	-	0.930	-	-
6	0.935	0.935	0.947	0.950
7	-	0.940	-	-
8	-	0.945	-	-
10	0.959	0.954	0.962	0.960
2nd derivative				
5	0.966	0.942	0.933	0.928
6	0.963	0.936	0.928	0.923
7	0.958	0.927	0.919	0.915
8	0.951	0.915	0.910	0.906
9	0.944	0.903*	0.913	0.914
10	0.937	0.911	0.923	0.925
11	0.923	0.926	0.936	0.938
12	0.940	0.938	0.947	0.949
15	0.966	0.965	0.969	0.970
3rd derivative				
2	0.972	0.966	0.966	0.966
5	0.976	0.961	0.958	0.958
8	-	0.937	-	-
9	-	0.919	0.918	0.920
10	0.959	0.909	0.906	0.910
11	-	0.899*	0.911	0.914
12	-	0.918	-	-
15	0.972	-	-	-
4th derivative				
2	0.971	0.969	0.969	0.969
5	0.978	0.968	0.966	0.965
8	0.978	-	-	-
9	-	0.932	0.935	0.936
10	0.970	0.916	0.923	0.925
11	0.961	0.898*	0.911	0.914
12	0.952	0.906	0.916	0.921
15	0.970	0.945	0.957	0.960
5th derivative				
2	0.973	0.971	0.971	0.971
5	0.976	0.969	0.968	0.968
9	-	0.942	0.943	0.943
10	0.978	0.931	0.933	0.934
11	-	0.917	0.921	0.924
12	-	0.903*	0.909	0.915
13		0.932	0.924	0.930
15	-	0.965	-	-

* Lowest r values for different pairs of solvents for each order of derivative

† data-point block size

‡ Note : actual starting wavelength varies with derivative

(b) Analysis for the second-derivative spectra measured on setup 5 (dispersive spectrophotometer with horizontal setup) using different wavelength ranges.

Wavelength/nm	*r* values	Corresponding solvents
1100 to 1900	0.935	Absolute alcohol/ethanol 96%
1100 to 2000	0.901	Absolute alcohol/propan-2-ol
1100 to 2100	0.909	Absolute alcohol/ethanol 96%
1100 to 2200	0.910	
1100 to 2300	0.975	
1100 to 2498	0.973	

(c) Analysis for spectra measured on setup 7 (FT spectrophotometer with fibre-optic probe) using different wavenumber ranges, order of derivatives and data-point block sizes

	r values			
	4500 to 9996 cm^{-1}	5000 to 9996 cm^{-1}	5500 to 9996 cm^{-1}	6000 to 9996 cm^{-1}
1st derivative				
†1	0.927	0.887*	0.990	0.948
2	0.932	0.907	0.989	0.953
3	0.940	0.913	0.988	0.957
5	0.957	0.938	0.985	0.979
2nd derivative				
1	0.947	0.941	0.997	
2	0.931	0.921	0.997	
3	0.923	0.936	0.997	-
5	0.932	0.902*	0.995	
7	-	0.934	-	
3rd derivative				
1	0.966	0.968	0.997	
2	0.951	0.949	0.997	
3	0.945	0.955	0.998	-
5	0.915*	0.918	0.997	
7	0.943	0.920	-	
4th derivative				
1	0.972	0.974	0.996	
2	0.956	0.957	0.997	
3	0.958	0.962	0.998	-
5	0.921	0.921	0.997	
7	0.924	0.906*	-	
9	-	0.956	-	

* Lowest *r* values comparing different solvents for each order of derivative

† data-point block size

Table 3.3 Correlation coefficients for spectral library measured using setup 1. Second derivative spectra (9 data-point block size). Wavelength range 1136 – 2000 nm.

	A	AN	B1OL	B2OL	CH	DCM	DMF	E	E96	EA	M	P1OL	P2OL	THF	TOL
Acetone (A)	1.000	0.001	–0.451	0.075	0.292	–0.057	0.009	0.403	0.330	0.5739	0.171	–0.238	0.372	0.227	0.573
Acetonitrile (AN)		1.000	0.212	–0.165	–0.494	–0.389	–0.142	0.178	–0.168	0.611	0.039	–0.034	–0.278	0.213	–0.068
Butan-1-ol (B1OL)			1.000	0.557	–0.053	0.290	0.500	0.406	0.343	–0.183	0.606	–0.894	0.199	0.565	–0.424
Butan-2-ol (B2OL)				1.000	0.674	0.750	0.807	0.890	0.742	0.089	0.811	0.839	0.894	0.494	0.284
Chloroform (CH)					1.000	0.794	0.583	0.651	0.506	0.037	0.473	0.329	0.774	0.036	0.617
Dichloroform (DCM)						1.000	0.645	0.608	0.501	–0.163	0.497	0.546	0.684	0.180	0.255
Dimethylformamide (DMF)							1.000	0.701	0.667	0.071	0.680	0.752	0.716	0.493	0.343
Ethanol (abs) (E)								1.000	0.853	0.258	0.853	0.699	0.903*	0.580	0.346
Ethanol (96%) (E96)									1.000	0.251	0.651	0.591	0.747	0.457	0.288
Ethyl acetate (EA)										1.000	0.146	–0.143	0.237	0.213	0.496
Methanol (M)											1.000	0.796	0.705	0.586	0.189
Propan-1-ol (P1OL)												1.000	0.560	0.596	–0.109
Propan-2-ol (P2OL)													1.000	0.406	0.514
Tetrahydrofuran (THF)														1.000	–0.204
Toluene (TOL)															1.000

* Highest correlation value

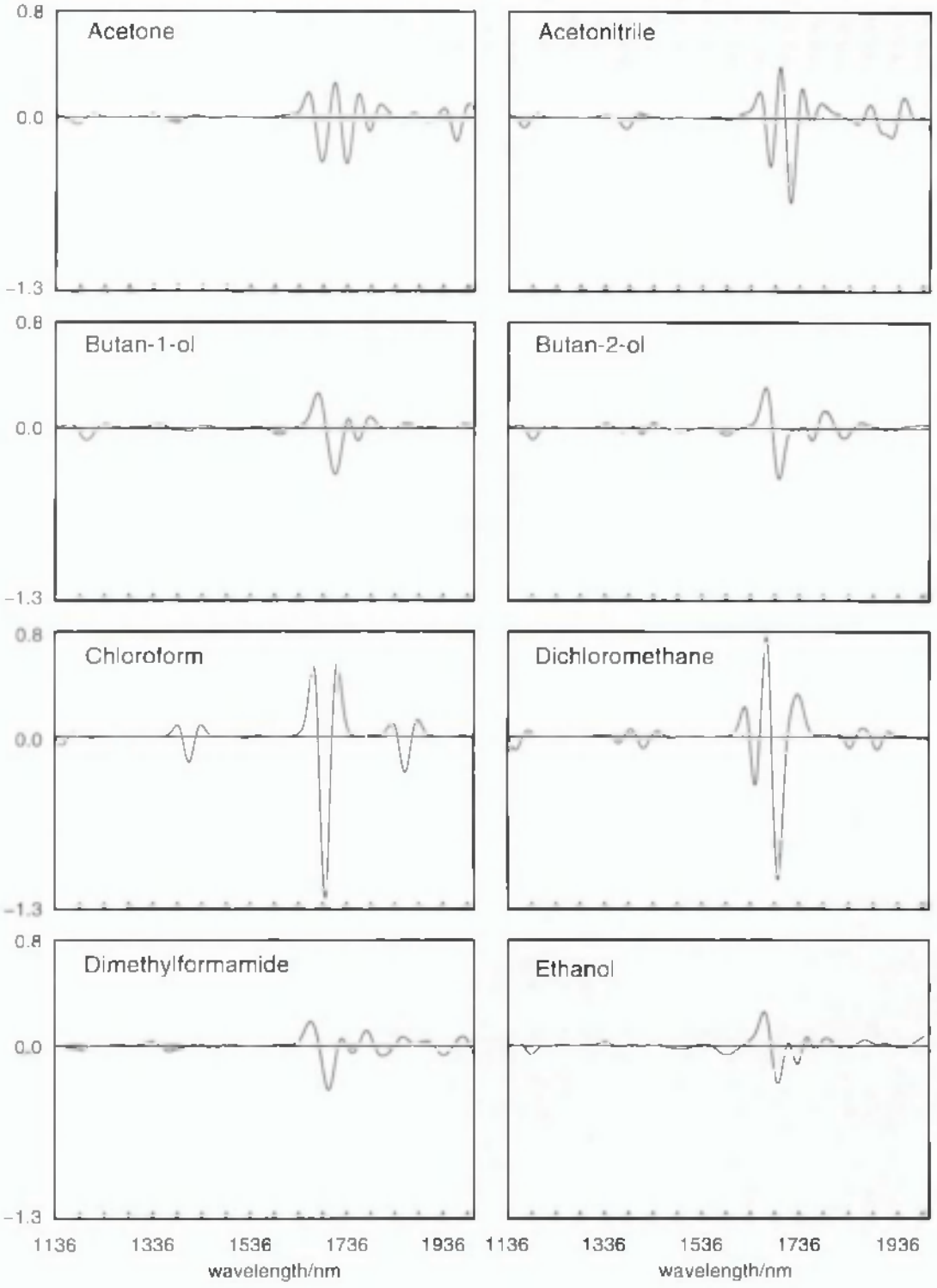


Fig. 3.7 Second derivative (9 data-points block size) spectra of the library. solvents. Setup 1.

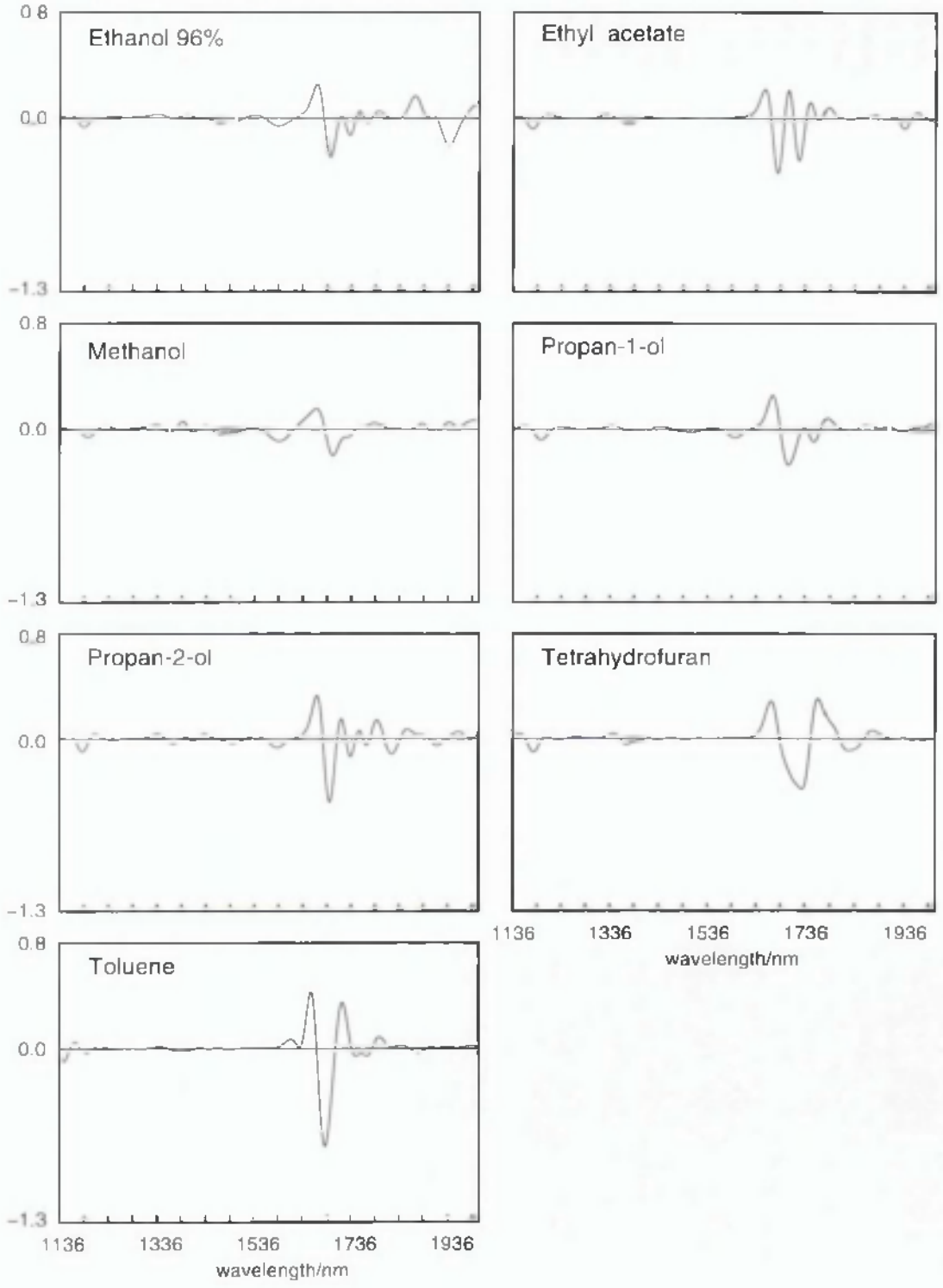


Fig. 3.7 Continued.

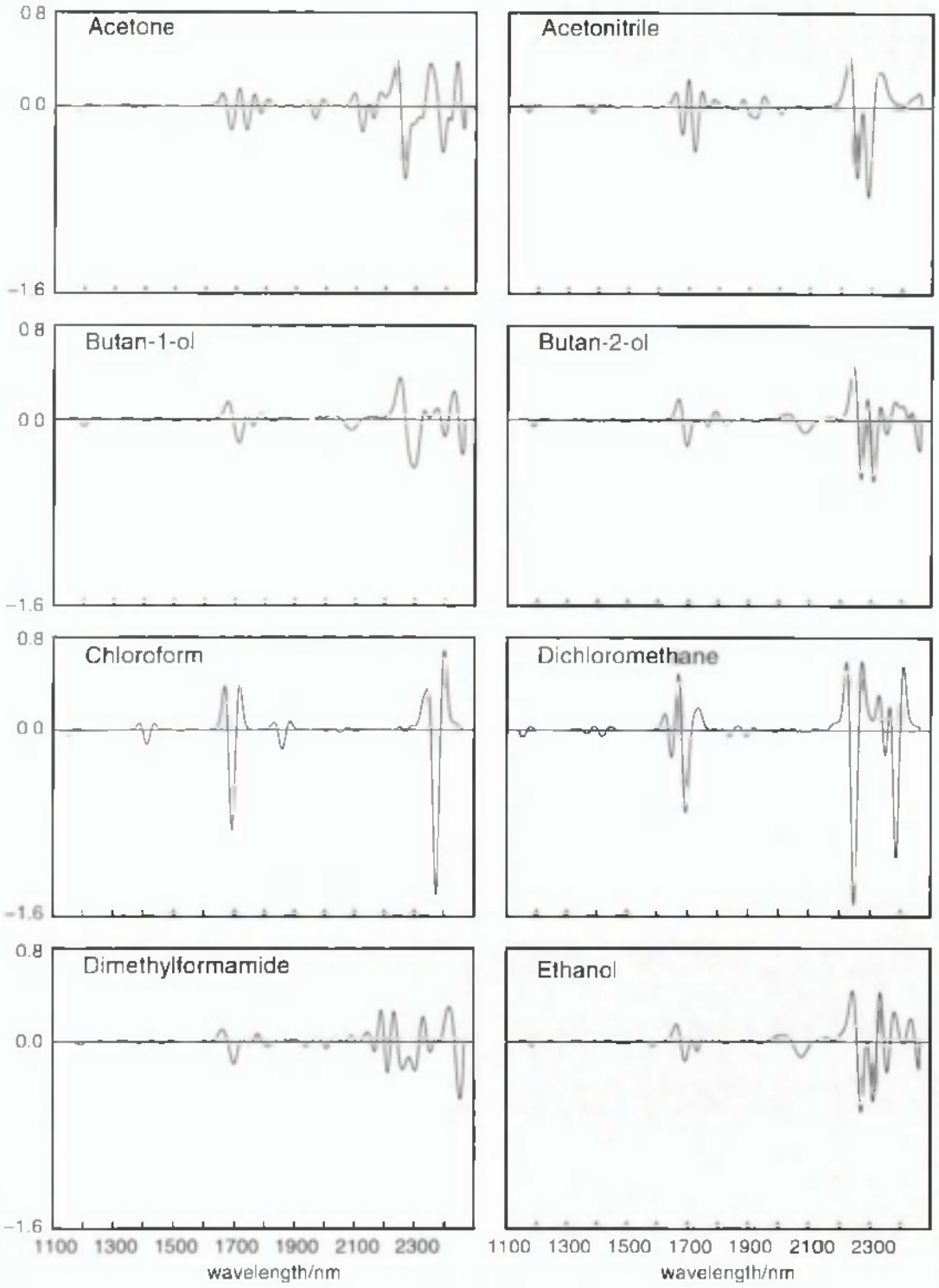


Fig. 3.8 Second derivative (9 data-point block size) spectra of the library solvents. Setup 5.

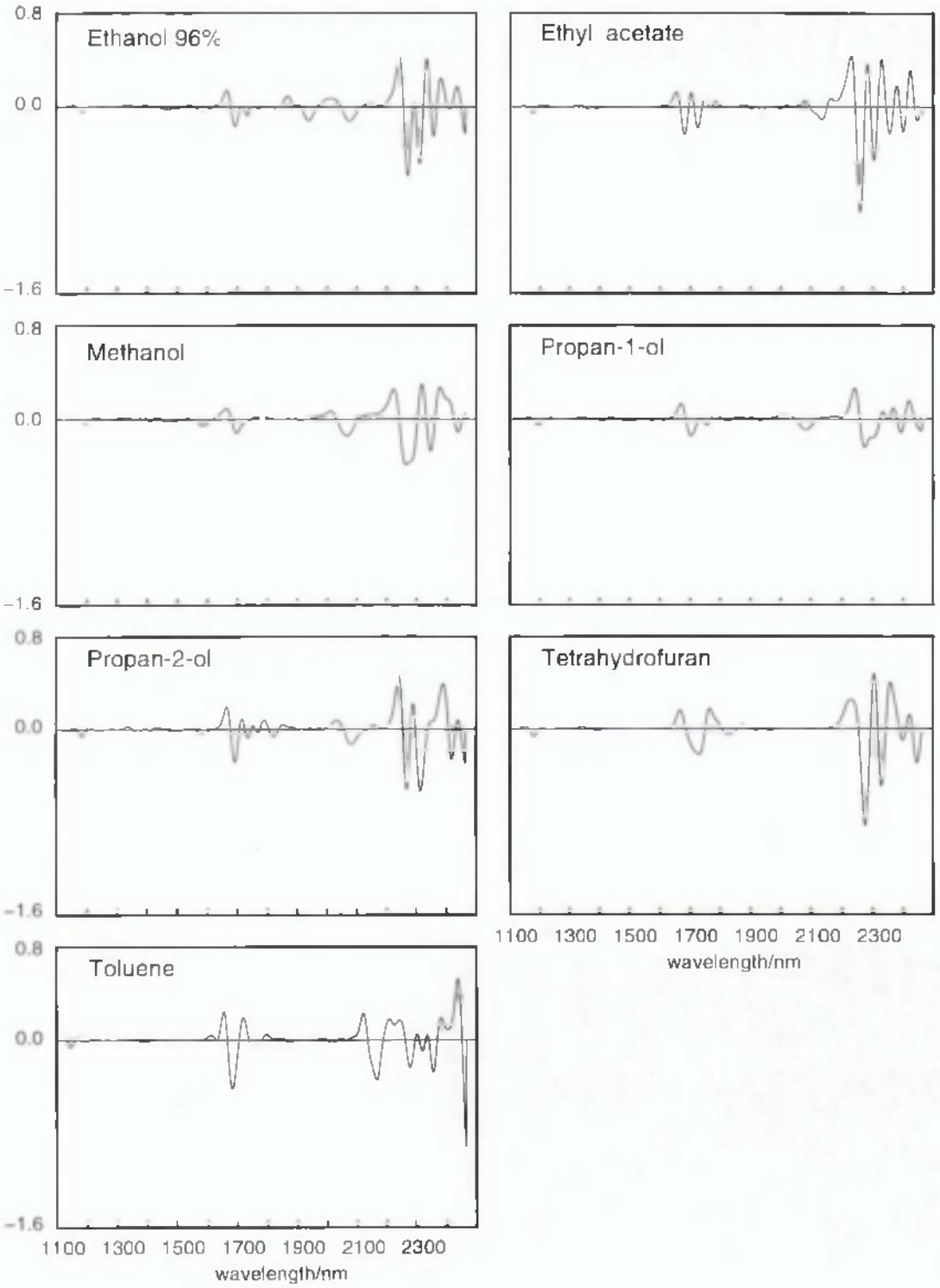


Fig. 3.8 Continued.

It should be noted that using the original wavenumber scaled spectra from setups 7 and 8, that is the two FT NIR instruments, the optimum conditions for distinguishing between the solvents is different, namely: 1st derivative absorbance spectra calculated using a 1 data point block size, and over the range 5000 - 9996 cm^{-1} (2000 – 1000 nm), (see **Table 3.2c**). The highest r value between different solvents was 0.887 (butan-2-ol/propan-2-ol) and is somewhat lower than for the wavelength scaled data.

A comparison of all different possible pairs of the 119 spectra measured on all eight setups was carried out. This gives 7021 different possible pairs of which 413 are between the same solvents but measured on different instruments.

> The number of different pairs is given by $C_{n\,x}$ where $n = 119$ and $x = 2$.
>
> $C_{n\,x} = n!/\{x!\,(n-x)!\}$
>
> $\therefore C_{119,2} = 7021$
>
> 14 of the solvents were each measured on all 8 setups, while 1 more solvent was measured on 7 setups. The number of correct matches is therefore given by
>
> $14 \times (7 + 6 + 5 + 4 + 3 + 2 + 1) + (6 + 5 + 4 + 3 + 2 + 1) = 413$

The correlation coefficients fall into two distinct groups with a clear gap between them (see **Table 3.4a**). There was no overlap between mis-matches and correct matches. The highest r value for a mis-match was 0.931 between ethanol (setup 6) and ethanol 96% (setup 8), while the lowest r for a correct match was 0.978 between dichloromethane samples on setups 2 and 7. Restricting the spectra to only those measured on the FOSS setups gave the highest r value between different solvents as 0.917 between ethanol (setup 6) and ethanol 96% (setup 3), and the lowest r value between the same solvent as 0.991 (ethanol, setups 6 and 4). This improvement is hardly surprising since the spectra from the FT instruments had had to undergo wavenumber to wavelength conversion. There is no indication that spectra differed in quality between any of the setups. Clearly a value of > 0.97 strongly suggests a positive identification though any value below 0.99 may indicate a difference in purity and should be treated with caution.

Table 3.4 Distribution of correlation coefficients calculated between all different pairs of spectra (library solvents) measured on the eight setups.

(a) Second derivative absorbance spectra (9 data-point block size), 1136 – 2000 nm

Correlation coefficient range	Number of pairs Setups 1 – 8	Number of pairs Setups 1 – 6	Notes
< 0.900	6535	3653	All pairs between different solvents
0.900 < 0.910	41	37	
0.910 < 0.920	25	6	
0.920 < 0.930	6	0	
0.930 < 0.940	1	0	
0.940 < 0.950	0	0	
0.950 < 0.960	0	0	
0.960 < 0.970	0	0	
0.970 < 0.980	1	0	All pairs between same solvents
0.980 < 0.990	7	0	
0.990 < 1.000	405	220	

(b) Third derivative absorbance spectra (9 data-point block size), 1154 – 2000 nm

Correlation coefficient range	Number of pairs Setups 1 – 8	Notes
< 0.900	6558	All pairs between different solvents
0.900 < 0.910	30	
0.910 < 0.920	11	
0.920 < 0.930	8	
0.930 < 0.940	1	
0.940 < 0.950	0	
0.950 < 0.960	0	
0.960 < 0.970	0	
0.970 < 0.980	1	All pairs between same solvents
0.980 < 0.990	7	
0.990 < 1.000	405	

(c) Fourth derivative absorbance spectra (9 data-point block size), 1172 – 2000 nm.

Correlation coefficient range	Number of pairs Setups 1 – 8	Notes
< 0.900	6564	All pairs between different solvents
0.900 < 0.910	30	
0.910 < 0.920	7	
0.920 < 0.930	0	
0.930 < 0.940	4	
0.940 < 0.950	3	
0.950 < 0.960	0	
0.960 < 0.970	0	
0.970 < 0.980	1	All pairs between same solvents
0.980 < 0.990	7	
0.990 < 1.000	405	

Again, the third and fourth derivative spectra with gave similar distribution patterns although the *r* values tended to concentrate more towards higher values than when using the second-derivative spectra (see **Table 3.4b** and **c** respectively).

3.3.1 Effect of water vapour on air reference spectrum

The effect of water vapour on the air reference spectrum was investigated by recording the spectrum of air dried with silica gel and air saturated with water. *r* values comparing spectra of propan-2-ol, dimethylformamide and acetonitrile measured under these two conditions were 0.999772, 0.999975 and 0.999947 respectively. With such small effects on the *r* values (< 0.0005), humidity changes are therefore unlikely to be important.

3.3.2 Effect of wavelength accuracy

Errors in wavelength calibration between different instruments will cause the *r* value to decrease for a given solvent. The effect of wavelength errors was simulated for three solvents by shifting the whole spectrum by small amounts using cubic spline interpolation. Again, solvent examples were chosen so as to cover both broad and sharp spectra. The results are shown in **Table 3.5a** from which it can be seen that wavelength shifts up to 1 nm cause little problem, as the *r* values remained > 0.99. This is within the tolerance specified for the FOSS monochromators of ± 0.3 nm and also the FT instruments (± 0.2 nm at 1000 nm to ± 1.2 nm at 2500 nm). Examining the six most intense negative peaks in the second-derivative spectra over all six FOSS setups and fifteen solvents indicated a mean maximum difference of 0.65 nm. Over all eight setups the mean maximum difference was 1.1 nm. The largest difference over all setups was 4.73 nm for the small poorly defined water (impurity) peak at ~ 1940 nm in dimethylformamide (see **Table 3.5b**). Clearly the effects of wavelength errors are small.

3.3.3 Effect of band-pass

Differences in band-passes between monochromators can also be expected to affect *r* values. This was simulated by convoluting the absorbance spectra with triangular filters of varying sizes (8, 9, 10, 11, 12, 14 and 16 nm, width at half the height), see **Fig. 3.9A**.

Table 3.5 Effects of wavelength errors on second-derivative absorbance spectra (9 data-point block size), 1136 – 2000 nm.

(a) Effects on *r* values with whole spectrum shifted to lower wavelengths by a specified amount. All *r* values with respect to unshifted spectrum.

Solvent	Wavelength shift/nm	Correlation coefficient
Butan-1-ol	0.5	0.999
	1.0	0.997
	2.0	0.987
	3.0	0.970
Dichloromethane	0.5	0.998
	1.0	0.994
	2.0	0.976
	3.0	0.946
Toluene	0.5	0.999
	1.0	0.996
	2.0	0.984
	3.0	0.964

(b) Six most intense peaks of dimethylformamide measured on all instrumental setups. Maximum wavelength shift between FOSS setups only and all setups are also indicated.

	Peak positions/nm					
	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6
Setup 1	1698.57	1810.21	1937.80	1748.21	1369.02	1192.20
Setup 2	1698.48	1810.10	1938.76	1748.34	1369.43	1192.89
Setup 3	1699.04	1810.20	1942.05	1748.35	1369.45	1192.28
Setup 4	1698.97	1810.25	1942.53	1748.36	1369.32	1192.67
Setup 5	1698.63	1810.56	1938.39	1748.60	1368.98	1192.41
Setup 6	1698.87	1810.01	1940.50	748.23	1369.17	1192.48
Setup 7	1698.45	1808.99	1941.66	1747.98	1368.55	193.58
Setup 8	1698.88	1809.13	1940.77	1748.21	1368.93	1193.44
Maximum difference for all setups	0.59	1.57	4.73	0.62	0.90	1.38
Maximum difference for Foss setups only	0.56	0.55	4.73	0.39	0.47	0.78

Note: Peak positions determined using program PEAK which used quadratic interpolation to calculated the peak wavelengths.

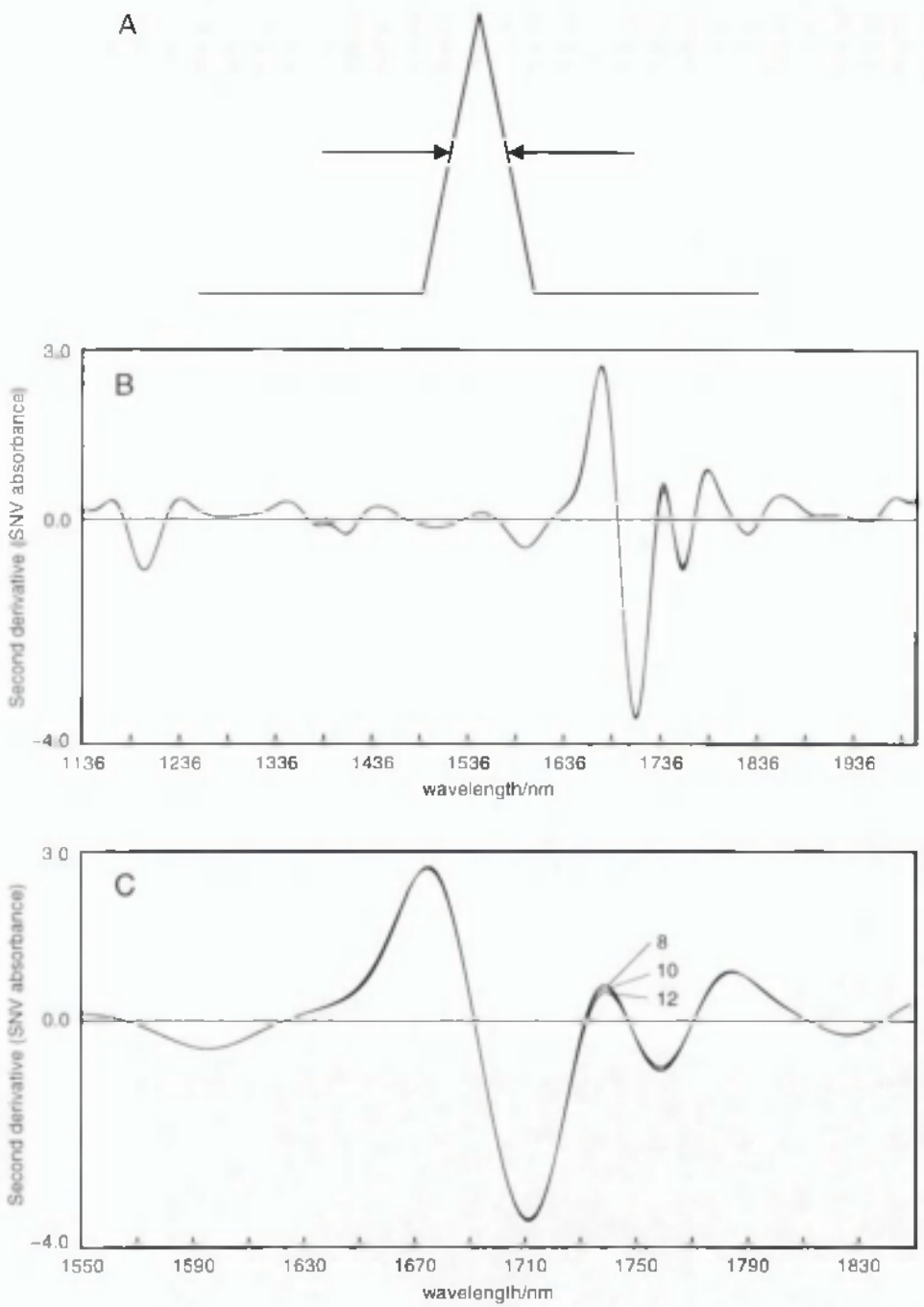


Fig. 3.9 Effects of band-pass errors
(A) Schematic diagram for the triangular band-pass filter used to simulate band-pass effects.
(B) Second derivative spectra of butan-1-ol after applying filters of 8, 10 and 12 nm.
(C) As B, but for the wavelength region 1550 to 1850 nm.

This was carried out using an in-house program called DECON. Visually, the effects of the above treatment on the spectra of butan-1-ol was minimal, **Fig. 3.9**. Program DECON did not normalise spectra directly, hence spectra in **Fig. 3.9** have been normalised using standard normal variate to bring all to a common basis for comparison. It was found that small changes in band-pass of ± 2 nm would not cause a problem with r remaining > 0.99 (see **Table 3.6**). This is within the tolerance set by FOSS for their monochromators of −1 to + 2 nm.

3.3.4 Effect of path-length

If the Beer-Lambert Law holds, identification using Correlation in Wavelength Space should not be effected by path-length changes. This was investigated by measuring the spectra of acetone, dichloromethane and methanol over the optical path-length range 1 to 6 mm using the Smartprobe (setup 3). **Fig. 3.10** shows spectra and absorbance *vs* path-length plots for acetone and dichloromethane.

Although the air reference spectrum was measured at the corresponding path-lengths, considerable baseline shifts to negative absorbances were observed and this presumably arose because of refractive index differences between the solvent and air. The Beer-Lambert Law assumes that the radiation impinging on the sample is a parallel beam and perpendicular to the sample thickness. This would require the situation indicated in **Fig. 3.11A**. In practice, the radiation from a fibre-optic probe is divergent, and the reflectors used commonly have a matt finish. With a divergent beam, the amount of radiation collected will depend upon the path-length, **Fig. 3.11B**. As the divergent beam passes through the probe window it will undergo refraction because of the difference between the refractive indices of the window and sample, **Fig. 3.11C**. There will be greater refraction for the air reference than for the solvent sample and consequently less radiation will be collected for the air reference, (I_o) than the sample, (I). Hence, even if the solvent has a very low or zero absorbance I will be greater than I_o, giving a negative absorbance value. The larger the path-length, the greater the effect.

Table 3.6 Effects of band-pass errors on correlation coefficient.. Spectra measured on setup 1. Second derivative spectra (9 data-point block size, 1136 – 2000 nm)

(a) *r* values with respect to original spectrum showing the effects of increasing the band-pass.

	Band-pass/nm						
	8	9	10	11	12	14	16
Toluene	0.9996	0.9993	0.9991	0.9985	0.9981	0.9964	0.9940
Chloroform	0.9992	0.9986	0.9981	0.9970	0.9961	0.9927	0.9877
Dichloromethane	0.9993	0.9987	0.9981	0.9970	0.9961	0.9927	0.9875
Butan-1-ol	0.9996	0.9993	0.9990	0.9985	0.9980	0.9963	0.9938

(b) *r* values with respect to spectrum convoluted with 11 nm band-pass filter showing the effects of decreasing and increasing the band-pass.

	Band-pass/nm						
	8	9	10	11	12	14	16
Toluene	0.9996	0.9998	0.9999	1.0000	>0.9999	0.9996	0.9984
Chloroform	0.9993	0.9997	0.9999	1.0000	>0.9999	0.9990	0.9970
Dichloromethane	0.9992	0.9997	0.9999	1.0000	>0.9999	0.9990	0.9967
Butan-1-ol	0.9996	0.9998	>0.9999	1.0000	>0.9999	0.9995	0.9984

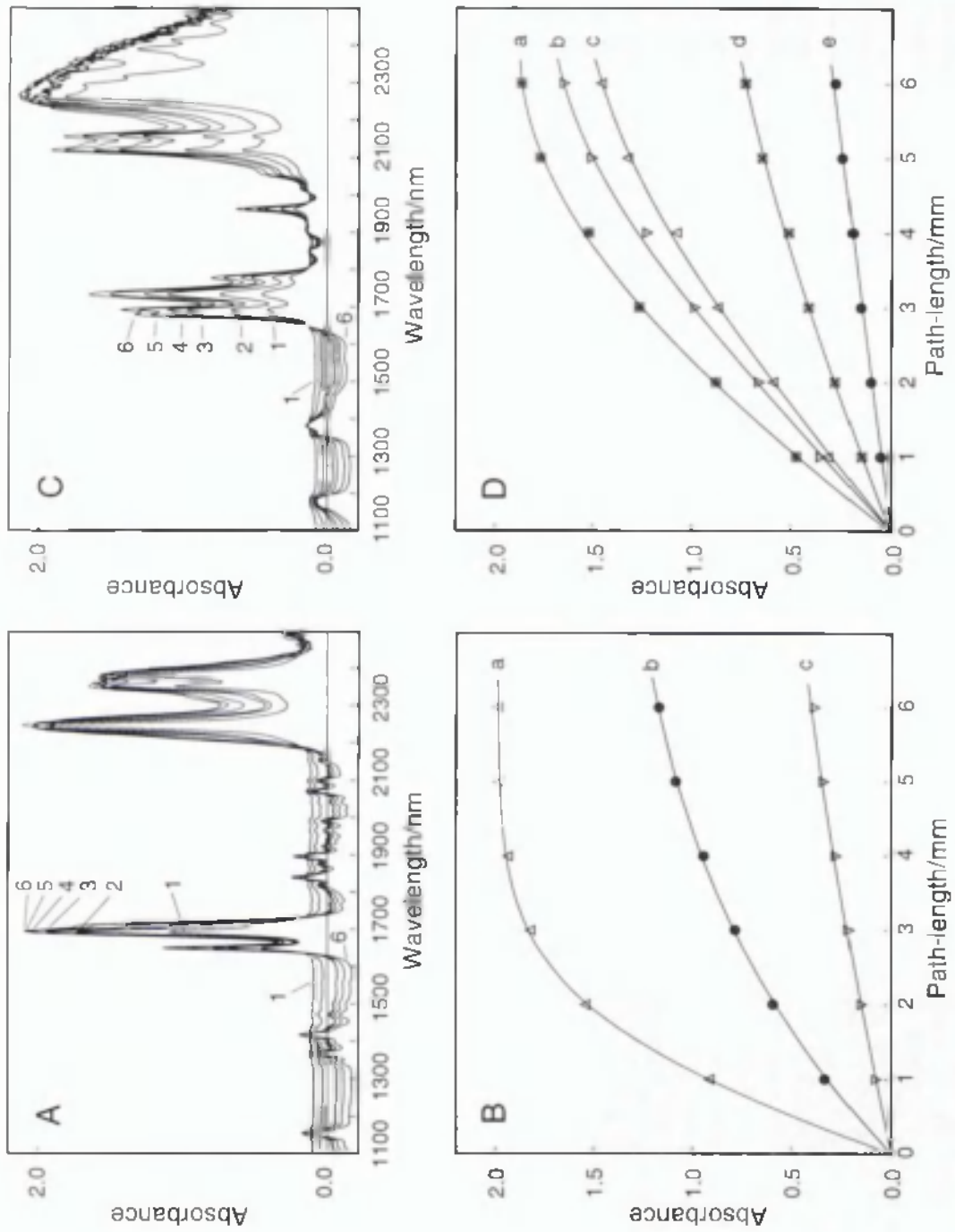


Fig. 3.10 Effects of path-length on absorbance spectra.
(A) Absorbance spectra for dichloromethane. (B) Absorbance *vs* path-length for dichloromethane: (a) 1692, (b) 1648 and (c) 1840 nm. Absorbances measured with respect to baseline at 1578 nm.
(C) Absorbance spectra for acetone. (D) Absorbance *vs* path-length for acetone (a) 2118, (b) 1732, (c) 1692, (d) 1962 and (e) 1378 nm. Absorbances measured with respect to baseline at 1268 nm.
Optical path-lengths: (1) 1, (2) 2, (3) 3, (4) 4, (5) 5 and (6) 6 mm.
All spectra measured using setup 3.

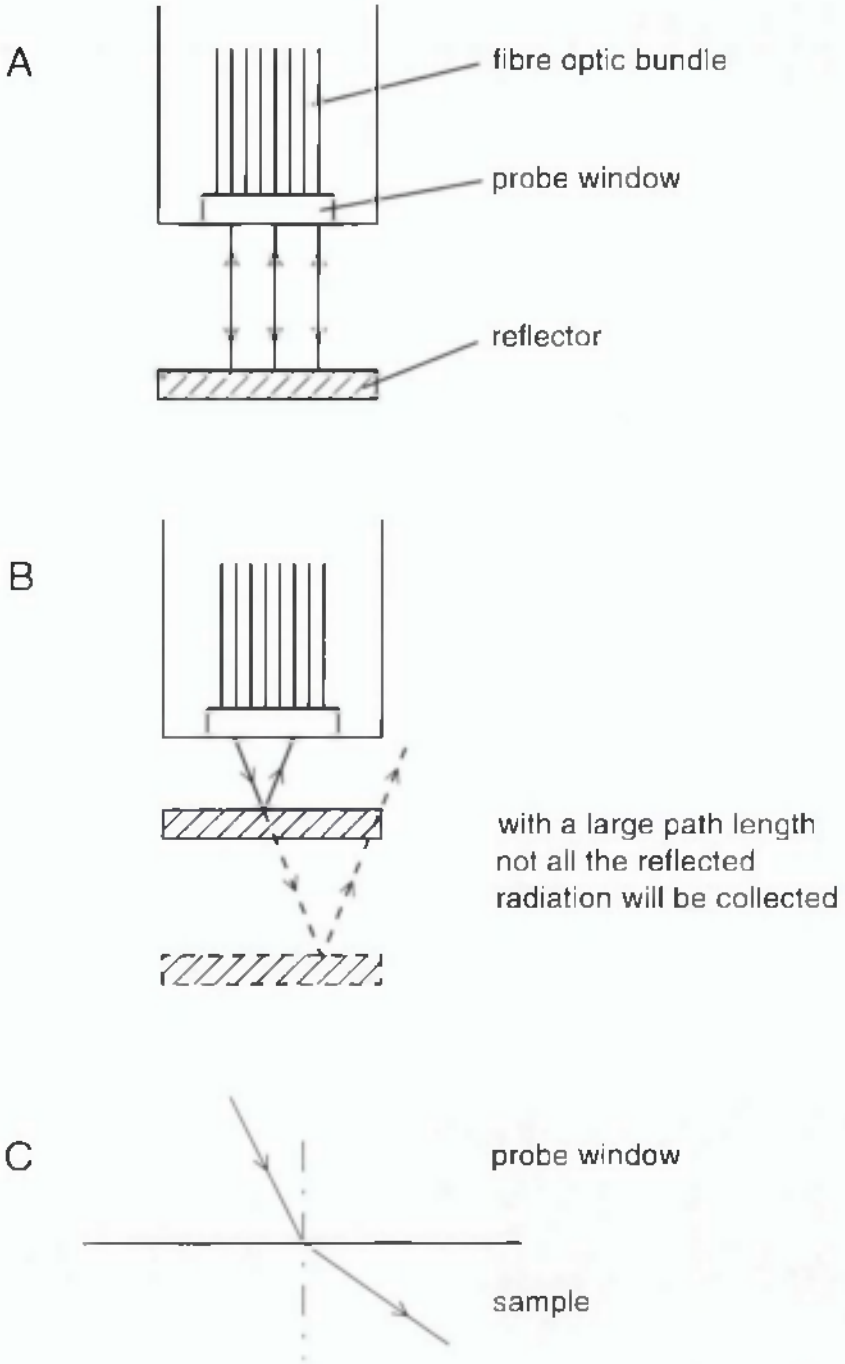


Fig. 3.11 Schematic diagram illustrating the radiation path when using a fibre optic probe.
(A) Ideal case for Beer-Lambert Law to hold.
(B) Real probe with divergent beam. Amount of radiation collected depends upon path-length.
(C) Radiation will be refracted as it passes from the probe window into into the sample/reference.

Only for small peaks were reasonably straight line plots of absorbance *vs* path-length obtained. Second-derivative absorbance *vs* path-length plots showed similar curvature (see **Fig. 3.12**). With too large a path-length it is easy to exceed the dynamic range of the instruments which varied considerably over the eight setups. The dynamic range was determined by measuring ethanol/water mixtures of concentrations 0, 20, 25, 45, 50, 60, 70, 80, 90, 96 & 100% v/v water in absolute alcohol and visually noting the maximum absorbance of the strongly absorbing water peak at ~ 1940 nm (see **Fig. 3.13**). For setups 1, 2, 4 and 5 the maximum measurable absorbance was approximately 1.9; 2.4 for setup 3 and 2.9 for setup 6. These values are lower than the photometric range of 3 AU that has been specified by FOSS NIRSystems (see chapter 2). Both FT instruments were software limited to absorbance values of $\leq$ 2.0. Even where the dynamic range of the instrument was not exceeded, non-linearity is clearly seen. Small changes in path-length can be tolerated provided the solvent absorbance is small, **Table 3.7a**. Similar values of *r* for acetone to those shown in **Table 3.7a** were obtained for measurements made on the Buhler instrument, **Table 3.7b**.

Small differences between the path-lengths used to measure the sample and reference were found not to be important. This is shown in **Table 3.7a** for acetone measured with respect to a fixed 2 mm air reference.

3.3.5. Effect of temperature

NIR spectra are known to be sensitive to temperature changes (Wulfert *et al.*, 1998; DeBraekeleer *et al.*, 1998). The effect of temperature was investigated using setup 1 by measuring spectra of all the solvents except 96% ethanol at approximately 0, 20 and 33 °C. These spectra were then compared against the spectra measured at ambient temperature (second-derivative absorbance, 9 data point block size). No significant effects on the correlation coefficients were observed and in no case did the *r* value fall below 0.9926 (see**Table 3.8**). Visually the most affected solvents were those in which hydrogen bonding occurred and absorbance changes in the OH regions could be easily seen. **Fig. 3.14** shows the effect of temperature on the spectrum of methanol, as an

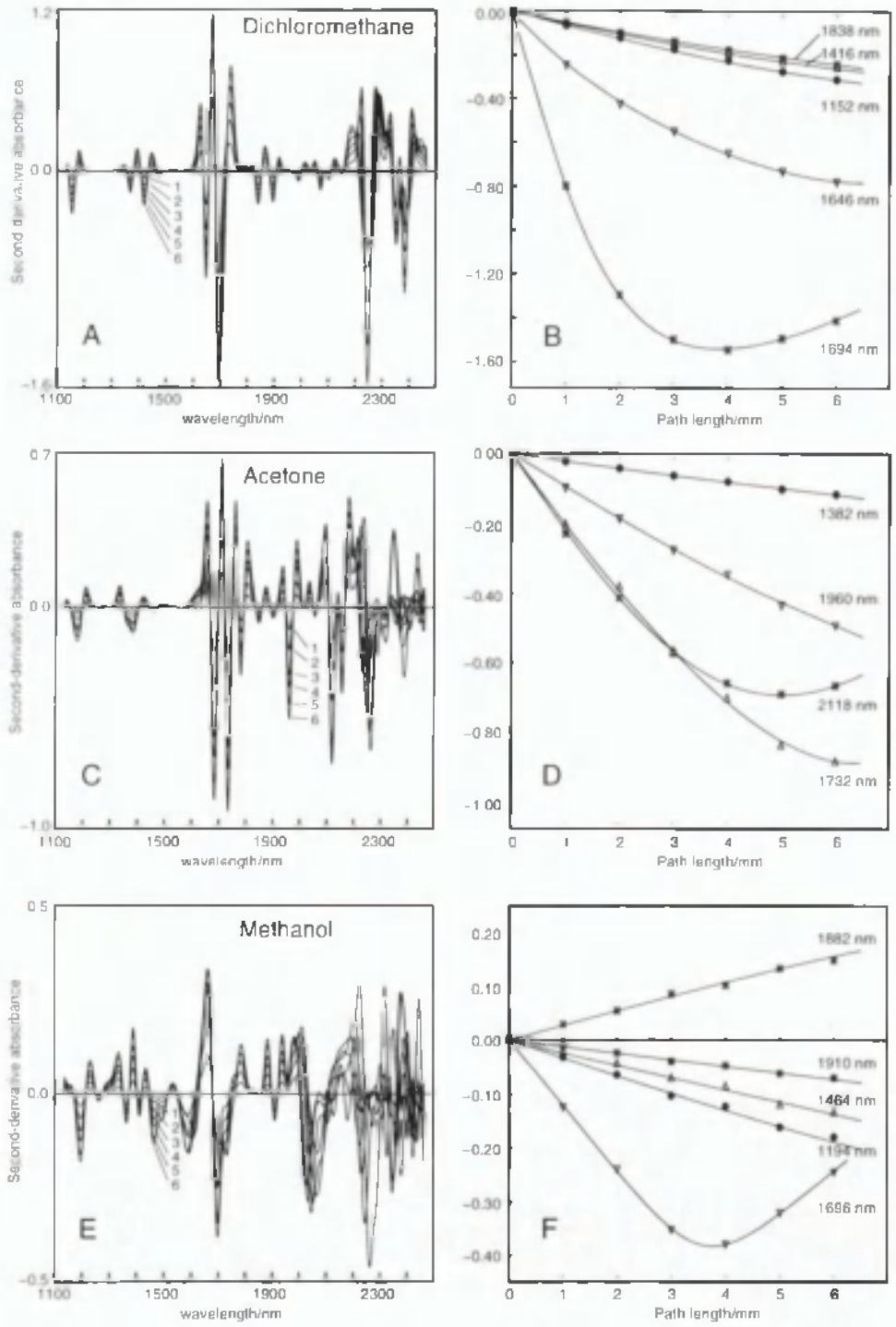


Fig. 3.12 Effect of path-length on second derivative spectra.
(A) Spectra for dichloromethane.
(B) Peak amplitude vs path-length for dichloromethane.
(C) Spectra for acetone.
(D) Peak amplitude vs path-length for acetone.
(E) Spectra for methanol.
(F) Peak amplitude vs path-length for methanol.
All spectra measured using setup 3 at path-lengths of 1, 2, 3, 4, 5 and 6 mm.

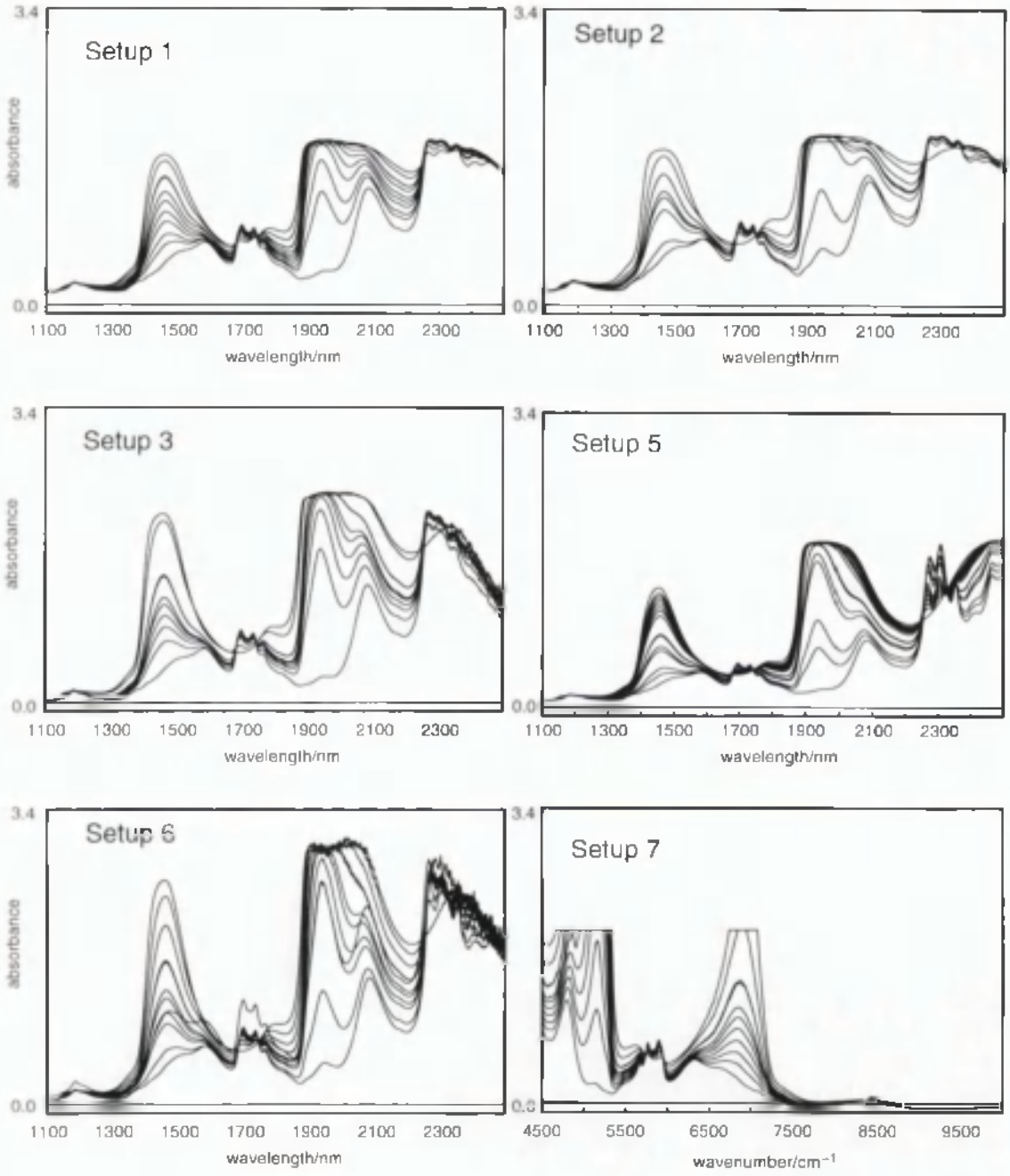


Fig. 3.13 Absorbance spectra for ethanol/water mixtures measured using various setups.

Table 3.7 Effect of path-length on the correlation coefficient.

(a) Second-derivative absorbance spectra (9 data-point block size), 1136 – 2000 nm.

All *r* values with reference to the 2 mm optical path-length. Measured on setup 3.

	Correlation coefficient					
Optical path-length/mm	1	2	3	4	5	6
Acetone	0.9998	1.0000	0.9998	0.9994	0.9985	0.9970
Acetone (air ref. Fixed at 2 mm)	0.9998	1.0000	0.9998	0.9995	0.9982	0.9969
Dichloromethane	0.9969	1.0000	0.9954	0.9799	0.9530	0.9283
Methanol	0.9999	1.0000	0.9991	0.9950	0.9464	0.8796

(b) Second-derivative absorbance spectra (1 data-point block size), 4992 – 9984 cm^{-1}. All *r* values with reference to the 2 mm optical path-length. Measured on setup 8.

	Correlation coefficient					
Optical path-length/mm	1	2	3	4	5	6
Butan-1-ol	0.9998	1.0000	0.9994	0.9992	0.9981	0.9954

Table 3.8 Effects of temperature on the correlation coefficients of different solvents. All *r* values calculated with reference to the spectrum measured at ambient temperature.

	Correlation coefficient	
	0 °C	33 °C
Acetone	0.9926	0.9932
Acetonitrile	0.9998	> 0.9999
Butan-1-ol	0.9997	0.9996
Butan-2-ol	0.9997	0.9987
Chloroform	> 0.9999	> 0.9999
Dichloromethane	> 0.9999	> 0.9999
Dimethylformamide	0.9944	0.9998
Absolute alcohol	0.9992	0.9987
Ethyl acetate	> 0.9999	> 0.9999
Methanol	0.9965	0.9970
Propan-1-ol	0.9987	0.9990
Propan-2-ol	0.9990	0.9993
Toluene	> 0.9999	0.9999
Tetrahydrofuran	0.9996	0.9997

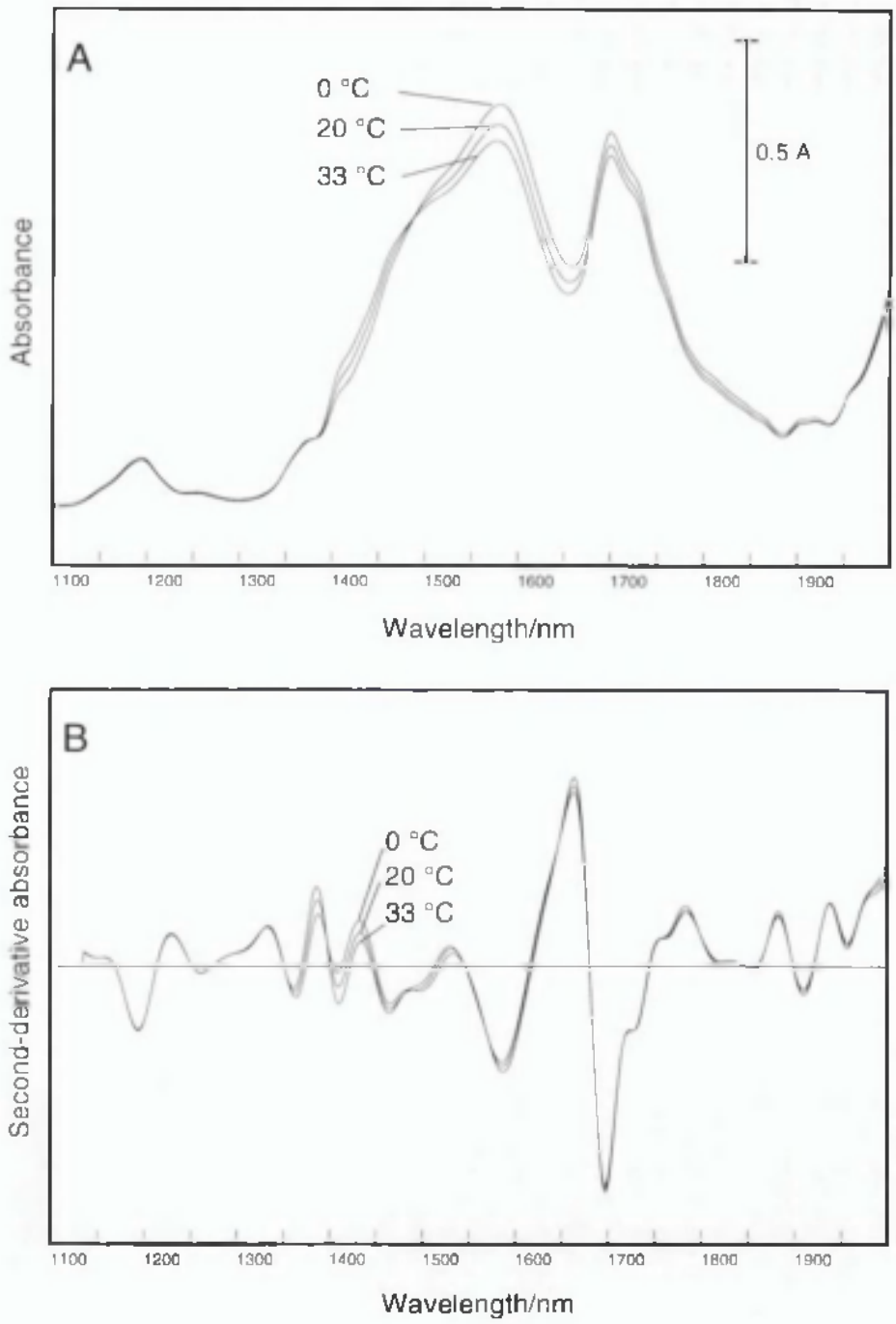


Fig. 3.14 Effect of temperature (0, 20 and 33 °C) on NIR spectra.
(A) Methanol, absorbance spectra, (B) methanol, second-derivative spectra (9 data-point block size), (C) dichloromethane, absorbance spectra, and (D) dichloromethane, second-derivative spectra (9 data-point block size).

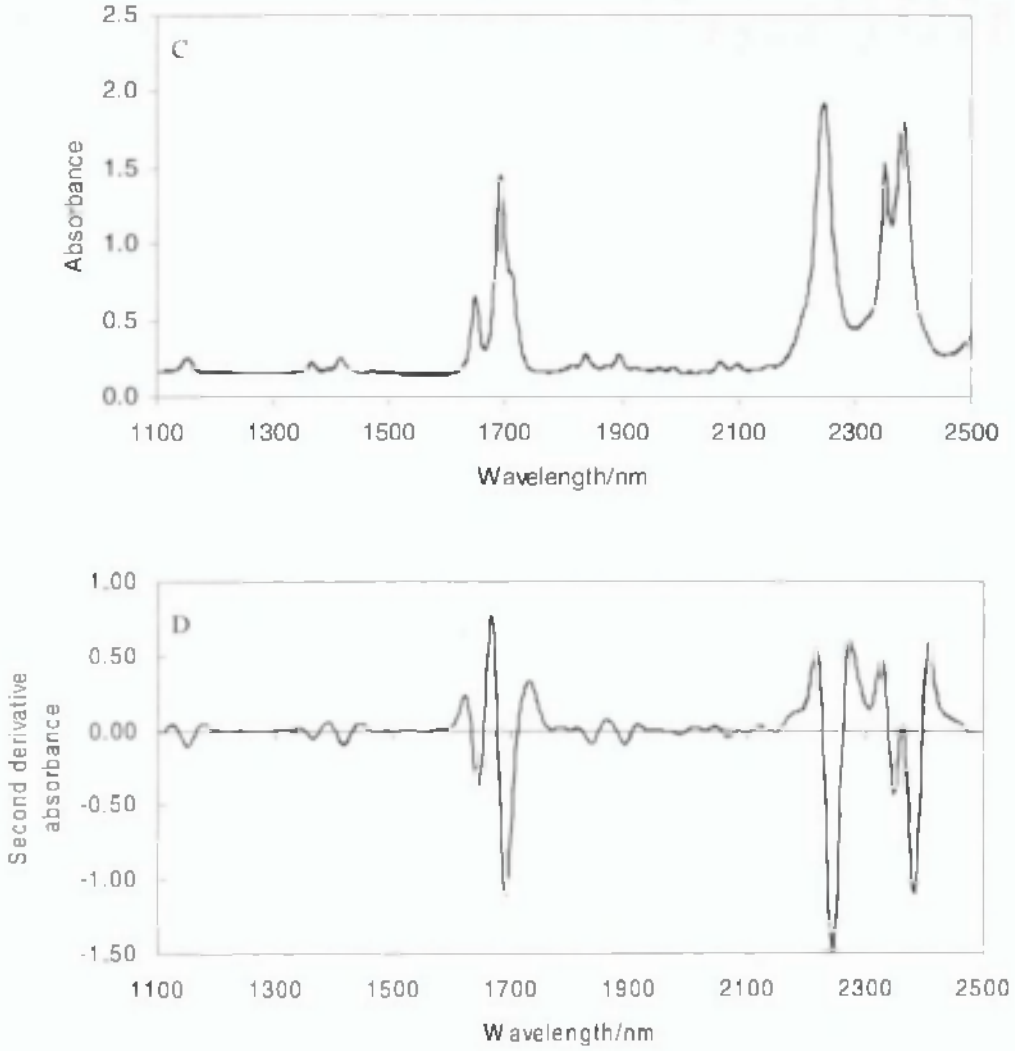


Fig. 3.14 Continued.

example of a hydrogen-bonded solvent and dichloromethane, as a non hydrogen-bonded solvent.

3.3.6 Effect of water

Water is a common impurity in most solvents and its effect on the ability to correctly identify the solvents was investigated. Samples of solvents were spiked with small quantities of water and the spectra compared with those for the corresponding original sample spectra. The quantity of water added corresponded approximately to × 0.5 and × 1 the maximum water content specified by the manufacture for the solvent. The effects on *r* were small and caused no problems in correctly identifying the solvents, **Table 3.9**.

Very small quantities of water could easily be detected by visual inspection of the spectra. Dimethylformamide is very hygroscopic and the uptake of water on exposure to air over a period of only 10 minutes is easily observed, **Fig. 3.15**. The uptake of water by acetone and methanol, while much less, could also be easily seen over a similar time period. The importance of visually comparing spectra and not simply relying on the *r* value can not be too strongly stressed.

3.3.7 Solvent mixtures

Although the library was developed primarily for the identification of pure solvent samples, the ability to reject a number of binary mixtures was investigated. The plots for correlation coefficients for acetone, methanol, propan-2-ol and water with ethanol are shown in **Fig. 3.16.** Correlation values did not fall below 0.99 for mixtures containing up to approximately 10% v/v acetone, 20% v/v methanol and 20% v/v propan-2-ol. Mixtures of similar solvents will always be difficult to correctly identify using Correlation in Wavelength Space and give rise to a high false positive rate. This needs always to be kept in mind when carrying out identifications by the proposed method.

3.3.8 Solvents not in the library

There is always the possibility that a solvent external to the reference library will give a good match to a library solvent and be wrongly identified. **Table 3.10** shows the

Table 3.9 Effect of water on the correlation coefficient. Solvent before adding water taken as reference. Second-derivative absorbance spectra (9 data-point block size), 1136 – 2000 nm.

Solvent	Water (% v/v)	Correlation coefficient
Acetone	0.25	0.9904
	0.50	0.9602
Acetonitrile	0.15	0.9978
	0.30	0.9913
Butan-1-ol	0.25	0.9997
	0.50	0.9990
Butan-2-ol	0.25	0.9995
	0.50	0.9982
Chloroform	0.05	0.9998
	0.1	0.9996
Dichloromethane	0.025	0.9994
	0.050	0.9995
Dimethylformamide	0.025	0.9980
	0.050	0.9947
Ethanol (absolute)	0.25	0.9995
	0.50	0.9981
Ethyl acetate	0.01	0.9964
	0.02	0.9923
Methanol	0.1	0.9997
	0.2	0.9980
Propan-1-ol	0.25	0.9996
	0.50	0.9981
Toluene	0.025	0.9999
	0.050	0.9999

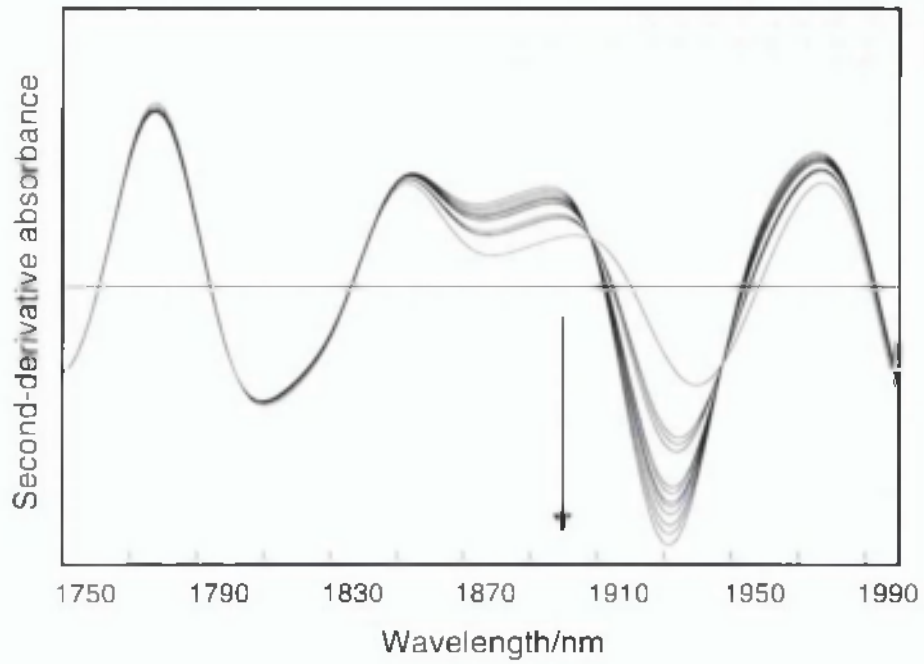


Fig. 3.15 Uptake of water by dimethylformamide over a period of about ten minutes. The direction of the arrow indicates increasing time.

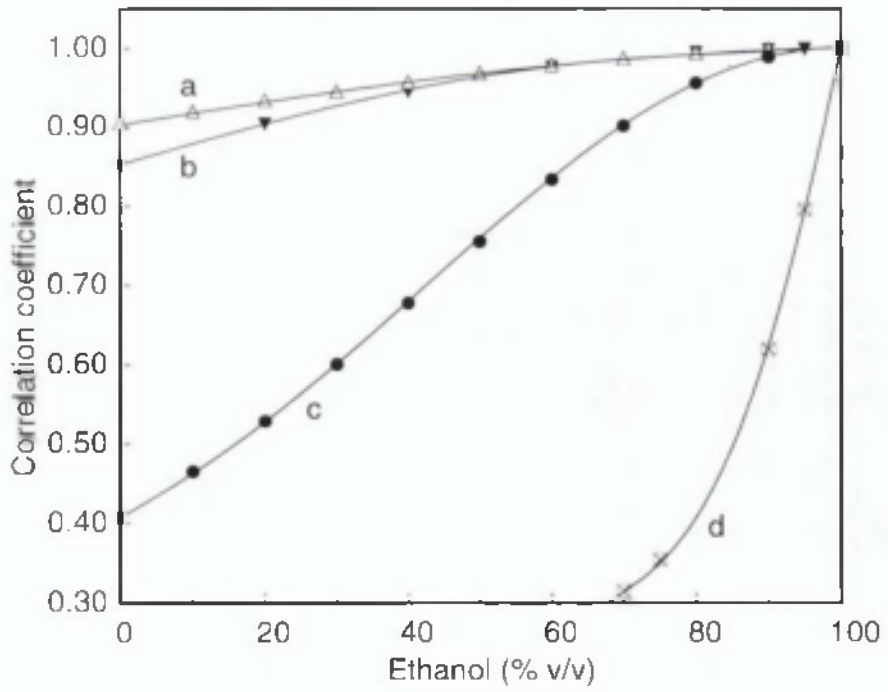


Fig. 3.16 Change of correlation coefficient with solvent composition: (a) ethanol/propan-2-ol, (b) ethanol/methanol, (c) ethanol/acetone and (d) ethanol/water mixtures. Correlation coefficient values with respect to ethanol. Second derivative spectra (9 data-point block size), 1136-2000 nm. Setup 3.

Table 3.10 Correlation coefficients for solvents not in the library. Second-derivative spectra (9 data-point block size), 1136 – 2000 nm.

Sample	Setup	Correlation coefficient	Closet matching solvent
Benzyl alcohol	5	0.889	Toluene
Decane	5	0.744	Butan-1-ol
Dichloroethane	5	0.723	Acetonitrile
	7	0.716	
Dodecane	5	0.677	Butan-1-ol
Ethyl methyl ketone	3	0.819	Acetone
	5	0.823	
	7	0.830	
Hexane	1	0.933	Butan-1-ol
	4	0.907	
	5	0.908	
	7	0.891	
Heptane	3	0.856	Butan-1-ol
	5	0.867	
	7	0.837	
Industrial methylated spirit 74 OP	5	0.994	Ethanol (abs)
Industrial methylated spirit 64 OP	5	0.981	Ethanol (96%)
Nonane	5	0.784	Butan-1-ol
Octane	5	0.824	Butan-1-ol
	7	0.783	
Pentane	1	0.929	Butan-1-ol
	4	0.929	
	5	0.925	
Iso-pentane	5	0.827	Propan-1-ol
	7	0.823	Butan-2-ol
n-Pentan-1-ol	5	0.938	Butan-1-ol
Undecane	5	0.710	Butan-1-ol
Water	5	0.097	Ethanol (96%)

correlation coefficients for a range of common solvents which were not included in the library and apart from industrial methylated spirits (IMS) no mis-identifications occurred. Not surprisingly IMS 74 OP (~95% ethanol, 4% methanol and 1% water) matched absolute ethanol and IMS 64 OP (~90% ethanol, ~4% methanol and 6% water) gave a fairly good match to ethanol 96% ($r = 0.981$). IMS represents a difficult problem and can not be solved using the simple Correlation in Wavelength Space algorithm used in this paper for identification.

3.3.9 External validation

To test the robustness of the library it was used to identify the spectra from 405 different batches of solvents collected over a period of about 3 years (setup 4). **Table 3.11** summarises the results showing that all batches of solvents were correctly identified. In all cases where the r value was lower than normal (< 0.99), visual inspection of the spectra clearly showed the presence of traces of water in the solvents (see **Fig. 3.17**).

Table 3.11 External validation of library. Second-derivative absorbance spectra (9 data-point block size), 1136 – 2000 nm.

Solvent	Number of batches	Correlation coefficient
Acetone	1	0.982
	40	> 0.992
Dimethylformamide	1	0.983
	18	> 0.991
Dichloromethane	42	> 0.998
Ethyl acetate	1	0.992
	68	> 0.996
Methanol	1	0.973
	1	0.990
	125	> 0.993
Propan-2-ol	31	> 0.998
Tetrahydrafuran	19	> 0.998
Toluene	57	> 0.998

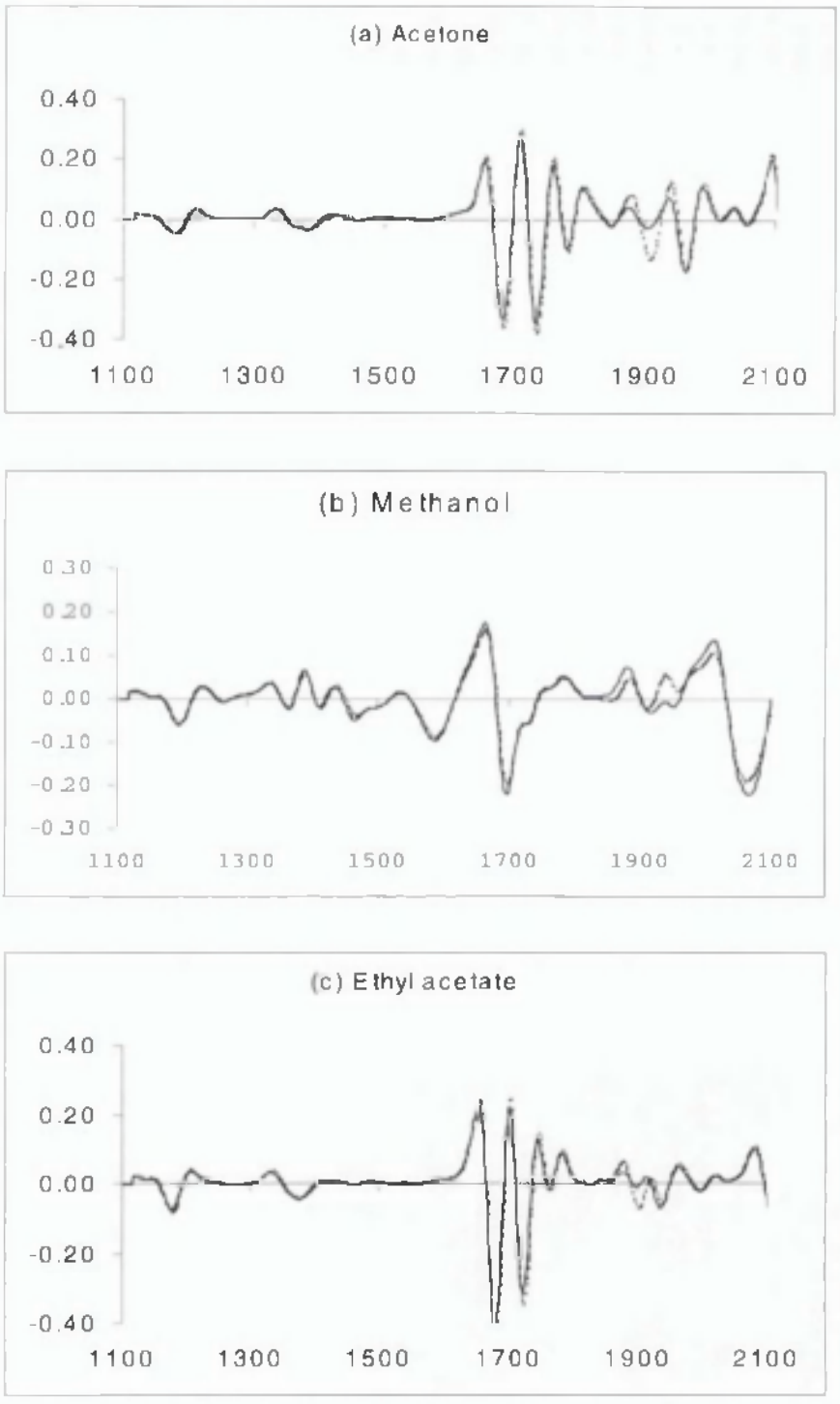


Fig. 3.17 Spectra of solvents from batches with r values < 0.99. Solid lines – reference spectrum, dashed lines – sample spectrum. Presence of water is indicated by spectral differences ~ 1940 nm.

3.4 Conclusion

NIR transflectance spectra measured with respect to an air reference have been shown to be transferable, without any correction, between different instruments not only from the same manufacturer, but also between grating and FT based systems. This was achieved by careful selection of wavelength range and also data-pretreatment parameters (i.e. order of derivative and data smoothing) to obtain a robust method whilst ensuring its effectiveness. A spectral library of fifteen commonly used pure solvents constructed on one instrument could be used to reliably identify solvent spectra measured on other instruments using Correlation in Wavelength Space. A correlation coefficient of > 0.97 was a clear indication of a positive identification, though r values below 0.99 were indicative of the presence of impurities. Small changes in path-length, temperature and trace impurities of water could all be tolerated. Clearly false negative or false positive identifications can never be ruled out, however, using an r value of > 0.99 the data in **Tables 3.4a** and **3.11** would suggest a false negative rate of 1 or 2%. It is important to include all closely related and other likely substances to those of direct interest in the library so that the optimum wavelength range, mathematical pre-treatment *etc.* of spectra are chosen to minimise these errors. The importance of visually inspecting spectra and not just relying on the r value can not be too strongly stressed.

CHAPTER 4: AN INTERLABORATORY TRIAL TO STUDY THE TRANSFERABILITY OF A SPECTRAL LIBRARY FOR THE IDENTIFICATION OF SOLVENTS USING NEAR-INFRARED SPECTROSCOPY

4.1 Introduction

Further to the successful study concerning the transferability of a spectral library between different instruments, it was decided that an inter-laboratory trial should be carried out to test the robustness of the method between different laboratories. Variation between laboratories such as differences in humidity, temperature and operator variability are well known sources of errors (Miller & Miller, 1993; Thompson *et al*, 1993). Therefore, a collaborative trial was considered important to confirm the transferability of the library and method.

A completely transferable spectral library is almost impossible to achieve. Even if all spectrophotometers measured 'perfect' spectra, there is still the problem of storing the data : number of data-points, wavelength or wavenumber scale, number of significant figures, file format *etc*. For a library to be useful it must be accessible to the user.

One common problem of establishing transferability lies in the file formats of digitally coded spectra. Proprietary software packages used for controlling instruments tend to have their own unique file formats. While facilities to export/import spectral data in a variety of formats are usually provided, finding one that is common to all software packages proved difficult. Similarly, no universally accepted method of comparing spectra for identification purposes exists.

The software from three known NIR instrument manufacturers i.e. FOSS NIRSystems, Buhler and Bran+Luebbe have no one common file format between them. FOSS NIRSystems store data in what are called da files (DOS file extension .da, which are binary coded files). Their *NSAS* version 3.52 software provides options for exporting/importing data in Data Interchange Format (.dif) and JDCAMP-DX (.jcm). The more recent FOSS software – *Vision* provides only an ASCII format. *NIRCAL* from Buhler provides *NIRCAL* spectral file (.nsf), Grams (.spc) and JCAMP-DX (.jdx or.dx) while the *Sesame* software from Bran+Luebbe provides IDAS (.dat), JCAMP-DX (.jdx), SpectraCalc (.spc), Unscrambler Ver. 2 (ASCII) and Data Interchange Format (.dif).

Even attempting to transfer a library from one instrument to another of the same manufacturer, but fitted with different sampling accessories caused problems. With

the IQ^2 software (FOSS NIRSystems) a library created on a system fitted with a Rapid Content Analyser could not be transferred to one using a fibre optic probe. The shorter wavelength range (1100 – 2000 nm) causes the noise test to fail; a test which is compulsory before any measurements can be made. Even a simple aspect as a difference in file extension (e.g. .jcm, .jdx or .jdx for JCAMP-DX files) can also easily cause a recognition failure by the software when attempting to transfer data. Although the JCAMP file format is common to most of the software packages, a further problem was encountered: not all the manufacturers actually followed the published JCAMP guidelines (MacDonald and Wilks, 1988) in full.

All these issues present problems when designing an inter-laboratory trial which ideally is open to participants using a range of instruments and software. For these reasons an in-house program was written called IDENT, which compares spectra using Correlation in Wavelength Space. This program could read JCAMP-DX and Vision ASCII files, so allowing data from any of the three NIR instruments manufacturers to be read. The JCAMP file reader was written to be tolerant of the known deviations from the published JCAMP guidelines. Program IDENT is a simple MS-DOS based program using a simple menu system to choose options. Sample spectra are compared against a library of reference spectra and the matches with r values above a pre-determined critical value displayed. An option to allow matched spectra to be visually inspected is also provided. Before spectra are compared, they may be converted to 1^{st}, 2^{nd} *etc.* derivative, using either a simple moving average difference algorithm or a Savitzky-Golay filter. The spectral range over which the correlation coefficient is to be calculated can also be specified. Details of spectral pre-treatments are stored in the library file.

An inter-laboratory trial was organised in which participants were asked to record NIR spectra of pure solvents from a source of their own choice. The recorded spectra were to be returned to the Centre for Pharmaceutical Analysis, The School of Pharmacy, University of London, Brunswick Square, London. The spectra were compared against the master library at the Centre using IDENT. Subsequent to the data analysis, but before the results were revealed to the participants, they were sent a copy of the program and manual. Participants were to asked analyse their own data and return the results to the Centre. Participants were encouraged to measure as many

different solvents as possible, including solvents not represented within the library so that the robustness of the procedure could be assessed.

Participants were recruited by open invitation at the FOSS NIRSystems Pharmaceutical User Group Meeting held at Eli Lilly, Windlesham on the 7th and 8th October 1998, along with invitations to known NIR spectroscopists in industry and academic institutions in Europe. Each participant was given a copy of the protocol and assigned a code for naming computer data files. A copy of the protocol is given in the Appendix I.

4.2. Protocol

The following details the framework and the underlying intentions to the design of the experimental protocol.

4.2.1 Instrumentation

Although the study was open to all NIR *dispersive* and *Fourier Transform* instruments, using either the transflectance or transmission method of measurement, it was assumed that they would be one of the following three:

(1) FOSS NIRSystems Inc (12101 Tech Road, Silver Spring, MD 20904, USA), using any combination of the following monochromators and accessories:

Spectrophotometer : monochromator model 6500 or 5000.

Sampling accessories: Rapid Content Analyser (P/N : AP 1300), Direct Content Analyser (P/N : AP 1200), Interactance Immersion Probe (P/N : NR 6770),Interactance Reflectance Probe (P/N : NR 6775), Smartprobe (P/N: NR 6770), Optiprobe (NR 6770) and the Liquid Cuvette Module (P/N : NR 6509, for transmission).

The Liquid Sampling Kit (P/N : NR 6543, comprising a liquid sampling cup and gold reflector) is also necessary to facilitate the measurement of liquids by transflectance when using the first two accessories listed above, (i.e. horizontal setups). Similarly, appropriate probe tips are required for the various types of probes. As for the Liquid Cuvette Module, a quartz cuvette of 1 mm pathlength should be used.

(2) Bran+Luebbe (GmbH, Norderstedt, Germany). Spectrophotometer : FT-NIR InfraProver II (P/N : 124-A020-01) Sampling accessories : diffuse reflectance fibre optic probe with transflectance cap (P/N: 124-B603-02).

(3) Buhler AG (Uzwil, Switzerland). Spectrophotometer : FT-NIR NIRVIS Sampling accessories : fibre optic probe with transflectance cap.

4.2.2 Solvents

These were to be provided by the participants from a source of their own choice provided they met all the following requirements:

- laboratory grade or better.
- purity ≥ 99 %.
- within expiry date.
- opened within at least one year of measurement.

4.2.3 Sample measurement

Before each measurement session, instruments were to be checked according to the instrument manufacturers' instructions. Participants were requested to submit the results of these checks with their spectra.

Participants were to measure the samples according to the standard operating procedure described in the manual for the instrument used, i.e. run instrument checks, measure reference spectrum and sample spectrum. When using a fibre optic probe the distance between the probe tip and reflector was to be set to using a 1 mm gauge. For measurements using a sample cup, a reflector giving a 1 mm optical path-length was to be used.

All spectra were to be measured with respect to air as reference and recorded over a minimum range of 1100 – 2000 nm (5000– 9996 cm^{-1}). A single spectrum for each solvent was requested.

4.2.4 Data export and collection

Participants were asked to save the spectrum of each solvent as a separate file using a specified coded name. Data files were to be converted to JCAMP format and transferred to a 3.5-inch PC formatted high density (1.44 Mb) disc before sending to the Centre. Alternatively, the JCAMP files could be sent over the Internet to the Centre.

4.3 Results and Discussion

4.3.1 Details about response to study

The protocol was sent to 12 laboratories of which participants from 11 laboratories responded. The laboratories were numbered as 1 to 10. Laboratory 2 scanned the solvents with two different instrumental accessories, the two sets of results were labelled as 2a and 2b respectively. Eight of the participants were from industrial analytical laboratories, one from an academic laboratory, one from a warehouse environment and one from an instrument manufacturer. All these laboratories are different from those encountered in chapter 3.

4.3.1.1 Instrumentation

Table 4.1 lists the instrumentation used by the participants in the trial. Nine laboratories used FOSS NIRSystems instrumentation and one a Bran+Luebbe FT NIR Infraprover spectrophotometer. All except one (laboratory 9) of the FOSS NIRSystems used standard accessories. Laboratory 9 used a Liquid Cuvette Module which had been modified to take 0.8 ml GC vials (P/N: 0.8CPV, external dimensions – 40 x 7 mm, Chromacol Ltd., Hertfordshire, UK) instead of the standard 2 mm path-length cuvettes. **Fig. 4.1** shows a schematic diagram of this module. Radiation from the source passes through the sample and after reflection from the ceramic block falls on to the detectors. This modified module would have a sample path-length of approximately 5 mm and therefore does not meet the requirements of the protocol. Laboratory 2b used a petri dish to contain the solvent rather than a FOSS sampling cup.

4.3.1.2 Solvent spectra

One hundred and seventy three spectra were returned to the Centre of which 139 were internal and 34 were external to the library. The solvents measured by each of the participating laboratories are listed in **Table 4.2** and **Table 4.3**. Appendix II, lists the various suppliers of solvents used by the participants.

Table 4.1 Instrumental setups of participating laboratories

Laboratory number	Spectrophotometer	Sampling accessory
1	FOSS NIRSystems 5000	Direct Content Analyser, Liquid sampling kit
2a	FOSS NIRSystems 6500	Smartprobe
2b	FOSS NIRSystems 6500	Direct Content Analyser, Liquid sampling kit (petri dish instead of liquid sampling cup)
3	FOSS NIRSystems 5000	Direct Content Analyser, Liquid sampling kit
4	FOSS NIRSystems 5000	Direct Content Analyser, Liquid sampling kit
5	FOSS NIRSystems 5000	Optiprobe
6	FOSS NIRSystems 6500	Interactance Immersion Probe
7	FOSS NIRSystems 6500	Direct Content Analyser, Liquid sampling kit
8	FOSS NIRSystems 6500	Transmission Liquid Module
9	FOSS NIRSystems 5000	Direct Content Analyser, fitted with liquid cuvette module
10	Bran+Luebbe, FT-NIR InfraProver	Diffuse reflectance fibre optic probe with transflectance cap

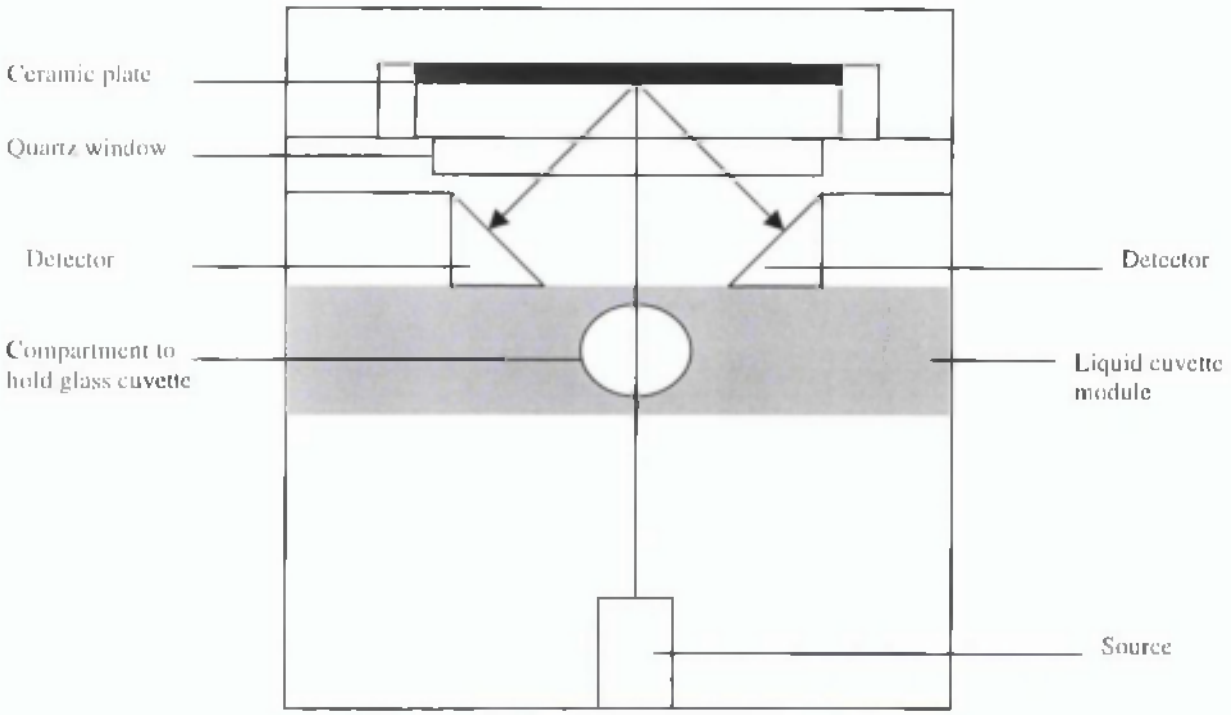


Fig. 4.1 Schematic diagram of the non-standard instrumental setup used in laboratory 9.

Table 4.2 Correlation coefficients obtained for data submitted by participants – internal solvents.

	Laboratory number											
	1	2a	2b	3	4	5	6	7	8	9†	10‡	Range for 1–8 (standard modules)
Acetone (A)	>0.999	0.998	0.999	0.998	0.999	>0.999	0.999	>0.999	-	0.996	(0.968) A	0.998 (n=8)
Acetonitrile (AN)	0.998	0.999	0.999	0.999	0.998	>0.999	>0.999	0.999	0.997	0.994	-	0.997 (n=9)
Butan-1-ol (B1OL)	>0.999	-	-	>0.999	>0.999	>0.999	>0.999	>0.999	-	0.995	(0.766) B1OL	0.999 (n=6)
Butan-2-ol (B2OL)	>0.999	-	-	0.999	-	-	>0.999	(0.904) P1OL	-	0.991	-	0.999 (n=3)
Chloroform (CH)	0.997	0.998	0.998	0.997	0.997	0.999	0.999	0.997	-	0.987	-	0.997 (n=8)
Dichloromethane (DCM)	0.997	0.999	>0.999	0.997	0.998	>0.999	>0.999	0.998	0.994	(0.965) DCM	-	> 0.994 (n=9)
Dimethylformamide (DMF)	0.997	0.996	0.998	0.999	-	0.997	0.998	0.997	-	0.992	(0.778) DCM	> 0.996 (n=7)
Ethanol (abs) (E)	-	-	-	0.999	>0.999	0.991	0.999	0.996	0.996	(0.992) E	0.978	0.995 (n=6)
Ethanol (96%) (E96)	0.999	(0.998) E	(0.999) E	0.999	>0.999	-	>0.999	0.999	0.998	0.996	-	> 0.998 (n=8)
Ethyl acetate (EA)	0.998	0.997	0.998	0.998	0.998	0.999	0.997	0.999	-	0.995	-	0.997 (n=8)
Methanol (M)	0.999	0.999	0.999	>0.999	>0.999	>0.999	>0.999	0.999	0.997	(0.966) M	(0.642) E96	0.997 (n=9)
Propan-1-ol (P1OL)	>0.999	0.999	0.999	>0.999	>0.999	>0.999	>0.999	>0.999	-	0.998	-	> 0.999 (n=8)
Propan-2-ol (P2OL)	0.999	0.998	0.999	0.999	>0.999	>0.999	>0.999	0.999	0.997	0.996	(0.705) E96	> 0.997 (n=9)
Tetrahydrafuran (THF)	0.999	0.999	0.999	>0.999	>0.999	>0.999	>0.999	0.999	0.995	0.975	-	> 0.995 (n=9)
Toluene (TOL)	0.999	0.999	0.999	0.999	>0.999	>0.999	>0.999	>0.999	0.999	0.990	-	> 0.999 (n=9)

Note > 0.999 indicates less than 1, but greater than 0.999

† non-standard module

‡ FT module, all samples regarded as outlier

- no spectrum measured

( ) spectrum incorrectly recognised or rejected, with highest matching solvent indicated.

Table 4.3 Correlation coefficients obtained for data submitted by participants – external solvents.

		Laboratory number										
	Highest matching solvent	1	2a	2b	3	4	5	6	7	8	9†	10‡
1,4 Dioxan	Butan-1-ol	0.781	-	-	-	-	-	-	0.781	-	0.747	-
2-ethoxyethanol	Tetrahydrafuran	-	-	-	0.736	-	-	-	-	-	-	-
2-methyl pentane	Propan-1-ol	-	-	-	-	-	-	-	-	-	-	0.681
Cyclohexane	Butan-1-ol	0.640	-	-	-	-	-	-	0.641	-	-	0.766
Diethylether	Acetone	0.701	-	-	-	-	-	-	0.699	-	0.676	-
DMSO	Acetone	-	0.821	0.816	-	-	-	-	-	-	-	-
Ethyl-methyl ketone	Acetone	-	-	-	-	-	-	-	0.824	-	-	-
Glycerol	Ethanol 95%	-	-	-	-	-	-	-	-	-	-	0.444
Hexane	Butan-1-ol	-	0.902	0.908	-	-	-	-	0.906	0.502	0.917	0.579
Hexene	Butan-1ol	-	-	-	-	-	-	-	-	-	-	0.532
IMS 74 OP	Absolute alcohol	-	0.994	0.993	-	-	-	-	-	-	-	-
n-heptane	Butan-1-ol	-	0.852	0.869	-	-	-	-	0.865	-	-	0.484
n-hexadecane	Butan-1-ol	-	-	-	-	-	-	-	-	-	0.767	-
Octane	Ethanol 96%	-	-	-	-	-	-	-	-	-	-	0.520
Petroleum spirit	Butan-1-ol	-	-	-	-	-	-	-	-	-	-	0.750
Tetrachloroethlynenet	Ethanol 96%	-	-	-	-	-	-	-	-	-	-	0.573
Triethylamine	Propan-2-ol	0.815	-	-	-	-	-	-	-	-	-	-
Xylene	Toluene	-	-	-	-	-	-	-	-	0.891	-	-

† laboratory using non standard module
‡ FT module, all samples regarded as outlier

4.3.2 Data analysis

Returned spectra were compared, using Correlation in Wavelength Space, against the master library of the fifteen solvents which had been recorded using a FOSS NIRSystems 6500 scanning spectrophotometer fitted with an Interactance Immersion Probe (setup 1; see chapter 3, section 3.2.2). The same optimum pre-treatments found in chapter 3 were used : 9 data-point block size, second derivative spectra and wavelength range 1136 to 2000 nm. The in-house written computer program IDENT was used to perform the data analysis.

Cubic spline interpolation was used to convert spectra measured in wavenumbers to equally spaced data-points in terms of wavelength. The wavenumber *vs* transmittance spectra were first converted to absorbance spectra and then absorbance values at 2 nm intervals over the range 1100 to 2498 nm to give 700 calculated values. Out of range vaues were set to zero (see chapter 3, section 3.2.4 for further details).

4.3.2.1 Recognition of internal solvents

For a positive identification, the threshold values for the dot product correlation coefficient was set to > 0.97 – optimum values deduced in chapter 3. **Table 4.2** summarises the correlation coefficient values for solvents internal to the reference library. Sample spectra with $r \leq 0.97$ and shown in parentheses were initially recorded by program IDENT as 'unknown'. Samples with values of $r > 0.97$ shown in parentheses were wrongly identified.

All the spectra from laboratory 10 except for the sample of absolute alcohol were classified as 'unknowns'. From visual inspection of the spectra, it was evident that there was a problem in the measurement procedure and it was not just a matter of low r values. Spectra were broad and similar to one another suggesting sample contamination and/or inadequate cleaning of the measuring probe. Although laboratory 10 used an FT instrument with the original data in wavenumbers, this was not considered to be the underlying reason for the discrepancy observed. Even direct comparison of the spectra with an FT spectral library (chapter 3, setup 7) gave no better results with r values remaining low (e.g. butan-1-ol, dichloromethane,

methanol, and propan-2-ol gave 0.584, 0.733, 0.793 and 0.844 respectively). For these reasons, the results from laboratory 10 were all considered as 'outliers' and not further considered in this study.

The butan-2-ol sample spectrum returned by laboratory 7 was classified as 'unknown'. Best match solvents were propan-1-ol (r = 0.904) followed by butan-2-ol (r = 0.884). Visual inspection of the spectrum suggested either the presence of an unknown impurity or an error during the measurement process, **Fig. 4.2**. The participant was requested to re-measure the sample, however, the original batch of butan-2-ol was no longer available. A new batch was measured and the spectrum correctly identified (r = 0.999).

Two spectra labelled 'ethanol 96%' from laboratories 2a and 2b were identified as absolute alcohol (r = 0.998 and 0.999 respectively). Checking with the participant revealed that the sample was in fact absolute alcohol and that the spectral files had been mis-labelled.

Sample spectra measured using the modified liquid cuvette module, laboratory 9, gave r values on the low side; two samples being recorded as 'unknown' (dichloromethane, r = 0.965 and methanol, r = 0.966). In all cases, however, the highest match solvent was correct. As previously noted the modified cuvette module did not meet the requirements of the protocol, the optical path-length being too long. As found in chapter 3, section 3.3.4, spectra of strongly absorbing solvents will be distorted when such long pathlengths are used. This is in full accord with the trend in r values given in **Table 4.2**. The r values for the strongly absorbing solvents; dichoromethane, chloroform, methanol and tetrahydrofuran being lower than for the more weakly absorbing solvents such as acetone and acetonitrile.

After the corrections for the problem spectra from laboratories 2a, 2b and 7, it is seen that all laboratories meeting the requirements of the protocol (laboratories 1 – 8) gave spectra that were correctly identified. Values for r were all very high, the lowest being

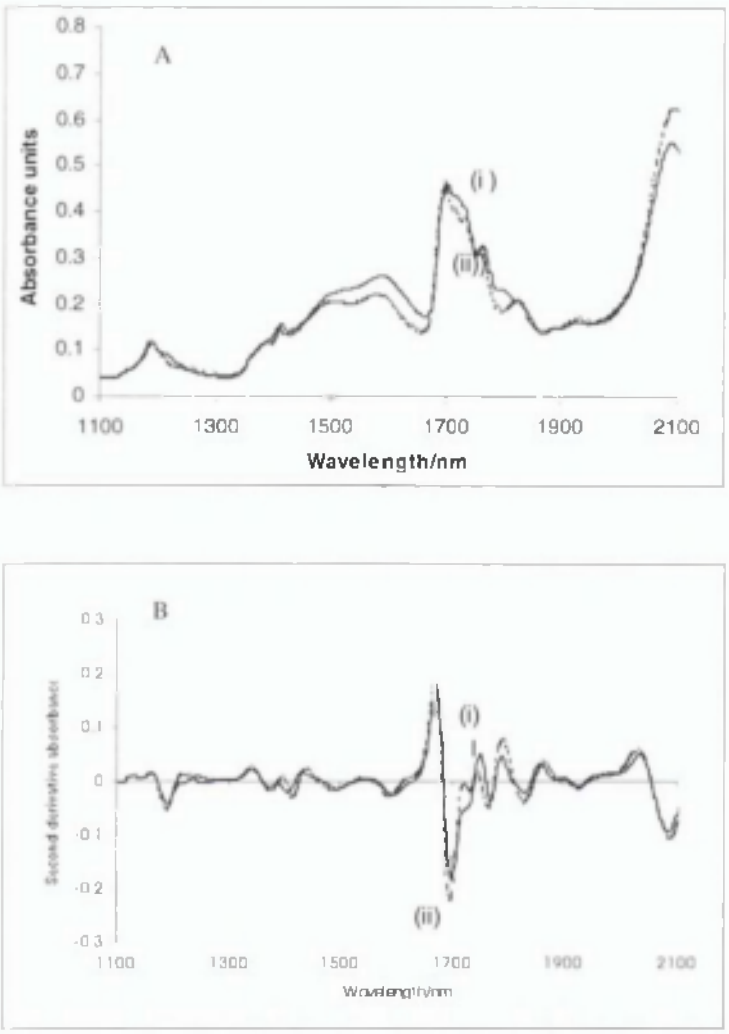


Fig. 4.2 Spectra of butan-2-ol from laboratory 7

(A) Absorbance spectra.
(B) 2nd derivative spectra, 9 data-point block size.
(i) Spectra of anomalous sample, $r = 0.904$.
(ii) Spectra of follow-up sample, $r = 0.999$.

0.994 (dichloromethane, laboratory 8), but still significantly above the pre-set threshold of > 0.97 required for identification.

4.3.2.2 Solvents external to the reference library

Thirty four spectra covering thirteen solvents not included in the reference library were returned to the Centre, **Table 4.3.** Apart from IMS 74 OP (Industrial Methylated Spirit), all solvent spectra were correctly identified as 'unknown' (i.e. $r \leq 0.97$). Hexane from laboratory 9 gave the closest match to any solvent in the library with an r value of 0.917 (hexane/butan-1-ol pair), but still well below the pre-set threshold.

IMS 74 OP was identified as absolute alcohol and presented an interesting challenge to the method. IMS grade 74 OP consists of 95% ethanol, 4% methanol and 1% water. The difference in spectra between it and absolute alcohol is minimal, but visible (see **Fig. 4.3**). The water absorption at ~ 1940 nm is clearly detectable. Differentiation between these two solvents would require the use of a more sensitive algorithm such Polar Qualification System or the Maximum Distance method as noted in chapter 3, section 3.3.8., it is not possible using Correlation in Wavelength Space.

4.3.2.3 Validation of IDENT software

To ensure the holistic transferability of the method, the IDENT software and a User Manual were also sent to the participants so they could check the identity of the solvents which they has previously measured for themselves. Four of the participants responded (the rest of the participants who did not respond cited a lack of time amongst their reasons for not being able to carry out this part of the trial). The results of their analysis were identical to that encountered in Table 4.2 and 4.3 and they also found the program simple to use.

The outcome of this part of the trial had demonstrated that, with the availability of a simple universal software, the transferability of a *method* based on NIR can be achieved.

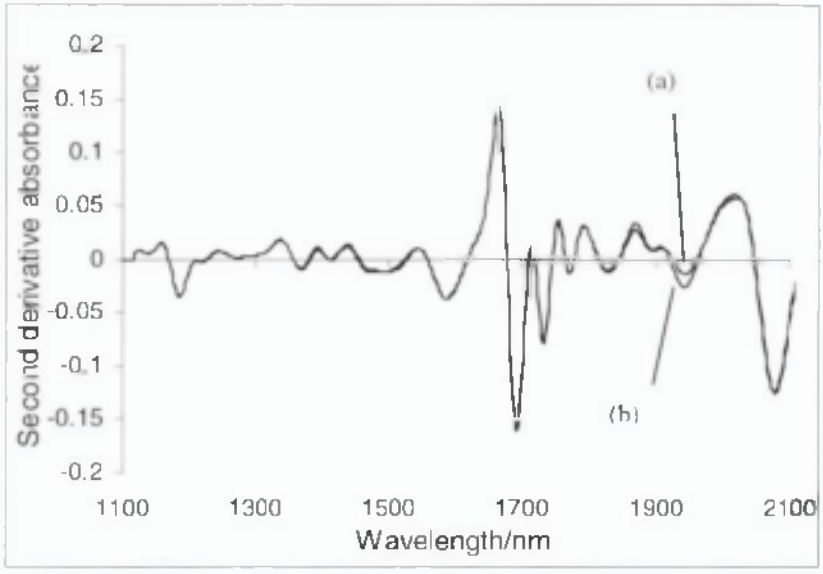


Fig. 4.3 Second derivative spectra of (a) absolute alcohol and (b) Industrial Methylated Spirit 74 OP.

4.4 Conclusion

A spectral library for the identification of common solvents using NIR spectroscopy has been successfully transferred between different laboratories. Solvents were identified with an effectively 100% success rate. Considering the wide range of sampling accessories and variety of solvent suppliers used the results clearly demonstrated robustness of the library and method.

Apart from Industrial Methylated Spirit 74 OP, a range of common solvents external to the library were correctly excluded. It was unfortunate that the data from the FT NIR instrument was not usable and had to be excluded as this would have allowed the transferability between different manufacturers' instruments in different laboratories to be established.

The formats used for storing spectral data and their compatibility with different software packages still remains an issue despite the publication of the JCAMP-DX file format specification as a transferable format of spectral data since more than a decade ago (MacDonald and Wilks, 1988). In this study, it could only be resolved by writing in-house software. Once again the importance of visually inspecting spectra and not just relying on the *r* value can not be too strongly stressed.

CHAPTER 5: EFFECTS OF SAMPLE PRESENTATION ON THE SPECTRA OF SOLID PHARMACEUTICAL EXCIPIENTS

5.1 Introduction

Relative diffuse reflectance measurements of solids are subject to more sources of variation than the transflectance measurements of liquids previously described (chapters 3 and 4). With solids, sample presentation is an important factor – the degree of compaction affecting the spectra (Mark and Workman, 1986). Also, unlike transflectance measurements which can be referenced to a universally available standard such as air, relative diffuse reflectance measurements of solids need to be made with respect to some reference standard. In this chapter, the effects of sample presentation on the spectra of powdered samples are examined.

The most commonly used sample presentation methods in reflectance NIR use either a fibre optic probe or sample cups. Despite their wide use, there is still little information concerning the necessity to standardise parameters such as cup size, amount of sample, pressure when inserting the probe *etc.* Experience gained from the food and agricultural industry suggests that sample presentation is a variable which must not be overlooked. Williams (1992) found that cell loading affected the precision of protein determination more than sample grinding. He also reported that variations in bulk density of the sample could lead to error. Mark and Tunnel (1985) reported that variations in packing affected the calibrations which they developed for the measurement of moisture, fat and protein in ground beef, mixed animal feed and breakfast cereal. It was necessary to make multiple measurements of the same sample to average out the variation due to sample presentation.

In this work, the effects of sample presentation when using a sample cup module on the reflectance spectra of some commonly used pharmaceutical excipients have been systematically examined. The effects of the resulting spectral changes on the Correlation in Wavelength Space and Maximum Wavelength Distance algorithms will also be looked into. By standardising and eliminating factors responsible for spectral variation, it is hoped that it will be possible to establish transferable libraries of spectra.

5.2 Experimental

5.2.1 Apparatus

A 6500 model spectrophotometer (FOSS NIRSystems, Silver Spring, USA) fitted with a Rapid Content Analyzer (RCA) or a Direct Content Analyzer (DCA) was used for the measurement of all reflectance spectra over the wavelength range of 1100 to 2500 nm. Other than where indicated the RCA attachment was used for all investigations. Each recorded spectrum was the average of 32 scans.

Samples were measured in flat based cups :
(a) quartz (standard sample cup, 52 mm diameter, FOSS catalogue number : NR7072)
(b) Pyrex glass (reflectance vessel, 40 mm diameter, FOSS catalogue number : NR6544)
(c) clear neutral glass (Fbg-Anchor glass vials, 21 mm diameter, catalogue number : BDH/Merck/215/0074/23)
(d) soda glass (Philip Harris specimen tubes, 23 mm diameter, catalogue number : PHI3 T82-528)

Particle sizes of the excipients were determined using Malvern MastersizerS (P/N: MAM 5004, Malvern Instruments, Malvern, UK) fitted with an automated dry powder feeder (P/N: MAM 2461), compressor (P/N: COM 0008) and vacuum cleaner (P/N: VAC 001). Particle shape was determined by examining the scanning electron micrographs taken using the Philips XL20 scanning electron microsocope (Philips Electron Optics, Eindhoven, Netherlands).

5.2.2 Materials

All excipients were of pharmaceutical grade: microcrystalline cellulose (Avicel PH101 & Avicel PH102, FMC Corp.), sodium starch glycollate (Explotab, Edward Mendells Corp.), anhydrous dibasic calcium phosphate (A-TAB, Albright and Wilson Ltd.), dibasic calcium phosphate dihydrate (Emcompress, Edward Mendells Corp.), lactose monohydrate (Broculo Whey Products UK Ltd.), hydroxypropyl methylcellulose (Methocel E5 Premium, Colorcon Ltd.), purified talc (Luzenac Europe), propyl and butyl *para* hydroxybenzoate (Nipa Laboratories Ltd) and Kollidon 25 & 30 (povidone or polyvinylpyrrolidone, BASF plc).

5.2.3 Data Treatment

The *NSAS* Version 3.52 software was used for the calculation of first derivative and second derivative spectra using a segment size of 20 data points and gap size of 0 data points. Standard normalisation of the absorbance spectra were carried out using an in-house program, DNA written in C-language.

The effects of sample presentation were quantified using Correlation in Wavelength Space and Maximum Wavelength Distance (FOSS NIRSystems, 1996). The correlation coefficient (r_{jk}) between the absorbances or mathematically transformed values, x, of two spectra j and k measured at p wavelengths was calculated according to **equation 1.19**. Maximum Wavelength Distance ($d_{jk,}$) was calculated using **equation 1.21.**

5.3 Results and Discussion

Reflectance spectra were found to be affected by sample cup diameter, sample thickness, cup material and packing method.

5.3.1 Cup diameter

This was investigated by placing a quartz sample cup on the adjustable iris diaphragm. This elevated the sample to 1 cm above the sample stage, (see **Fig. 5.1**). Spectra were measured over the cup diameter range of 4 to 50 mm. Increasing the sample diameter resulted in downward multiplicative shifts to the absorption spectra, with absorbance values at the shorter wavelengths decreasing much more rapidly than at longer wavelengths, (e.g. **Fig. 5.2**). Peaks also became increasingly well-defined, with increasing diameter. All these effects stabilised towards an 'infinite diameter' which was dependent upon detector-geometry. For the RCA, this 'infinite diameter' was also found to be a linear function of the elevation from the quartz window. The infinite diameters for an elevation of 10, 15 and 20 mm were found to 28, 32 and 36 mm respectively. These values can be represented by the equation:

$$y = 0.8x + 20$$

Where y is the infinite diameter (mm)

x is the elevation from the sample stage (mm)

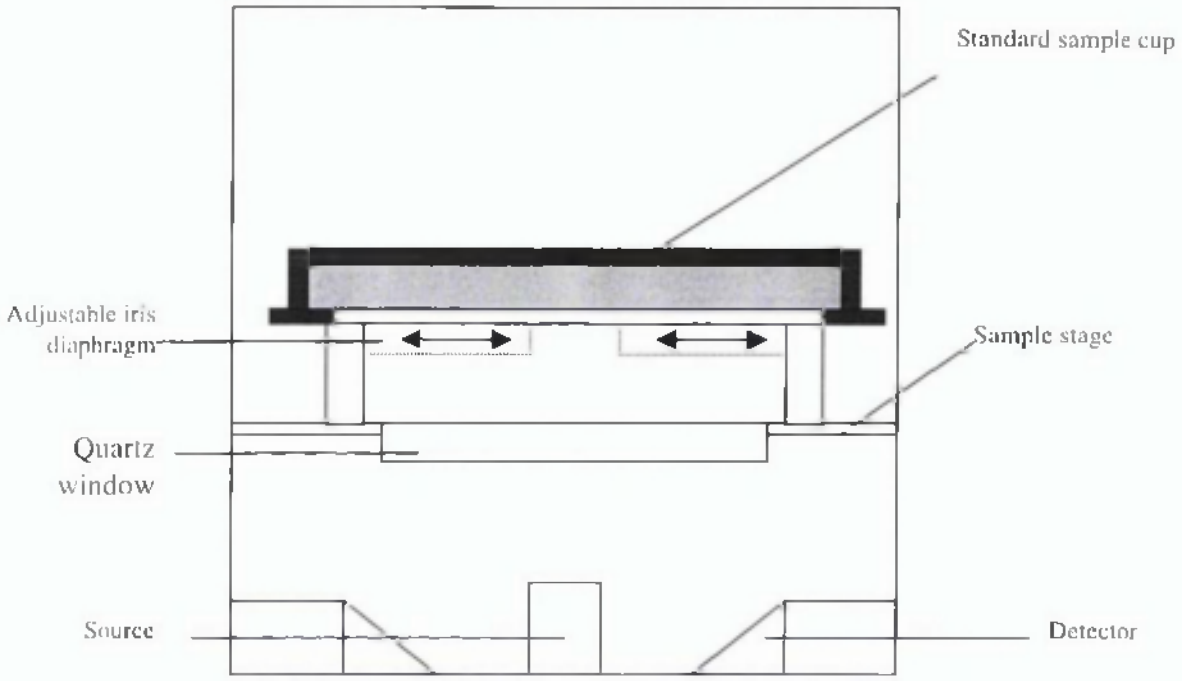


Fig. 5.1 Schematic diagram illustrating the measuring procedure used to investigate the effects of sample cup diameter. A standard sample cup was placed above the iris diaphragm which allowed the effective sample diameter to be adjusted. Diagram not drawn to scale.

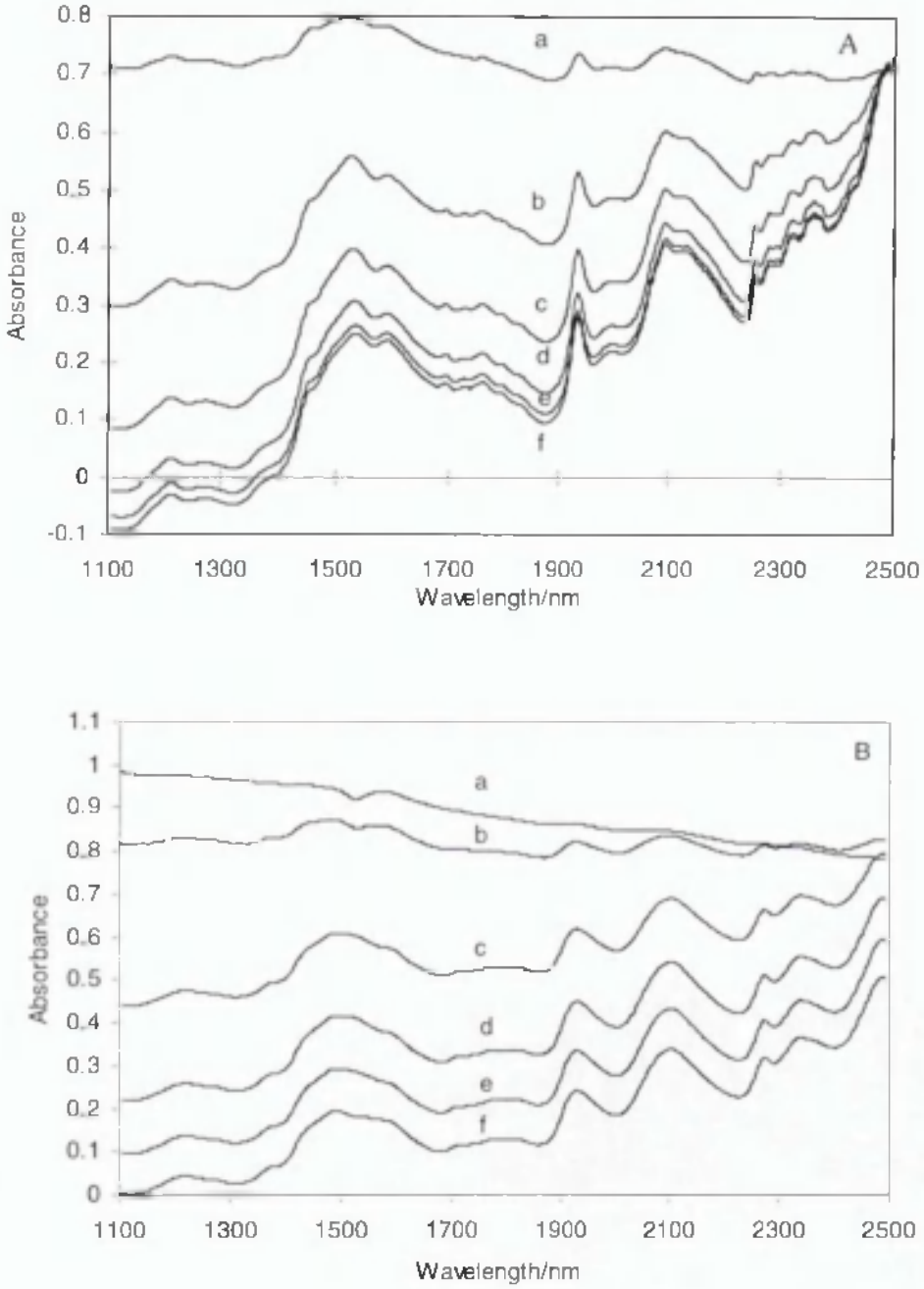


Fig. 5.2 NIR spectra of (A) lactose monohydrate and (B) microcrystralline cellulose using different sample cup diameters. a, 4 mm; b 8 mm; c 12 mm; d 16 mm; e, 20 mm and f, 50 mm.

For zero elevation (normal sample position), this gives an 'infinite diameter' of 20 mm. The DCA which has a different detector geometry was found to have an 'infinite diameter'of 35 mm.

Below the'infinite diameter', spectral shape was found to be a function of sample diameter and this was most noticeable with second derivative spectra. Peak amplitudes relative to the most intense peak in the second-derivative spectrum varied with sample diameter. This is illustrated in **Fig. 5.3A** which shows the variation of peak amplitude for a selection of peaks in the spectrum of Kollidon 25 as a function of sample diameter. Surprisingly, peak amplitudes relative to the most intense peak in the second-derivative spectrum varied with sample diameter. This effect is seen in **Fig. 5.3B** and **C** in which for clarity the peak ratios have all been normalised to 1 at the maximum sample diameter measured. Generally, for peaks at wavelengths greater than the position of the most intense peak, the ratio generally increased with diameter, while peaks at shorter wavelengths showed the opposite effect. The cause of this effect is not clear although probably related to stray-light effects. It was observed for all the excipients investigated. As the sample diameter is reduced below the 'infinite diameter', Wood's peak (1520 nm in second derivative absorbance spectra; Wood 1902) appears and becomes increasingly prominent with decreasing diameter. Greater peak position shifts were also observed with spectra obtained with smaller cup diameters. At 8 mm where most peaks become sufficiently defined to be discernible on the second derivative spectra, peak position shifts of up to 6.6 nm were observed (see **Table 5.1**), proving to be significant as the wavelength accuracy of the instrument was to be within 0.3 nm. All the spectral distortions mentioned above (i.e. shifts in absorbance values, changes in spectral shape, occurrence of Wood's peak and peak position shifts) could not be compensated with mathematical treatments such as derivatisation or standard normal variate normalisation.

Changes of sample diameter were found to have pronounced effects on the values of the identification algorithms; Correlation in Wavelength Space and Maximum Wavelength Distance which are commonly used for the identification of excipients in the pharmaceutical industry. This is illustrated in **Fig. 5.4**, which shows how r_{jk} and d_{jk} vary with sample diameter using spectra recorded using a sample diameter of 50 mm as reference. Larger diameter cups provide more spectral information which can

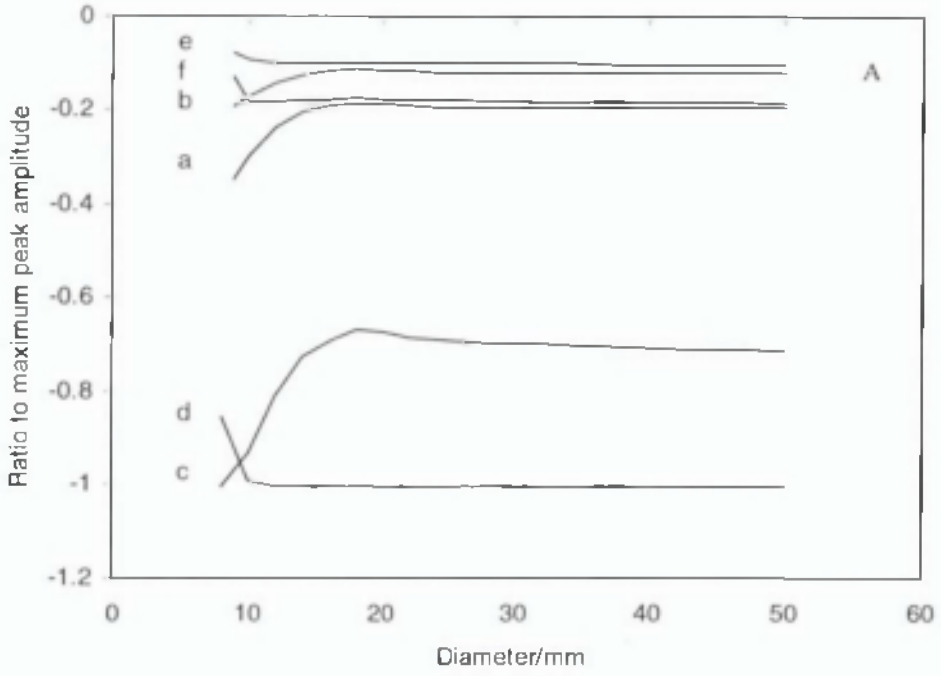
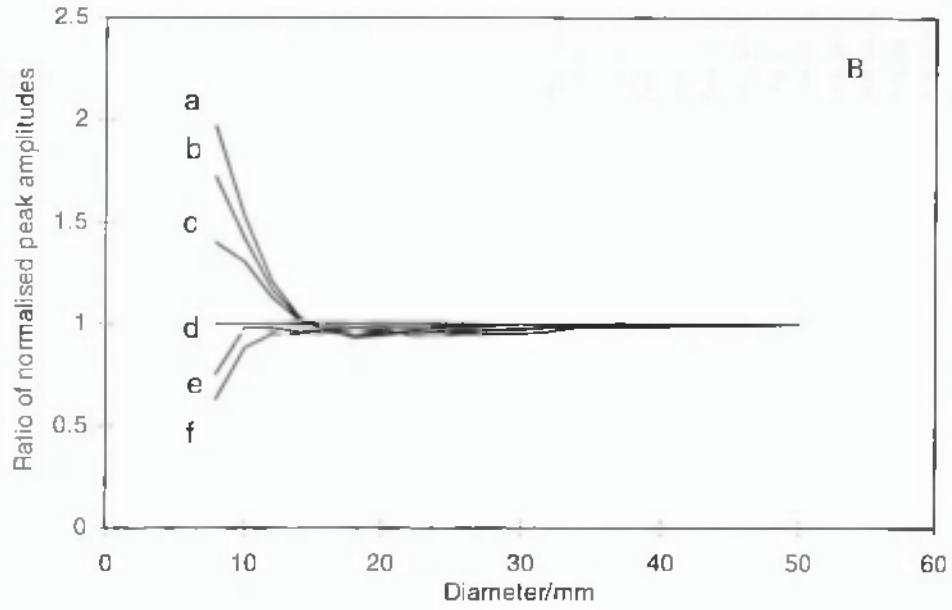


Fig. 5.3 (A) Effects of varying the sample diameter on the relative peak amplitudes for the second derivative absorbance spectra for Kollidon 25 at wavelengths a, 1372 nm; b, 1430 nm; c, 1695 nm; d, 2274 nm e, 2374 nm and f, 2465 nm.



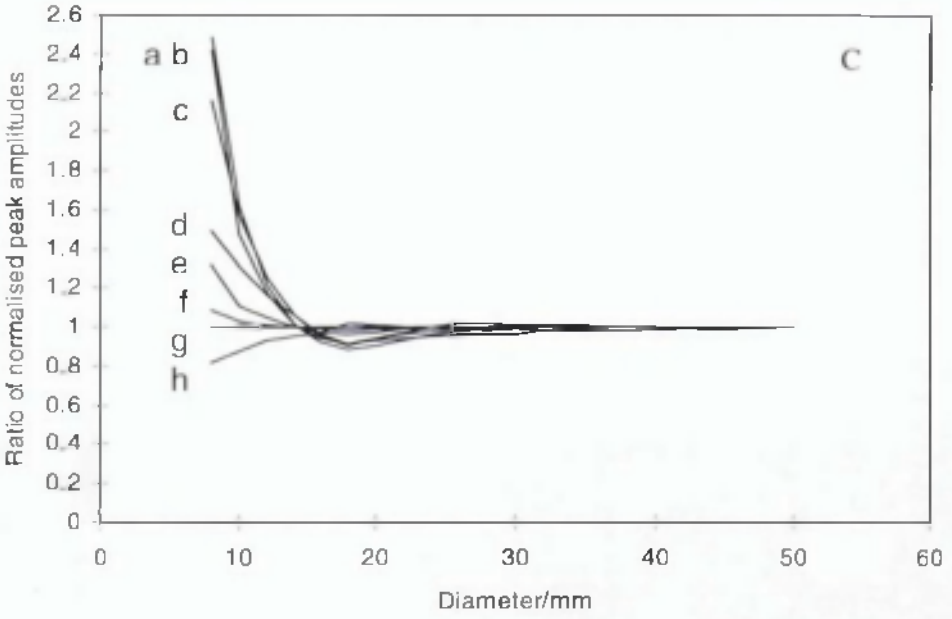


Fig. 5.3 Continued.
Effects of varying the sample diameter on the normalised relative peak amplitudes for the second derivative absorbance spectra for:
(B) Kollidon 25 at wavelengths a, 1372 nm; b, 1430 nm; c, 1695 nm; d, 2274 nm (maximum peak); e, 2374 nm and f, 2465 nm.
(C) Emcompress at wavelengths a, 1172 nm; b, 1479 nm; c,1431 nm; d,1772 nm; e, 2085 nm; f, 2000 nm; g,1930 nm (maximum peak) and h,2455 nm.

Table 5.1 Effect of sample diameter on the position of the six most intense peaks of *Emcompress*.

Diameter/mm	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5	Peak 6
8	1929.3†	1430.5†	-	1770.0	1172.0†	2449.2†
10	1929.3	1430.8	1478.4	1770.9†	1172.2	2452.5
12	1929.4	1430.9	1476.7	1771.3	1172.2	2452.6
14	1929.5	1431.0	1476.2	1771.5	1172.2	2453.4
16	1929.6‡	1431.1‡	1476.2†	1771.7	1172.2	2454.6
18	1929.6	1431.1	1476.5	1771.7	1172.2	2455.3
20	1929.6	1431.0	1477.2	1771.7	1172.2	2455.3
22	1929.5	1431.0	1477.9	1771.6	1172.1	2454.5
24	1929.5	1431.0	1478.6	1771.6	1172.1	2454.8
26	1929.5	1431.0	1479.0	1771.6	1172.1	2454.4
28	1929.5	1431.0	1479.2‡	1771.6	1172.1	2454.6
30	1929.5	1431.0	1479.2	1771.6	1172.2	2454.6
32	1929.5	1431.0	1479.2	1771.7	1172.2	2454.9
34	1929.5	1431.0	1479.2	1771.8	1172.3	2455.7
38	1929.5	1431.0	1479.2	1771.9	1172.5	2455.1
42	1929.5	1431.0	1479.2	1772.0	1172.5	2455.7
46	1929.5	1431.0	1479.2	1772.0	1172.6	2455.8
50	1929.5	1431.0	1479.2	1772.0‡	1172.6‡	2454.8‡
Maximum difference/nm	0.3	0.6	3.0	1.1	0.6	6.6

†minimum wavelength/nm
‡maximum wavelength/nm

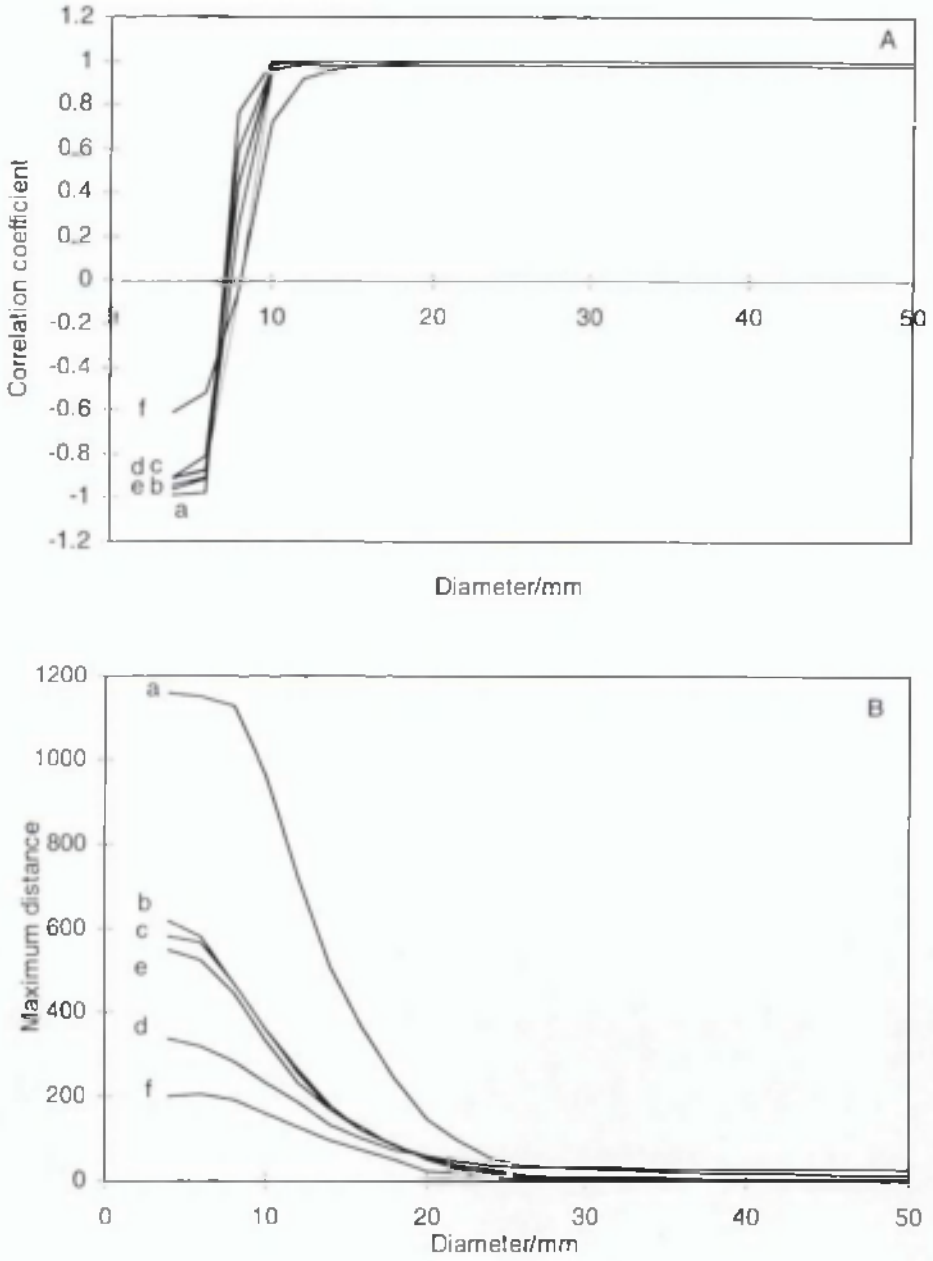


Fig. 5.4 Effects of varying the sample diameter on (A): Correlation in Wavelength Space and (B): Maximum Distance values, compared with reference spectra measured using a diameter of 50 mm. Excipients : a, Emcompress; b, A-TAB; c, Avicel PH102; d, lactose monohydrate; e, Methocel E5 Premium and f, purified talc.

enhance the ability to distinguish between closely related substances. **Table 5.2** shows the effects of diameter on the Correlation in Wavlelength Space and Maximum Wavelength Distance parameters for three closely related pairs of compounds (see **Fig. 5.5**). Avicel PH101 and 102 differ only in their nominal mean particle size (50 μm and 100 μm respectively) and are just distinguishable when using the large diameter cup by the Maximum Wavelength Distance parameter. A value of $d_{jk} > 6$ is generally considered to indicate significant difference (FOSS NIRSystems, 1996). Propyl and butyl *para* hydroxybenzoates are clearly distinguishable by both r_{jk} and d_{jk} parameters using the larger diameter cups. A value of $r_{jk} < 0.95$ is considered significant (FOSS NIRSystems, 1996). Kollidon 25 and 30 differ in relative molecular mass (30 000 and 50 000 respectively) and again are distinguishable using d_{jk}.

5.3.2 Sample thickness

The effects of changing sample thickness were investigated over the range of 1 to 25 mm by weighing increasing masses of sample into a sample cup (clear neutral glass, 21 mm diameter), measuring the sample thickness and recording the spectra. The spectral features became increasingly well-defined and the absorbance baseline shifted downwards with increasing sample thickness, though the effect was less pronounced than seen with changes of sample diameter (**Fig. 5.6A**). Reflectance values became independent of sample thickness above a certain value -'infinite thickness' (**Fig. 5.6B**). However, unlike the position with sample diameter, the 'infinite thickness' was dependent on the sample material (see **Table 5.3A**). Both identification algorithms were found to be sensitive to the effects of sample thickness. **Fig. 5.7** shows this for Kollidon 25.

The existence of an 'infinite thickness' has long been recognised and known to be affected by the samples' physical characteristics (i.e. particle size, distribution, shape, bulk density *etc.*) and chemical nature i.e. absorptivity (Williams, 1992; Olinger and Griffiths, 1993). Plots of bulk density, degree of stratification and mean particle size against 'infinite sample thickness' were all very scattered indicating no simple relationships between the variables, **Table 5.3B** and **Fig. 5.8**. Attempts to predict the infinite thickness of a sample directly from any measurable physical properties such as bulk density, degree of stratification upon tapping (by determining the decrease in

Table 5.2 Effect of sample diameter on the ability to distinguish between closely related substances using NIR absorbance spectra

Pairs of closely related substance	Mean Correlation Coefficient*		Mean Wavelength Distance*	
	Sample diameter/mm		Sample diameter/mm	
	6	52	6	52
Avicel PH101 and Avicel PH102	1.000	1.000	2.809	7.934
Propyl and butyl *para* hydroxybenzoate	0.996	0.948	8.08	39.338
Povidone 25 and Povidone 30	0.999	0.994	8.368	14.625

* Mean spectrum of first excipient named was used as the reference spectrum

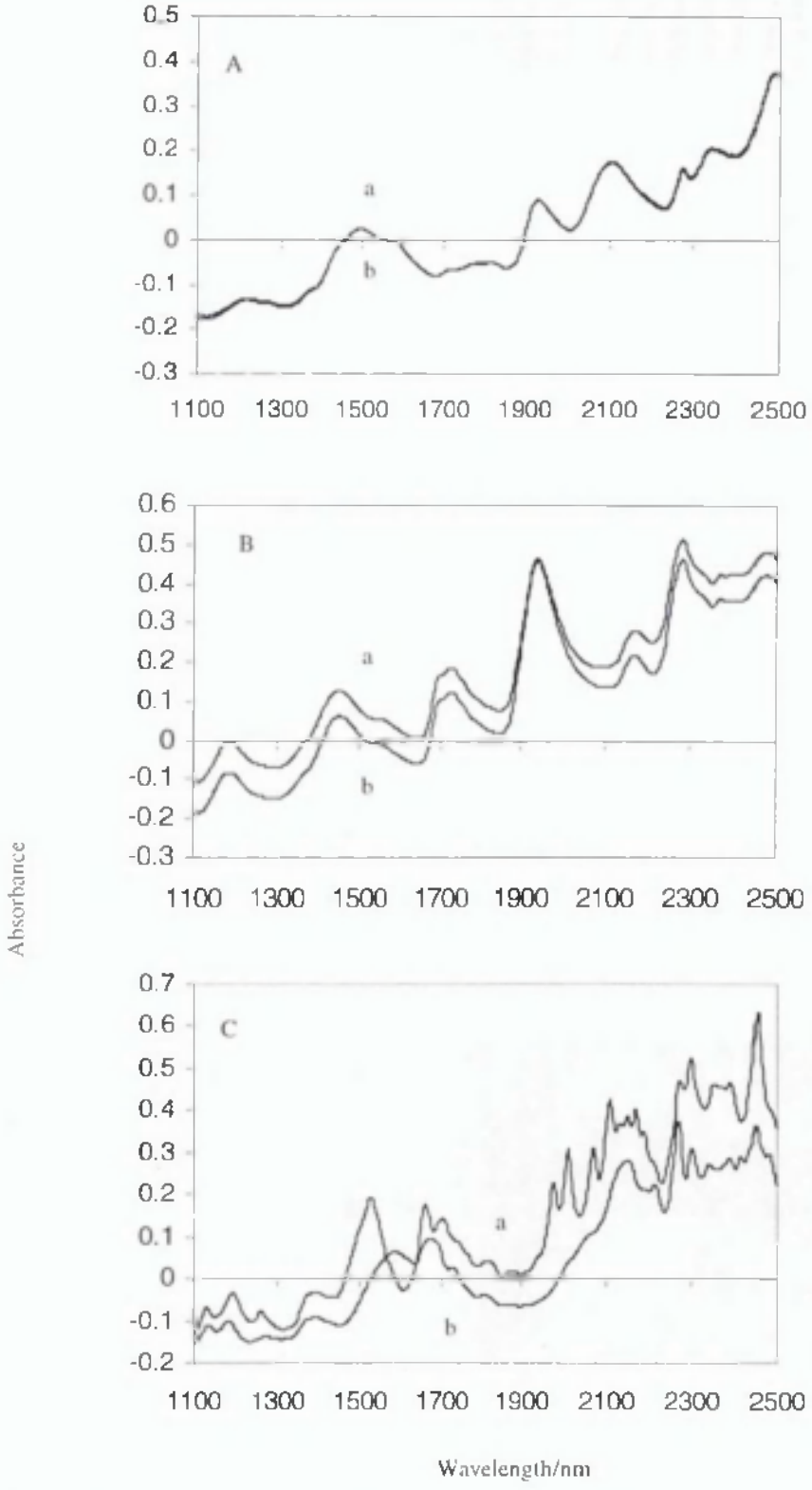


Fig 5.5 Absorbance spectra of:
(A) a, Avicel PH101 and b, Avicel PH102
(B) a, Povidone 25 and b, Povidone 30
(C) a, butyl *para* hydroxybenzoate and b, propyl *para* hydroxybenzoate

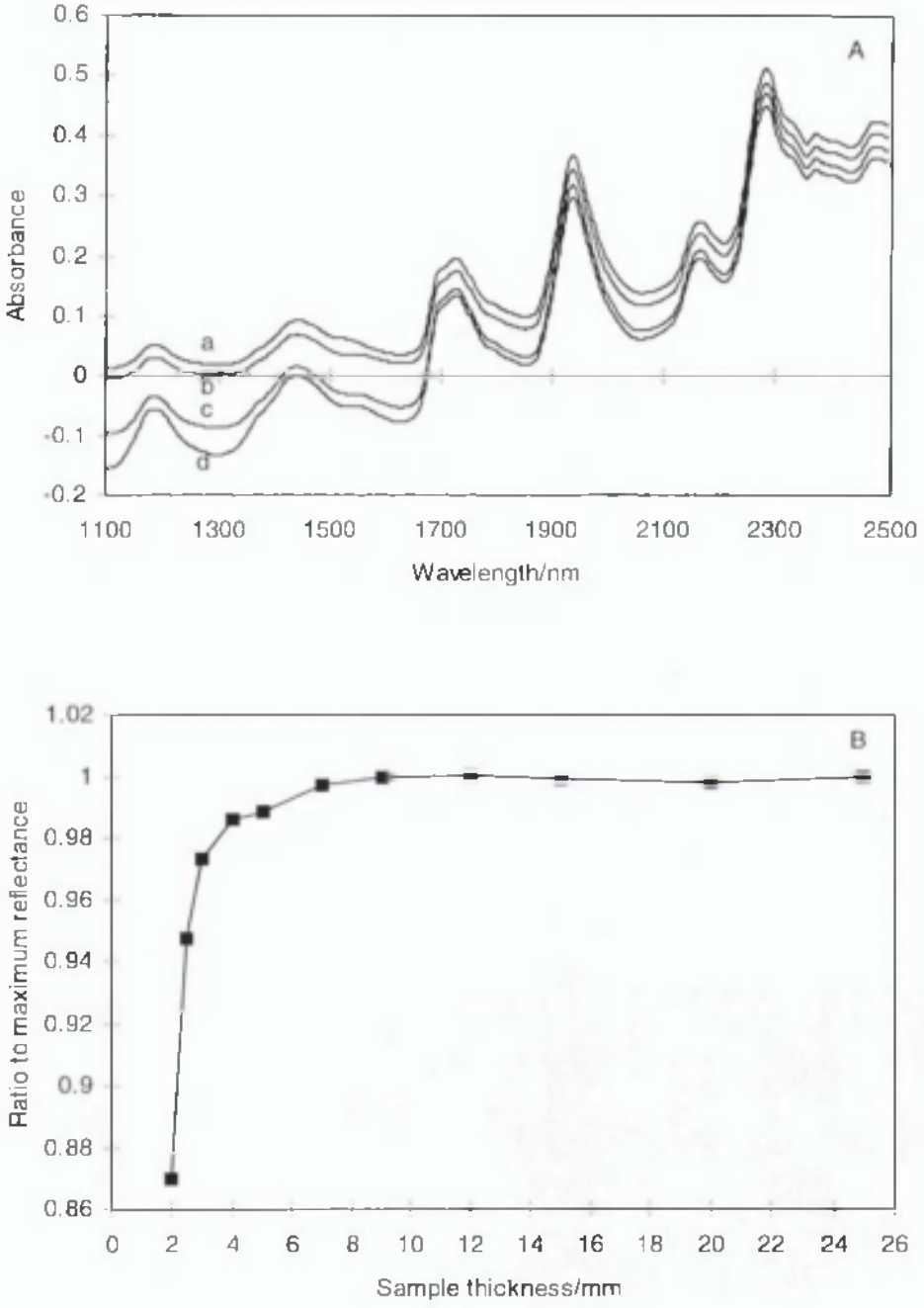


Fig. 5.6 Spectra for Kollidon 25 at different sample thicknesses, (A): a, 1 mm; b, 2 mm; c, 3 mm and d, 25 mm, and (B): dependence of reflectance values at 1100 nm on sample thickness for Kollidon 25. Reference spectra recorded at 25 mm thickness.

Table 5.3

(a) Infinite sample thickness of various pharmaceutical excipients.

Excipient	Infinite sample thickness (mm)
Sodium starch glycollate (Explotab)	3
Lactose monohydrate, regular	3.5
Dicalcium phosphate dihydrate (Emcompress)	4
Microcrystalline cellulose (Avicel PH 102)	5
Hydroxy propyl methylcellulose (Methocel E5)	5
Purified talc	6
Dicalcium phosphate anhydrous (A-TAB)	6
Povidone (Kollidon 25)	9

(b) Various physical parameters of samples

Excipient	†Bulk density (gcm^{-1})	‡Percentage decrease in bulk volume before and after tapping (%)	Mean particle size (μm)
Sodium starch glycollate (Explotab)	0.74	9.8	45.3
Lactose monohydrate, regular	0.62	10.0	87.14
Dicalcium phosphate dihydrate (Emcompress)	1.06	31.0	80.14
Microcrystalline cellulose (Avicel PH 102)	0.37	13.2	-
Hydroxy propyl methylcellulose (Methocel E5)	0.57	15.4	97.07
Purified talc	0.45	31.9	27.61
Dicalcium phosphate anhydrous (A-TAB)	0.52	31.0	-
Povidone (Kollidon 25)	0.39	14.1	-

†Determined using ~ 25 g of material placed in a 250 ml graduated cylinder.

‡Determined by placing ~ 200 ml of material in a 250 ml graduated cylinder. After reading the loose bulk volume, the sample was placed on a Jolting Volumeter Model JV1 (Tap Density Tester, Copley Instruments, Nottingham). The volume of each 100 tap increment was recorded. The tap volume was defined at the volume where three consecutive readings were the same.

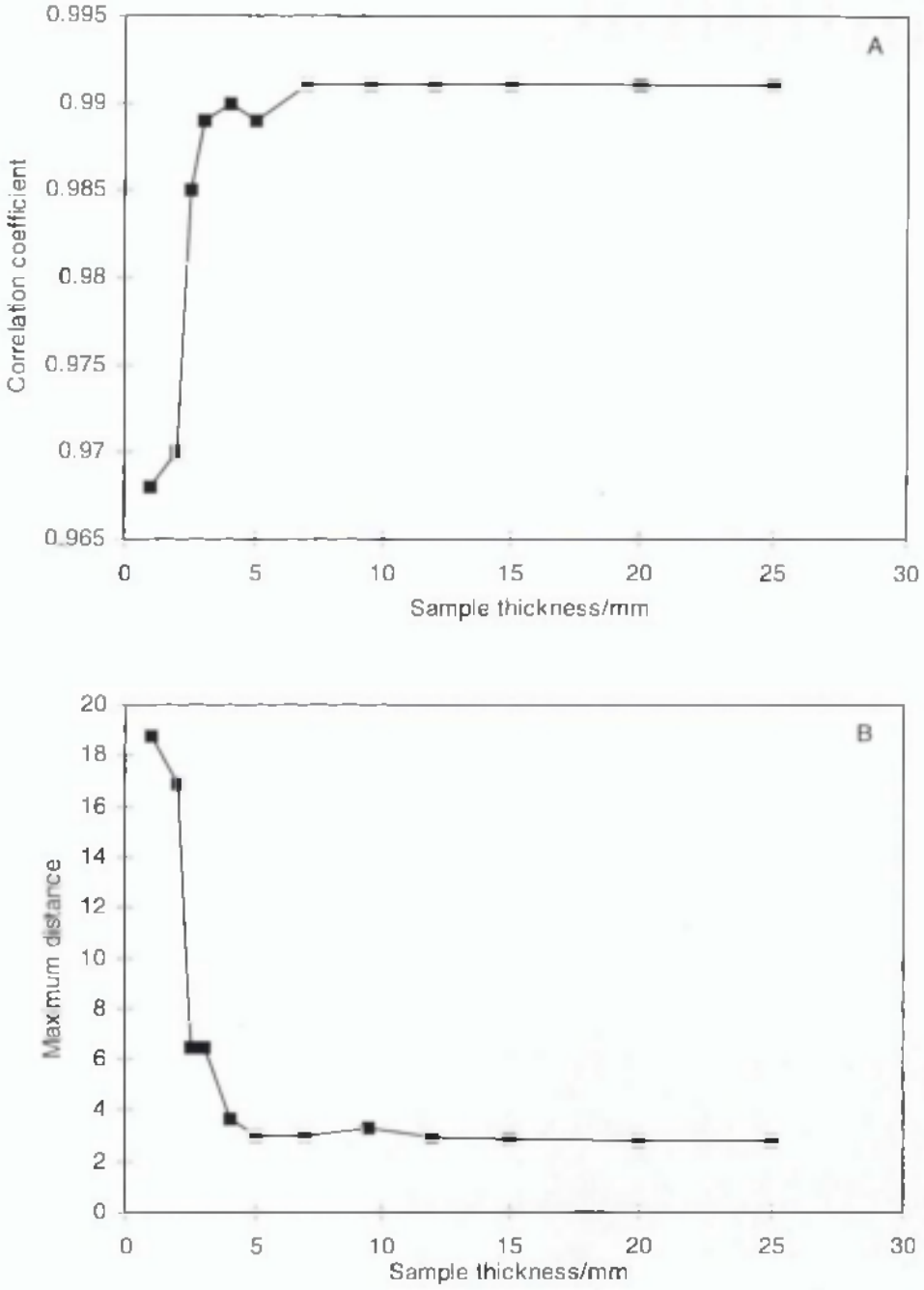


Fig. 5.7 The effect of sample thickness on (A): Correlation coefficient and (B): Maximum Distance for Kollidon 25 using absorbance spectra, with a 25 mm-thick sample used as the reference spectra.

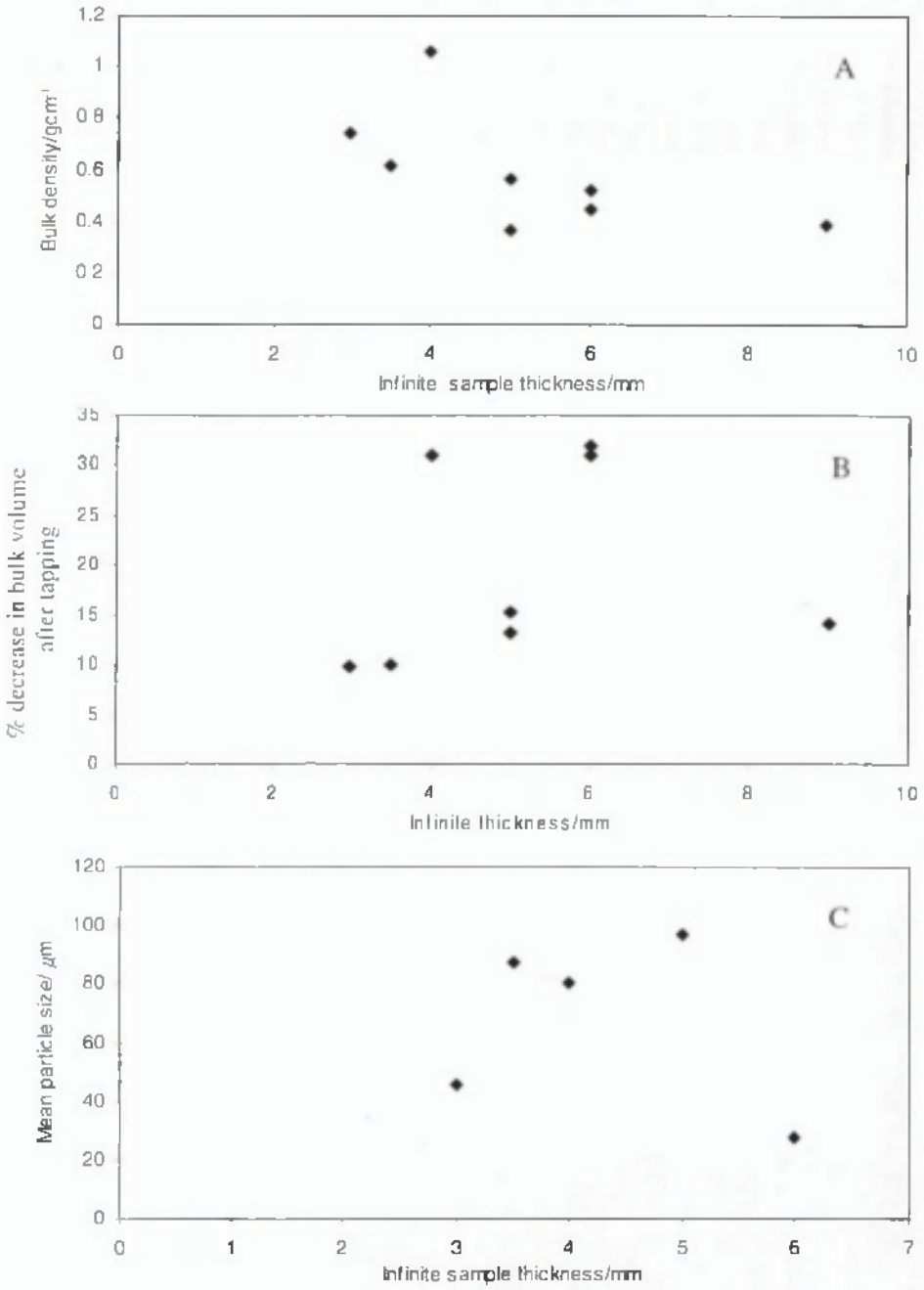


Fig 5.8 Dependence of various physical parameters on 'infinite thickness'. Correlation coefficients, r are also indicated. (A) bulk density, r = -0.611; (B) percentage decrease in bulk density after tapping, r = 0.186 and (C) mean particle size, r = -0.265.
Note: All r values suggest no evidence for correlation at 5% significance level.

the bulk volume after tapping), mean particle size have not been successful in this work (see **Table 5.3B** and **Fig. 5.8**).

Spectra measured using thicker samples improved the discrimination of closely-related excipients (**Table 5.4**). Greater differences in spectra were observed for both absorbances and second derivative absorbance spectra with increasing sample thickness. Also, better reproducibility of spectra were obtained with samples of greater than 'infinite thickness' than with thin samples because of the difficulty of uniformly filling sample cups (see **Fig. 5.9** and **Table 5.5**). Values for r_{jk} could be up to 10 times more variable when using sample thickness of 1 mm as compared to using 10 mm, as observed when comparing propyl and butyl *para* hydroxybenzoates.

5.3.3 Cup material

The ideal sample cup should not absorb near-infrared radiation, be easy to fill, disposable and cheap. Commonly used materials are quartz and various types of glass, but none of the materials currently in use fits the requirements above. Customized quartz sample cups are minimally absorptive but are hardly a cost-effective choice, particularly for large scale identification in a warehouse situation. Thus, commercially available glass vials were examined to determine their fitness for use in NIR applications.

Fig. 5.10 shows the second derivative absorbance spectra for a range of cups as measured by transflectance (i.e. by placing the ceramic reflectance reference over an empty cup). Quartz showed the least absorbance, followed by soda glass, clear neutral glass and Pyrex glass in that order. The peaks in glass at approximately 1400 nm and 2200 nm can be assigned to the O-H first overtone bands from the SiOH and also C=O forming combination bands, possibly from the carbonates of calcium and sodium.

The spectrum of the cup material was found to be additively superimposed upon the spectrum of the sample and can therefore cause serious distortion of the sample spectrum for poorly absorbing materials. **Fig. 5.11A** clearly shows this for the spectrum of A-TAB measured in quartz and Pyrex glass cups. **Table 5.6** illustrates the effect of such distortion on Correlation in Wavelength Space and Maximum

Table 5.4 Effect of sample thickness on the ability to distinguish between closely related substances using NIR absorbance spectra.

Pairs of closely related substances	Mean Correlation Coefficient*		Mean Wavelength Distance*	
	Sample thickness/mm		Sample thickness/mm	
	1	10	1	10
Avicel PH101 and Avicel PH102	1.000	1.000	0.576	9.818
Propyl and butyl *para* hydroxybenzoate	0.943	0.954	9.037	46.67
Povidone 25 and Povidone 30	0.997	0.996	3.599	7.608

* Mean spectrum of first excipient named was used as the reference spectrum

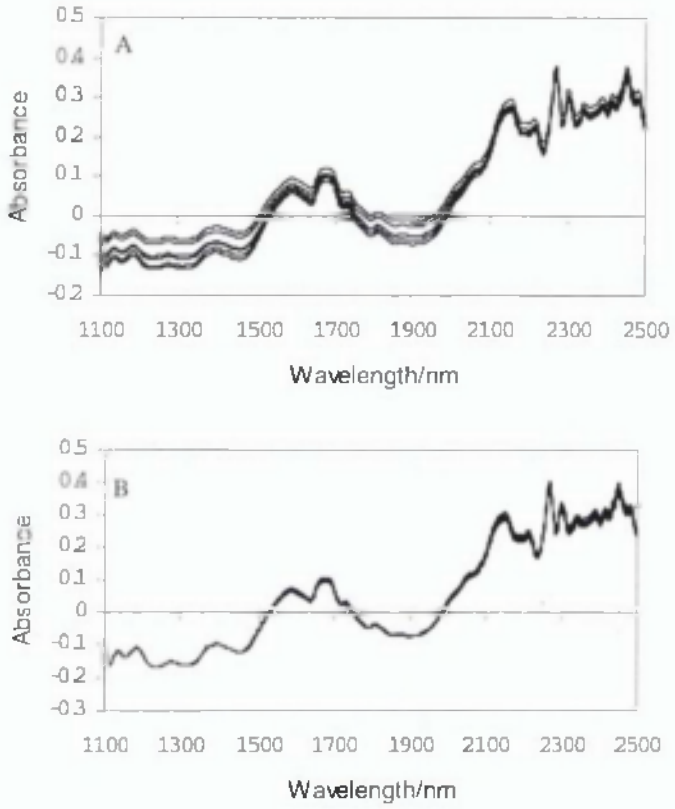


Fig. 5.9 Six repeat measurements of the spectra of propyl *para* hydroxybenzoate when sample was filled to a depth of (A) 1 mm and (B) 10 mm. Samples were refilled for each measurement.

Table 5.5 Effect of sample thickness on the reproducibility of *r* values.

Sample cup refilled for each spectrum.

Pairs of closely related substance	Correlation coefficient*	
	Sample thickness of 1 mm	Sample thickness of 10 mm
Avicel PH101and Avicel PH102		
Spectrum 1	0.999	0.998
Spectrum 2	0.994	0.999
Spectrum 3	0.999	0.999
Spectrum 4	0.991	0.998
Spectrum 5	0.998	0.999
Spectrum 6	0.998	0.998
Mean	0.9965	0.9985
Standard deviation	0.0033	0.00055
Povidone 25 and Povidone 30		
Spectrum 1	0.978	0.996
Spectrum 2	0.975	0.996
Spectrum 3	0.974	0.996
Spectrum 4	0.979	0.996
Spectrum 5	0.979	0.997
Spectrum 6	0.978	0.997
Mean	0.977	0.996
Standard deviation	0.0021	0.00052
Propyl *para* hydroxybenzoate and butyl *para* hydroxybenzoate		
Spectrum 1	0.959	0.956
Spectrum 2	0.957	0.956
Spectrum 3	0.941	0.954
Spectrum 4	0.932	0.954
Spectrum 5	0.934	0.954
Spectrum 6	0.934	0.953
Mean	0.943	0.9545
Standard deviation	0.0122	0.0012

* Mean spectrum of first named excipient was used as the reference spectrum

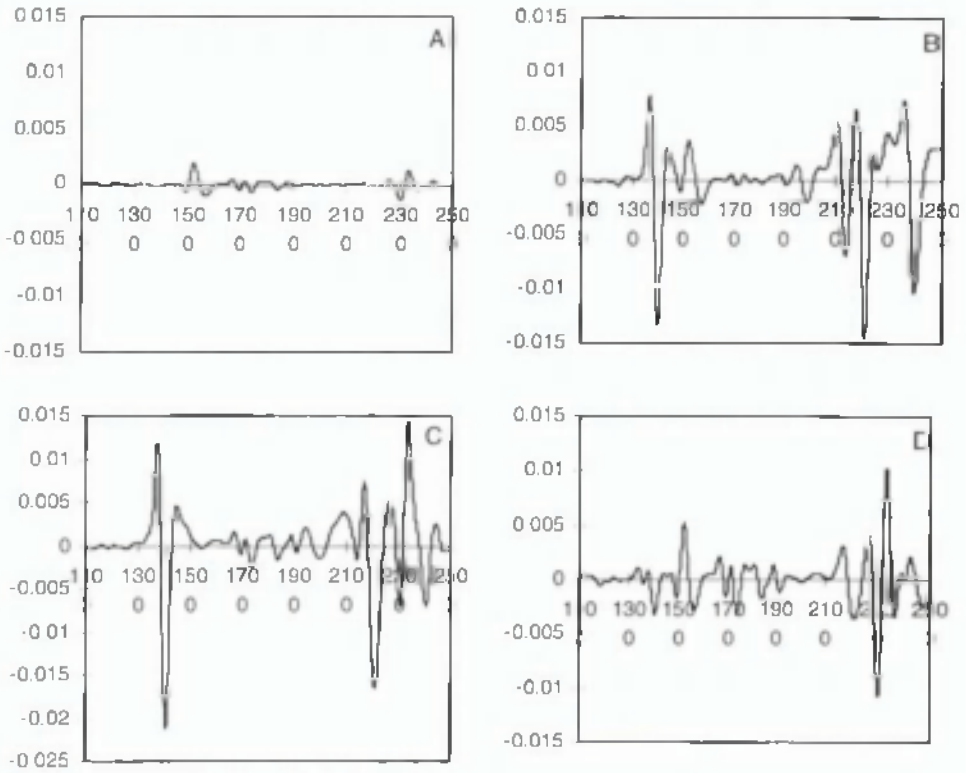


Fig. 5.10 Second derivative absorbance versus wavelength/nm.
A, quartz; B, Pyrex glass; C, clear neutral glass and D, soda glass.

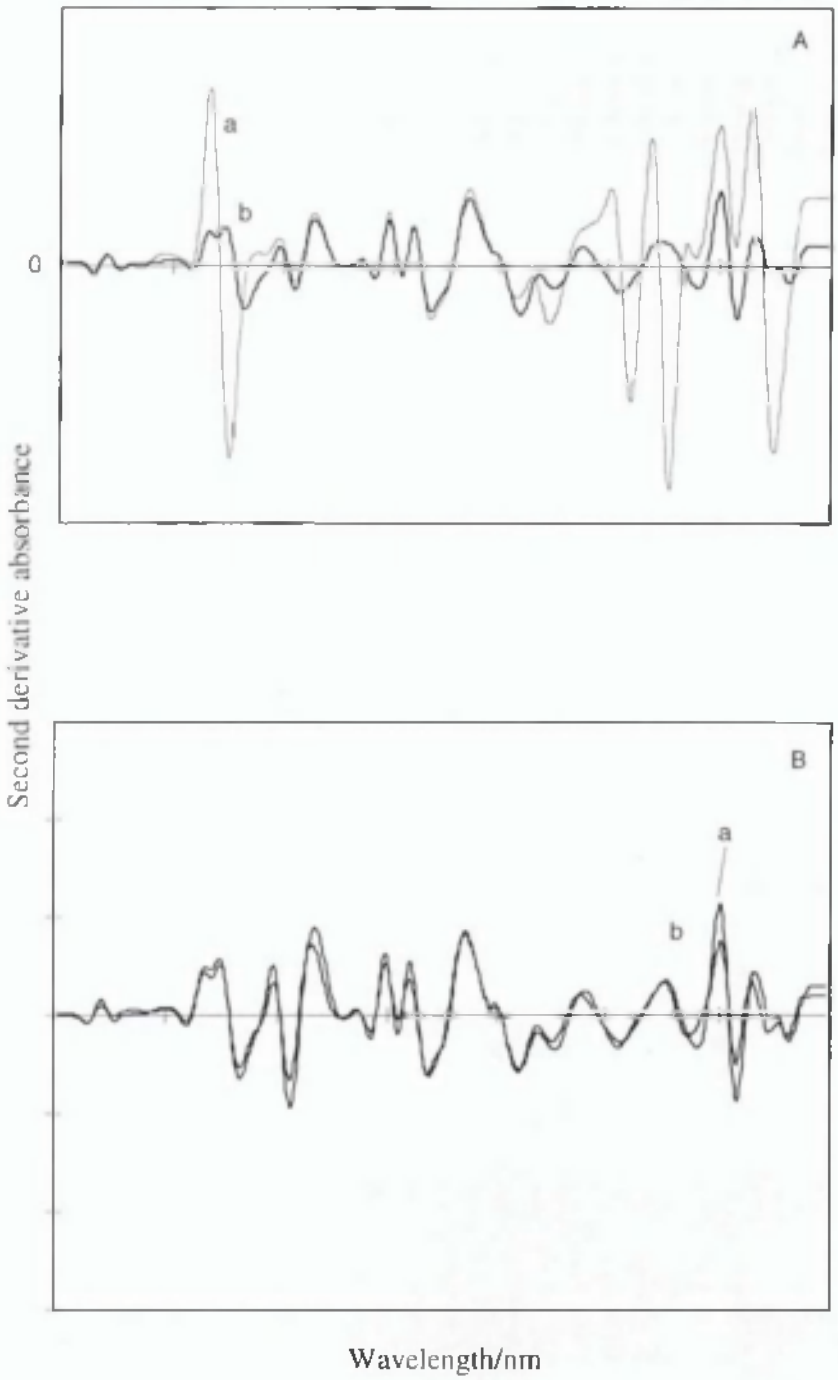


Fig. 5.11 (A): Second derivative spectra of A-TAB in a, Pyrex glass and b, quartz sample cup. (B): as in (A) but after subtraction of cup spectra.

Table 5.6 Effect of cup material on identification parameters, r_{jk} and d_{jk}. Spectra measured in the quartz cup taken as the reference. Values are mean of 6 spectra.

	A-TAB				Purified talc			
	Cup material				Cup material			
	Quartz	Pyrex glass	CNG*	Soda glass	Quartz	Pyrex glass	CNG*	Soda glass
Absorbance spectra								
r_{jk}	1	0.995	0.999	0.998	1	0.963	0.992	0.995
d_{jk}	0.870	4.881	3.236	2.037	0.963	44.637	22.251	13.058
2nd derivative absorbance spectra								
r_{jk}	1	0.416	0.836	0.939	1	0.999	1	1
d_{jk}	1.494	332.019	86.873	35.462	1.485	112.802	48.112	30.423

* Clear neutral glass.

Wavelength Distances. It is also obvious that more strongly absorbing excipients are affected to a lesser extent.

While it was possible to compensate for the cup spectrum by subtraction (**Fig. 5.11B**) the success is, however, very much dependent upon obtaining a representative spectrum for the empty cup. This is difficult as the spectrum obtained by transflectance is dependent upon the height at which the ceramic reflectance reference is placed above the cup base. Generally, this is set by the physical dimensions of the cup and it cannot be placed in direct contact with the cup base as required.

Apart from establishing the necessity to standardise on a particular cup material, the effect of cup material reproducibility could be just as crucial as it is impractical to use the same cup for all samples. The background spectra in the form of second derivative absorbance, for 6 different cups of each material used i.e. quartz, clear neutral glass, soda glass gave maximum standard deviation values of 1.4×10^{-4}, 8.92×10^{-4} and 8.48×10^{-4} respectively. These values, whilst minimal, are, however above the maximum value of the instrument noise of 4.0×10^{-5}, determining by taking 6 replicate measurements of the ceramic reference standard (see **Fig. 5.12**). Interestingly, Gemperline *et al.* (1989) obtained higher values of standard deviation, i.e. 7.52×10^{-3} and 1.76×10^{-3} for absorbance and first derivative spectra of samples measured in three different sample cups. Therefore, whilst the spectral reproducibility of glass vials are acceptable for identification purposes, depending on the accuracy of quantitative measurements required, the use of quartz cups may be necessary.

5.3.4 Packing method

Powder packing has been recognised as a source of random variation which can result in small spectral shifts (Mark and Tunnel, 1985). However, it is questionable if the packing method affects the spectra systematically. For example, tapping a powdered sample can cause stratification of the sample giving a greater density at the bottom of the sample and hence a greater reflectance. Recognising that such effects were important, three sample packing methods were examined: tapping, compression and pouring. The particles examined had a wide variety of particle shapes and sizes as can be seen from scanning electron micrographs of the materials, **Fig. 5.13**.

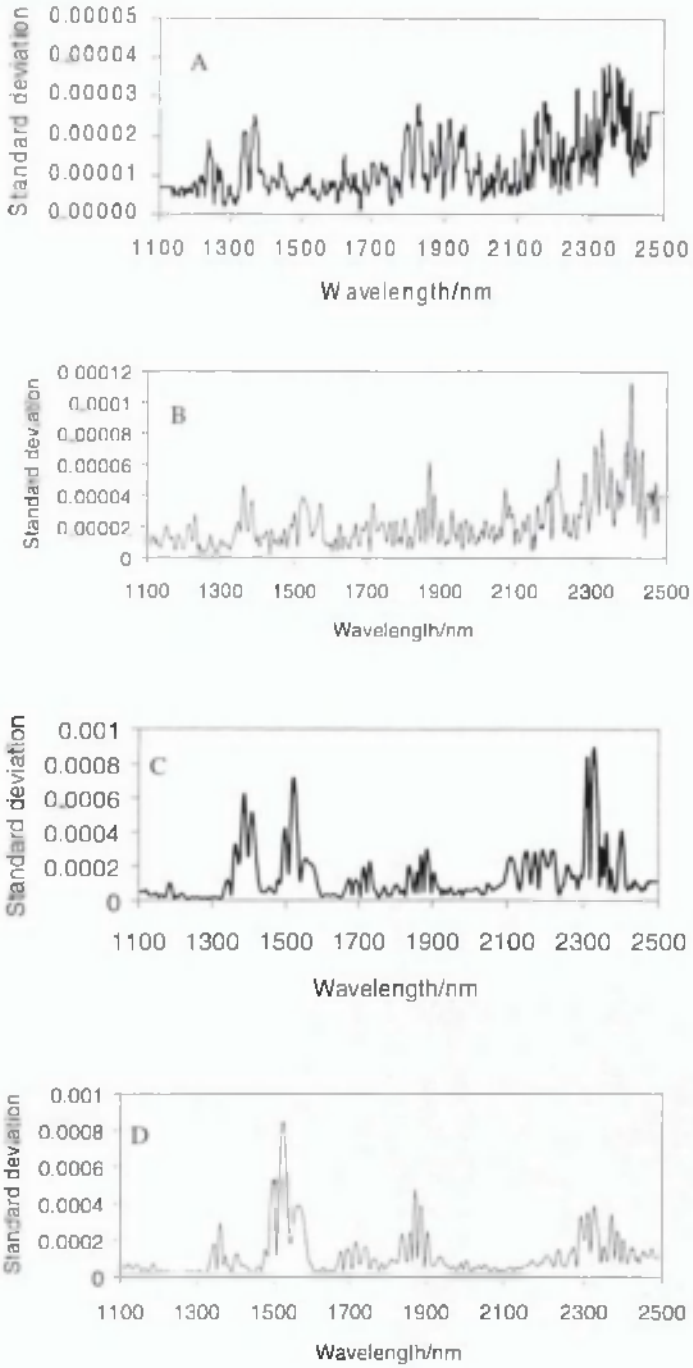
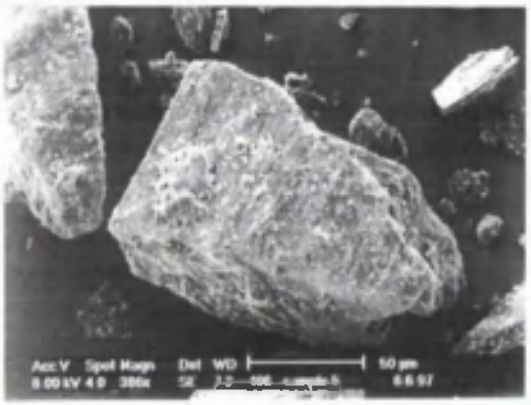
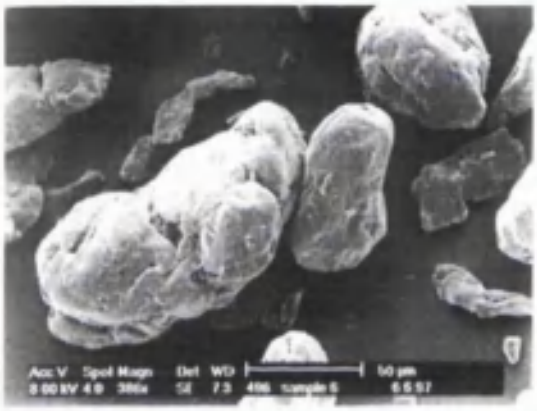


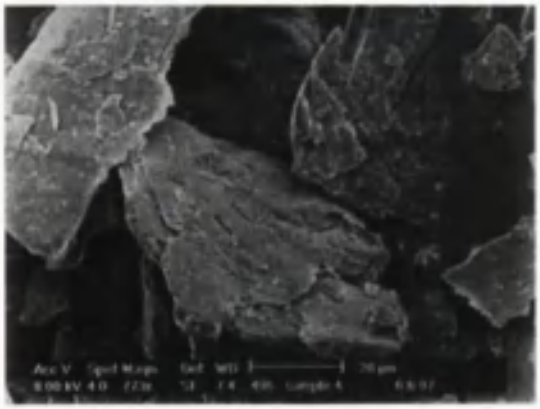
Fig. 5.12 Standard deviation *vs* wavelength for second derivative absorbance spectra measured using different sample cups. ($n = 6$) . A, noise; B, quartz; C, clear neutral glass and D, soda glass.



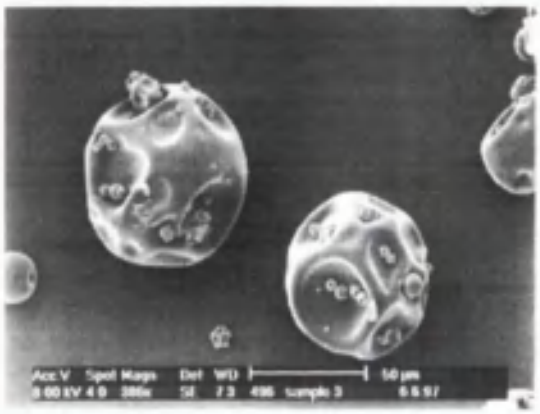
Lactose monohydrate
(Broculo Whey Products UK Ltd.)



Hydroxy propyl methyl cellulose
(Methocel E5 Premium, Colorcon Ltd.)

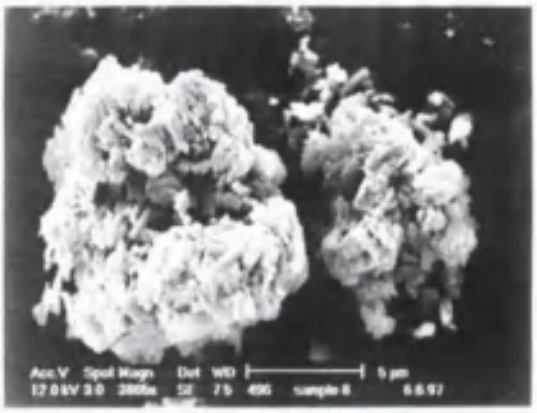


Purified talc
(Luzenac Europe)

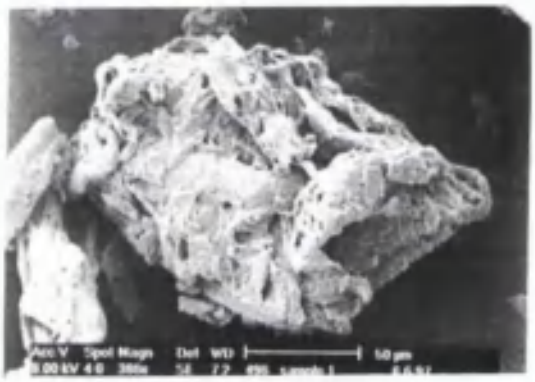


Povidone
(Kollidon 25, BASF plc)

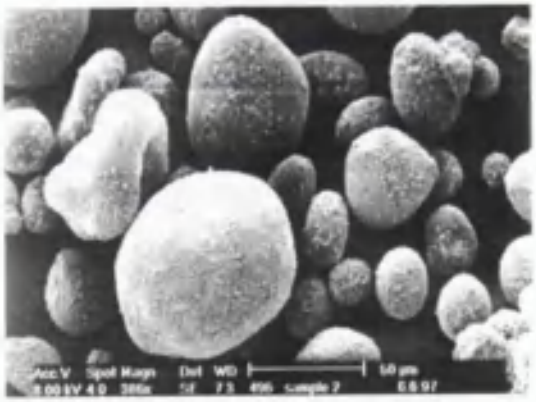
Fig. 5.13 Scanning electron micrographs of various excipients.



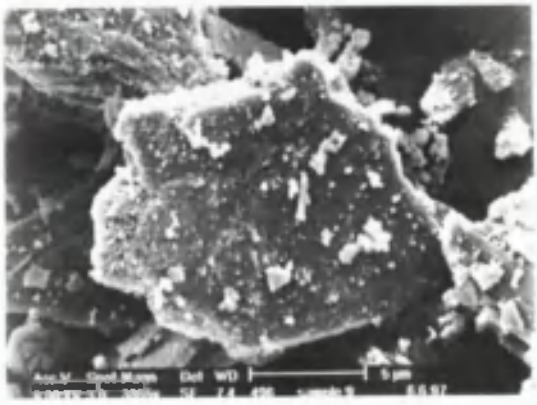
Anhydrous dibasic calcium phosphate
(A-Tab. Albright and Wilson Ltd.)



Microcrystalline cellulose
(Avicel PH102. FMC Corp.)



Sodium starch glycollate
(Explotab. Edward Mendells Corp.)



Dibasic calcium phosphate dihydrate
(Emcompress, Edward Mendells Corp.)

Fig 5.13 Continued

Tapping entailed knocking the base of the sample cup gently 10 times after filling. For compression, a 5000 Nm^{-2} pressure was applied to the powder. This was carried out by placing a 500 g weight on the back-plate of the standard sample cup. Pouring involved no treatment after filling. Each packing method was repeated 10 times and spectra were recorded. For all excipients examined, the differences between the mean spectra for the various packing procedures were only just detectable by eye. Tapped samples gave slightly stronger reflectances as compared to poured samples because of increased bulk density. Compression had no observable effect.

Correlation in Wavelength Space was calculated between the various mean spectra (absorbance, first and second derivative absorbance, standard normal variate) were not significantly differently from 1. The largest difference was observed for the first derivative spectra of purified talc. The correlation coefficient between tapped and pouring procedures was 0.994, proving that the packing method does not affect identification processes significantly.

The slight changes in absorbance values described above can, however, be important in quantitative analysis. A two sampled Students's *t*-test, **equation 5.1**, was used to compare the mean values at corresponding wavelengths between the various packing methods. An example of a *t*-plot is shown in **Fig. 5.14**.

To calculate a two sampled Student's *t* value,

$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s^2 = \frac{s_1^2\left(n_1 - 1\right) + s_2^2\left(n_2 - 1\right)}{n_1 + n_2 - 2} \qquad (5.1)$$

where s_1 is the standard deviation for the first method

s_2 is the standard deviation for the second method

$\bar{x}_1$ is the mean value for the first method

$\bar{x}_2$ is the mean values for the second method

n_1 is the number of values for the first method

n_2 is the number of values for the second method

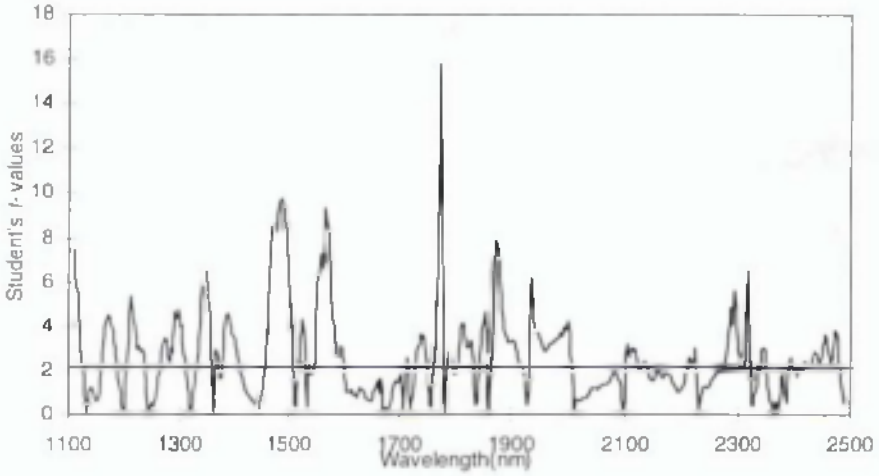


Fig. 5.14 Plot of Student's *t*-values over wavelength for the first derivative spectra of Explotab. Critical value for 18 degrees of freedom at the 5% significance level is 2.1 and shown by the horizontal line.

Significant differences in values were observed with absorbance and mathematically-treated spectra of tapped and poured samples across the majority of the wavelengths (**Table 5.7**). Purified talc showed gross differences across most of the spectrum. Purified talc has plate-like-shaped particles (see **Fig. 5.13**) and tapping can cause considerable reorientation of the particles resulting in differences in light-particle interactions.

To determine if the precision of poured samples differed from tapped and compressed samples, the two-sided F-test, **equation 5.2** was used at significance level of 5%. F values outside the range of 0.25 to 4.0 indicate that the variances are significantly different. **Table 5.8a** and **b** show that, in most cases, the precision for tapping and compression are different.

To calculate F values,

$$F = s_1^2 / s_2^2 \qquad (5.2)$$

Where s_1 is the standard deviation for the first data-set.

s_2 is the standard deviation for the second data-set.

Table 5.7 Student's t values for the two-sampled t-test for the comparison of mean values of tapped ($n=10$), poured ($n=10$) and compressed ($n=10$) samples. The table gives the maximum and minimum values of t observed across the complete wavelength range (1100 to 2498 nm). The critical value for t at 5% significance level is 2.1.

(a) Pouring and tapping

Excipients	Data treatment			
	Absorbance spectra	1st derivative absorbance	2nd derivative absorbance	Standard normal variate absorbance
Avicel PH102	0.02 - 4.87	0.67 - 10.55	0.03 - 11.43	0.03 - 12.86
Explotab	3.09 - 4.86	0 - 15.76	0 - 25.14	0.01 - 11.45
A-TAB	0 - 2.84	0.04 - 11.18	0 - 11.57	0.11 - 10.22
Emcompress	1.35 - 3.04	0.08 - 8.07	0 - 5.67	0 - 5.94
Purified Talc	0.08 - 8.26	0.6 - 327.65	0 - 282.13	0 - 100.63

(b)Pouring and compression

Excipients	Data treatment			
	Absorbance spectra	1st derivative absorbance	2nd derivative absorbance	Standard normal variate absorbance
Avicel PH102	1.07-2.29	0.07-2.73	3.16-0.0043	0.000004 -0.01
Explotab	0.0004-0.37	0-3.07	0-5.17	0.08-2.73
A-TAB	1.26-6.32	0-2.24	0-11.57	0.089-12.27
Emcompress	1.02-4.38	0.02-11.74	0-2.08	0.009-3.15
Purified Talc	0.0014-6.21	0.0159.98	0-9.54	0.004-4.32

Table 5.8 Effect of packing method on the repeatability of spectra. The table shows the maximum and minimum F values for the variance ratio test observed across the complete wavelength range (1100 to 2498 nm). Critical values for the 5% significance level are $F < 0.25$ and $F > 4.0$ (9 degrees freedom in each data-set).

Note : There is a 2.5% possibility of an F value being < 0.25 or > 4.0 There is a 5.0% possibility of an F values being outside the range 0.25 to 4.0.

(a) Pouring and tapping.

Excipients	Data treatment			
	Absorbance spectra	1st derivative absorbance	2nd derivative absorbance	Standard normal variate absorbance
Avicel PH102	0.302-29.44	0.16-17.34	0.168-16.89	0.18-22.41
Explotab	1.18-78.64	0.14-25.25	0.17-21.78	0.07-2.98
A-TAB	0.44-0.83	0.15-9.34	0.03-13.95	0.099-16.18
Emcompress	0.08-2.25	0.03-2.35	0.034-8.09	0.023-1.39
Purified Talc	0.56-12.69	0.15-50.48	0.08-27.56	0.39-13.47

(b) Pouring and compressing.

Excipients	Data treatment			
	Absorbance spectra	1st derivative absorbance	2nd derivative absorbance	Standard normal variate absorbance
Avicel PH102	0.99-1.4	0.36-5.12	0.2-9.33	0.39-2.98
Explotab	0.67-1.18	0.28-4.5	0.12-9	0.24-2.46
A-TAB	1.33-4.7	0.35-56.67	0.23-3.67	0.14-57.68
Emcompress	1.34-2.24	0.27-4.19	0.13-13.95	0.39-2.35
Purified Talc	0.74-1.14	0.35-9.89	0.19-7.19	0.54-2.37

5.4 Conclusion

The work in this chapter has clearly shown that sample presentation can have a significant effect on the near infrared reflectance spectrum of a substance. As the requirement on the accuracy and precision of a spectrum is dependent on a particular application, it is important to evaluate sample presentation effects as part of any methodology development. In all cases, cup diameter, sample thickness, cup material and packing method should be optimised and standardised. In the example used for this work, a cup diameter exceeding 20 mm should be used for the RCA. Also, powders should be filled to thickness greater than their 'infinite thickness'. For most pharmaceutical excipients, a thickness of 10 mm would provide sufficient material to ensure the reproducibility of spectra. Soda glass and clear neutral glasses absorb NIR radiation minimally, and can be used as cheaper alternatives when compared to quartz sample cups, except for applications of the most exacting nature. Tapping and compressing could sometimes cause significant ($P < 0.05$) spectral variation as when compared to simply pouring the sample into the sample cup. Therefore, multiple measurements to average out the variation would be necessary even for pharmaceutical powders. These results are in agreement with the findings reported by Mark and Tunnel (1985) for agricultural products.

CHAPTER 6: IDENTIFICATION AND QUALIFICATION OF POWDERED PHARMACEUTICAL EXCIPIENTS BY NEAR INFRA-RED SPECTROSCOPY

6.1 Introduction

The routine identification and qualification of pharmaceutical excipients is a demanding analytical task. Most excipients are powders and are also often available in various grades which differ only in terms of particle size, degree of polymerization, water content *etc.* It is therefore desirable to develop NIR methods which are sensitive to such differences, but at the same time robust enough to stand sample packing effects and batch to batch variations. Furthermore it is to be preferred if the methods can be transferred between different instruments without the need for re-calibration.

Correlation in Wavelength Space is generally not sufficiently sensitive to distinguish reliably between different grades of excipients and more complex chemometric pattern recognition methods such as SIMCA, Mahalanobis Distance and Maximum Wavelength Distance have had to be used (Gemperline *et al.*, 1989 & 1995). Unfortunately the heavy reliance on such chemometric methods has not meet the approval by some regulatory authorities (Moffat *et al.*, 1997). In addition, the Maximum Wavelength Distance method, which only requires one data-point to fall outside the threshold for a sample to be excluded, has been reported to be sensitive to slight instrumental differences (Gerhausser and Kovar, 1997).

In this chapter the possibility of using the 'Polar Qualification System' (PQS) to distinguish between different excipients and grades of excipients is investigated. As described in chapter 1, PQS is an algorithm which can be easily visualised and hopefully would be more acceptable to the regulatory authorities. The transferability of PQS between different instruments will also be studied. Finally the problems of manufacturers using different reference reflectance standards and the effects of stray radiation on spectra will be looked into. Suggestions for their standardisation to aid transferability of spectra will also be considered.

6.2 Experimental

6.2.1 Materials

All excipients used were of pharmaceutical grade and had been shown to meet the British Pharmacopoeia and/or United States Pharmacopeia specifications. The excipients investigated were: microcrystalline cellulose (Avicel PH102, Honeywill & Stein, Surrey, UK), dicalcium phosphate dihydrate (Emcompress, Edward Mendell Co Inc, Surrey,UK), sodium starch glycollate (Explotab, Edward Mendell Co Inc, Surrey, UK), regular lactose monohydrate and anhydrous lactose (DMV International, Veghel, Netherlands), lactose monohydrate, *Tablettose* (Meggle GmbH, Wasserburg, Germany), hydroxypropyl methylcellulose (Methocel, Colorcon, Orpington, UK), Opadry White YS-1-7003 (Colorcon, Orpington, UK), povidone (Kollidon 30, BASF Plc, Cheshire, UK), methyl-, propyl- and butyl- *para* hydroxybenzoate (Nipa Laboratories Ltd, Mid Glamorgan, UK), magnesium stearate (Akros Chemicals, Manchester, UK) and talc (Luzenac Europe, Toulouse, France).

For the stray radiation investigations, polyvinyl chloride (low *Mr* approx. 100 000, particle size - 100% passes BS 60 mesh, 74% passes BS 200 mesh; BDH, Poole, UK), barium sulphate (precipitated 99 %, Avocado Research Chemicals Ltd, Hyesham, UK) and carbon black were used. Carbon black was prepared by allowing a luminous flame from a Bunsen burner to impinge upon a water cooled glass surface. The carbon black deposit was scraped off and collected.

6.2.2 Apparatus

Three different instrumental setups were used. Instruments I and II were based on the FOSS NIRSystems (Silver Springs, MD, USA), using the 6500 spectrophotometer fitted with a Rapid Content Analyzer (RCA). Instrument III was a Bran+Luebbe Infraprover II Fourier Transform Polarisation Spectrophotometer (GmbH, Norderstedt, Germany) fitted with a Sample Presentation Accessory (SPA).

6.2.3 Measurement of spectra

Samples of excipients were measured in clear neutral glass vials (C/N : BDH/Merck/215/0074/23, 25 mm in diameter) to a depth of about 10 mm. Twelve replicates of each sample were measured to average out spectral variation arising from sample packing.

Stray radiation investigations were carried out using a number of different types of sample cups: clear neutral glass (as above), brown glass (diameter of 11 mm), Waters 1 ml vials (C/N: Waters/WAT025051, 6 mm in diameter) and Waters 4 ml (C/N: Waters/WAT025054, 12.5 mm in diameter) vials were used.

Each recorded spectrum measured on the FOSS NIRSystems instruments was an average of 32 scans, measured over the range 1100 to 2498 nm (2 nm intervals, 700 data-points). Spectra were measured relative to a ceramic reference standard.

For the Bran+Luebbe instrument, spectra were measured over the range 4000 to 9996 cm^{-1} (12 cm^{-1} intervals, 500 datapoints) and each recorded spectrum was an average of 6 scans. Spectra were measured relative to a Spectralon® standard.

6.2.4 Data treatment

Derivative spectra were calculated by simple difference procedures using program DERIV as described in chapter 3. A block size of 10 and gap size of 0 were used. Normalisation of the maximum absorbance or absolute second derivative peak to 1 according to **equation 1.17** and normalisation by standard normal variate **(equation 1.18)** were performed using program DNA.

Cubic spline interpolation was used to convert spectra recorded in wavenumber to spectra with equally spaced wavelength data-points using program CUBIC as described in chapter 3.

Program DNA was used for the transformation of spectra to polar coordinates and the calculation of centre of gravity points.

6.3 Results and Discussion

6.3.1 Identification of excipients

The feasibility of NIR spectroscopy to differentiate between eleven commonly used excipients was initially investigated. One batch of each excipient was measured twelve times using instrument I. The sample bottle was shaken between each measurement to see the effects of re-packing on the spectra. Original spectra for the excipients are shown in **Fig. 6.1**. The effects of re-pack are clearly seen as shifts in the baseline between successive scans. Taking the second derivative and normalisation is particularly good at removing this effect, **Fig. 6.2**.

The ability of Correlation in Wavelength Space to differentiate between these excipients is not particularly good. For example, the dot product Correlation Coefficient between pre-gelatinised starch and sodium starch glycollate (Explotab) is 0.972 for their second-derivative spectra over the wavelength range 1100 - 2498 nm. Similarly comparing Opadry White YS-1-7003 (a proprietary mixture of hydroxymethyl propyl cellulose, pigment and plasticizer) to Methocel (hydroxy propyl methylcellulose) was 0.981. The full correlation matrix is shown in **Table 6.1**, clearly there are values above the critical value of 0.95 commonly used to denote positive identification by correlation.

By comparison the PQS method gave good separation between the majority of the excipients when using second-derivative absorbance spectra, **Fig. 6.3A**. The centre of gravity points for each excipient can be clearly differentiated from one another except for the three celluloses (Avicel, Methocel and Opadry White YS-1-7003). The ellipses shown are 'equal frequency ellipses' and represent the area within which 95% of the population of measurements for each excipient would be expected to fall. Powder re-pack caused considerable spectral variation for the celluloses resulting in overlap between their equal frequency ellipses. Normalisation to 1 removes much of this

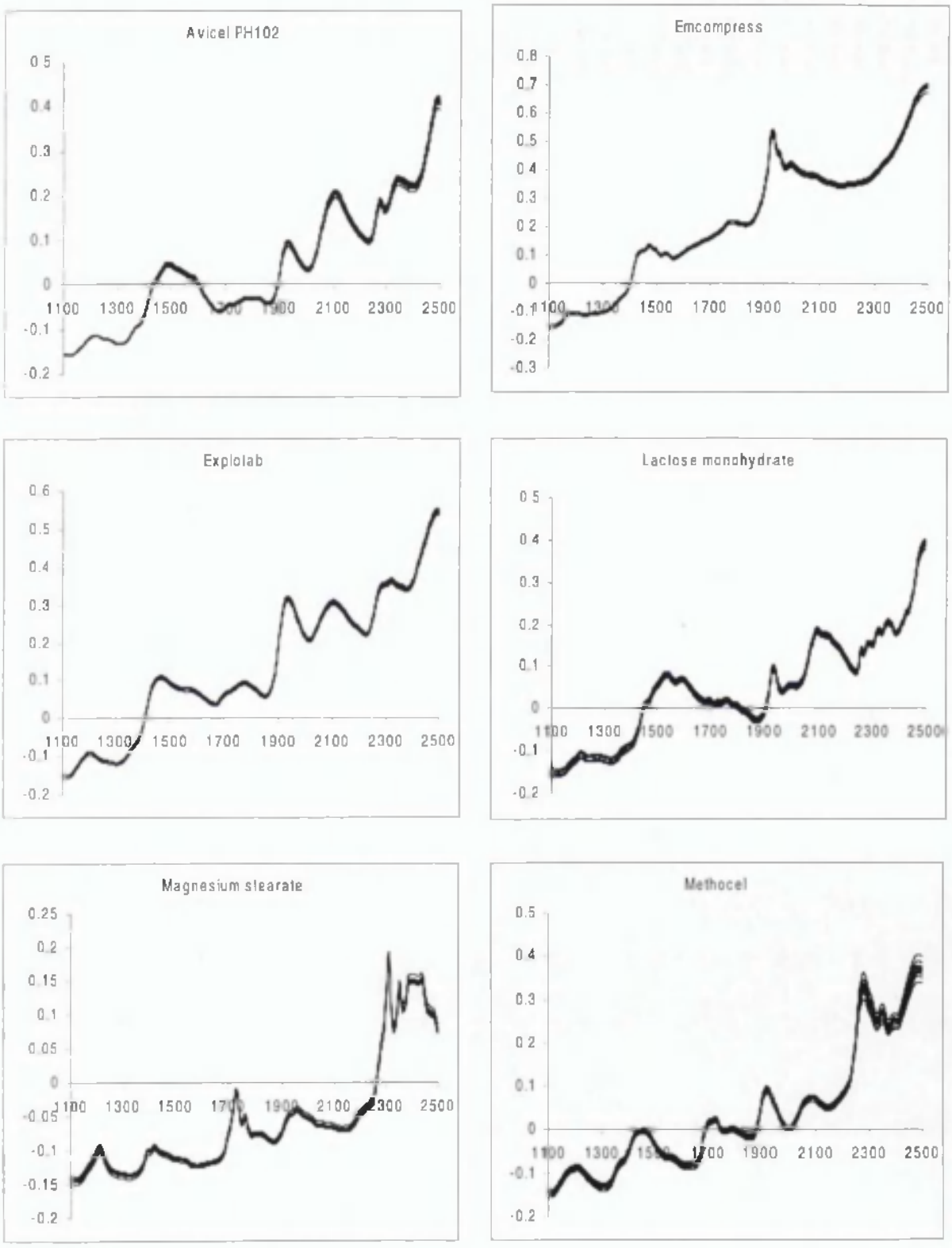


Fig. 6.1 Absorbance *vs* wavelength/nm.
Spectra (12 replicates) of various powdered excipients.

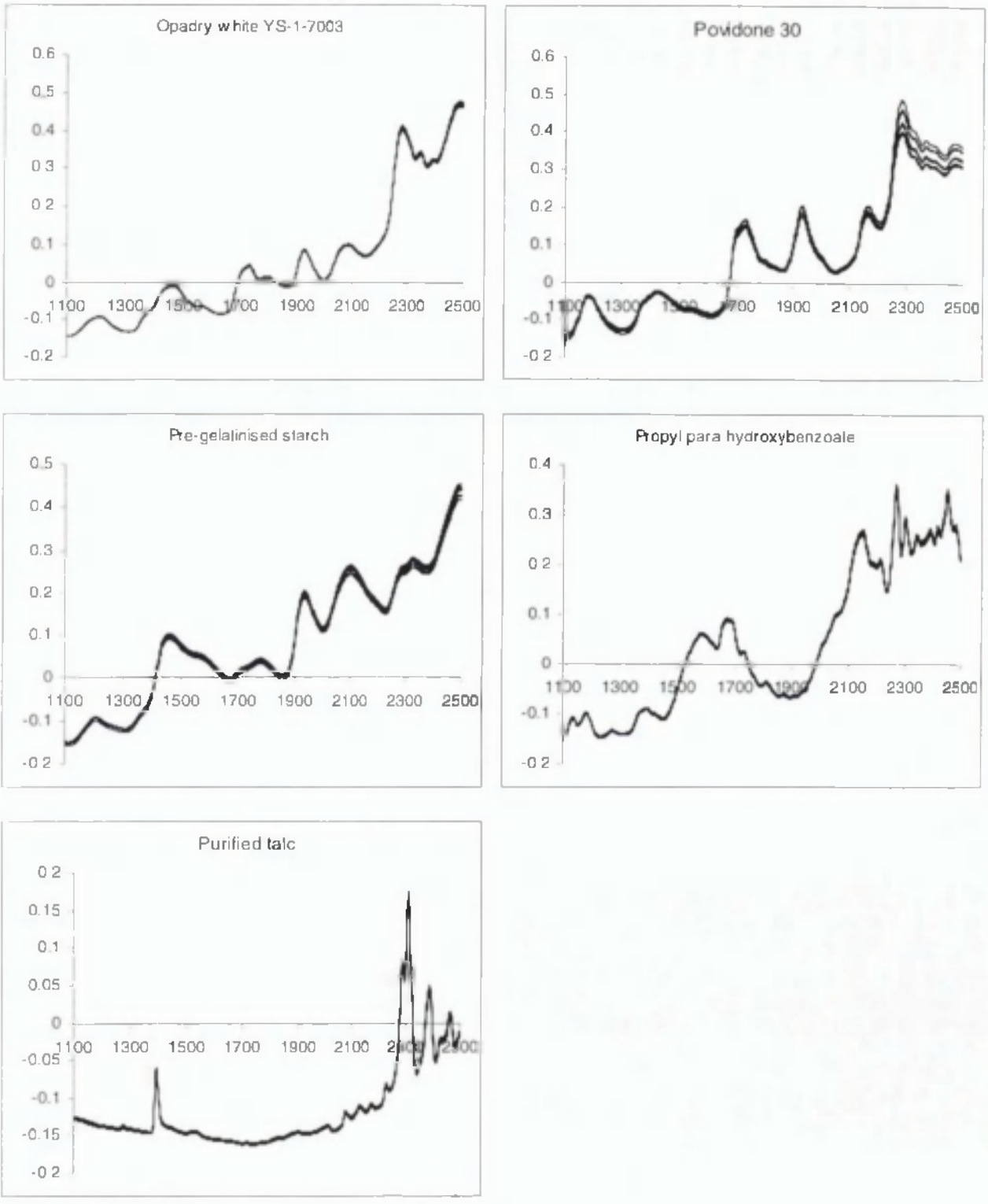


Fig.6.1 Continued. Absorbance *vs* wavelength/nm.

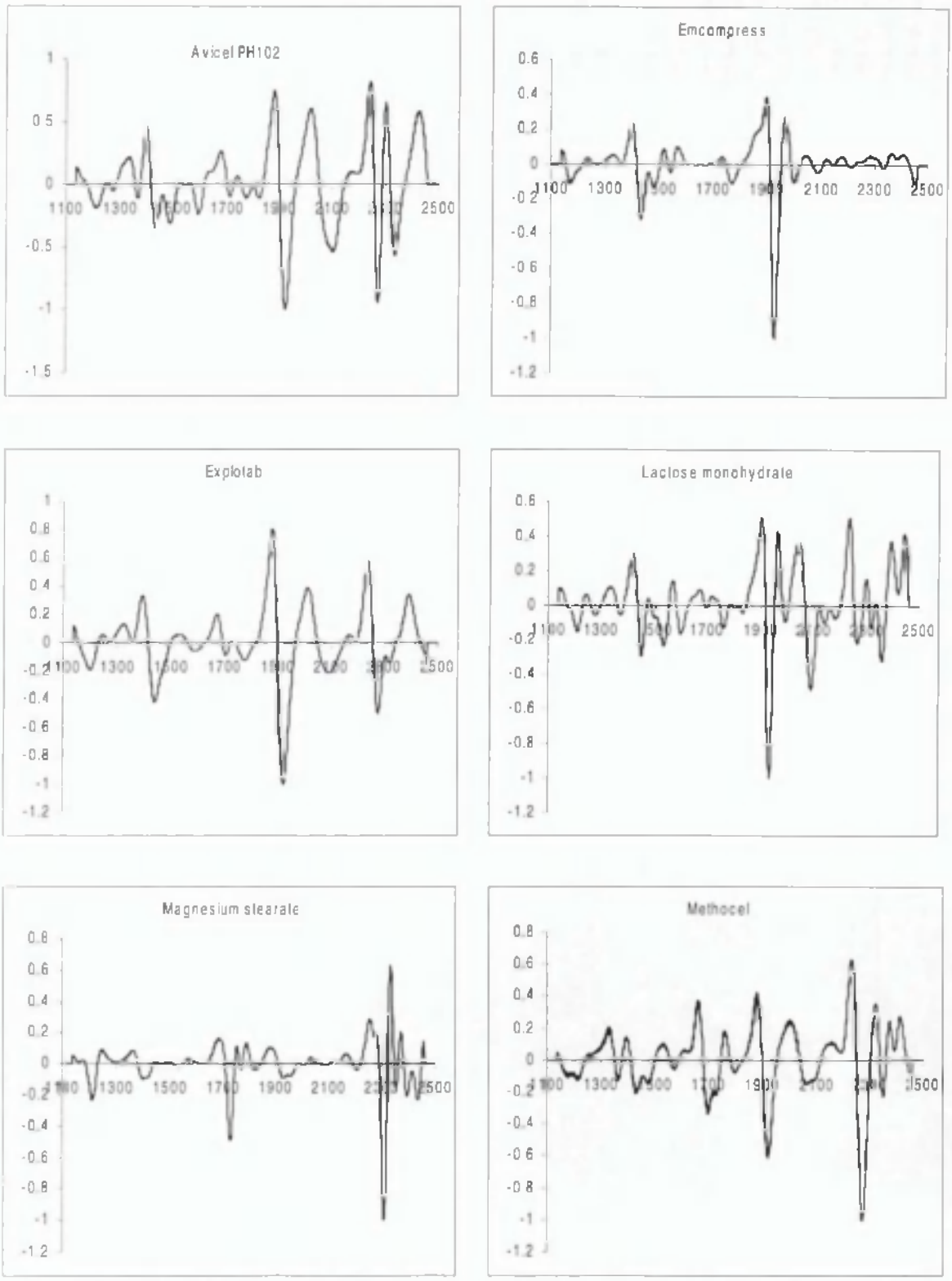


Fig. 6.2 Second derivative (normalised to 1) absorbance *vs* wavelength/nm. Spectra (12 replicates) for various powdered excipients.

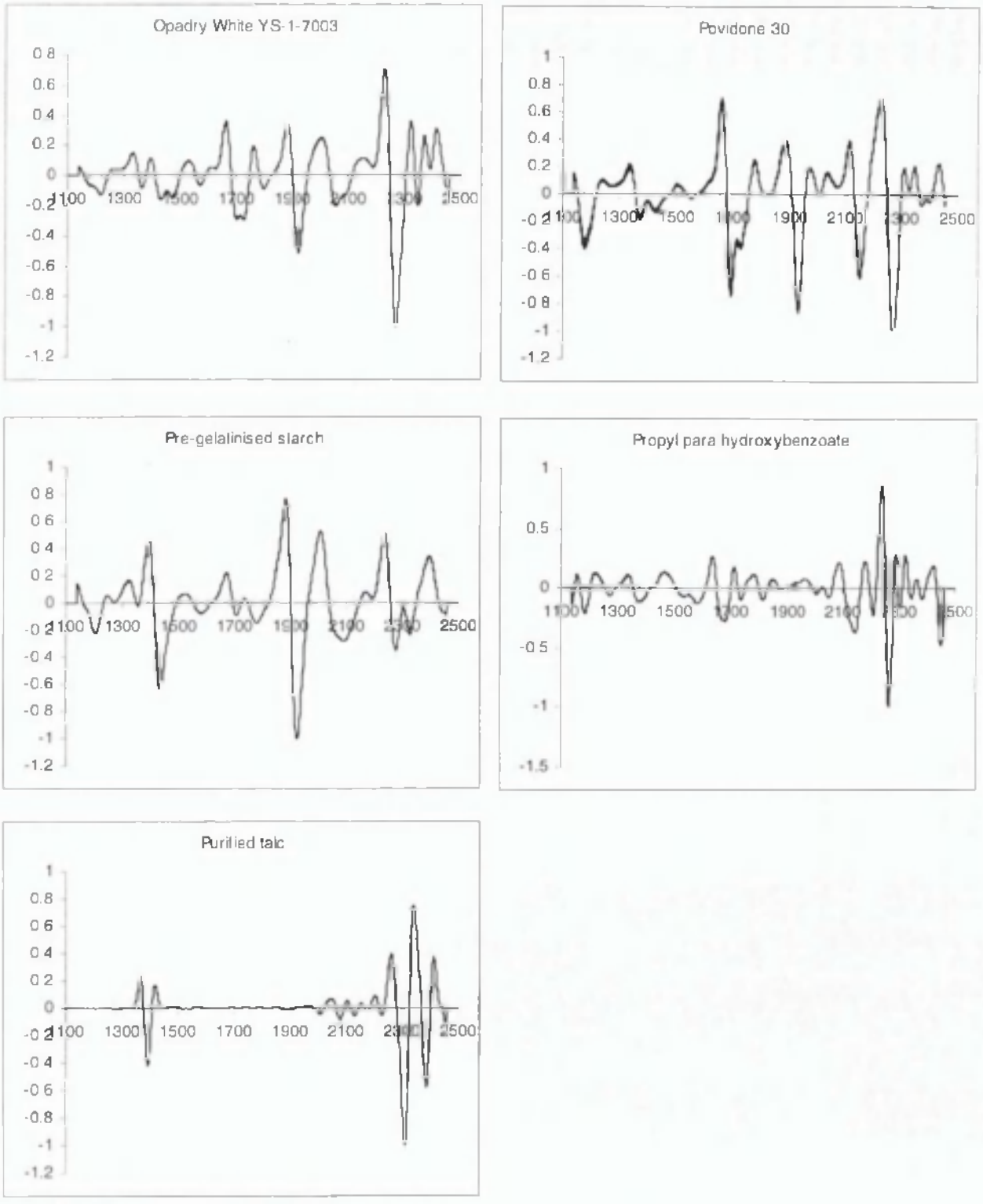


Fig. 6.2 Continued. Second derivative (normalised to 1) absorbance *vs.* wavelength.

Table 6.1 Correlation coefficients for various excipients measured using instrument 1. Second derivative spectra (10 data-point block size). Wavelength range 1136 – 2462 nm.

	Avicel PH102	Emcompress	Explotab	Lactose monohydrate	Magnesium stearate	Methocel	Opadry	Povidone 30	Pre-gelatinised starch	Propyl parabens	Talc
Avicel PH102	1.000	0.508	0.824	0.628	-0.176	0.691	0.654	0.454	0.825	0.219	-0.233
Emcompress		1.000	0.663	0.606	0.036	0.363	0.321	0.381	0.674	-0.020	-0.037
Explotab			1.000	0.553	0.099	0.720	0.696	0.551	0.972*	0.133	-0.026
Lactose monohydrate				1.000	-0.022	0.387	0.390	0.434	0.538	0.102	-0.098
Magnesium stearate					1.000	-0.089	0.045	0.001	0.069	-0.054	0.529
Methocel						1.000	0.981*	0.751	0.678	0.390	-0.202
Opadry							1.000	0.745	0.649	0.395	-0.108
Povidone 30								1.000	0.471	0.326	-0.073
Pre-gelatinised starch									1.000	0.072	-0.049
Propyl parabens										1.000	-0.038
Talc											1.000

* Correlation values > 0.95 for different excipients

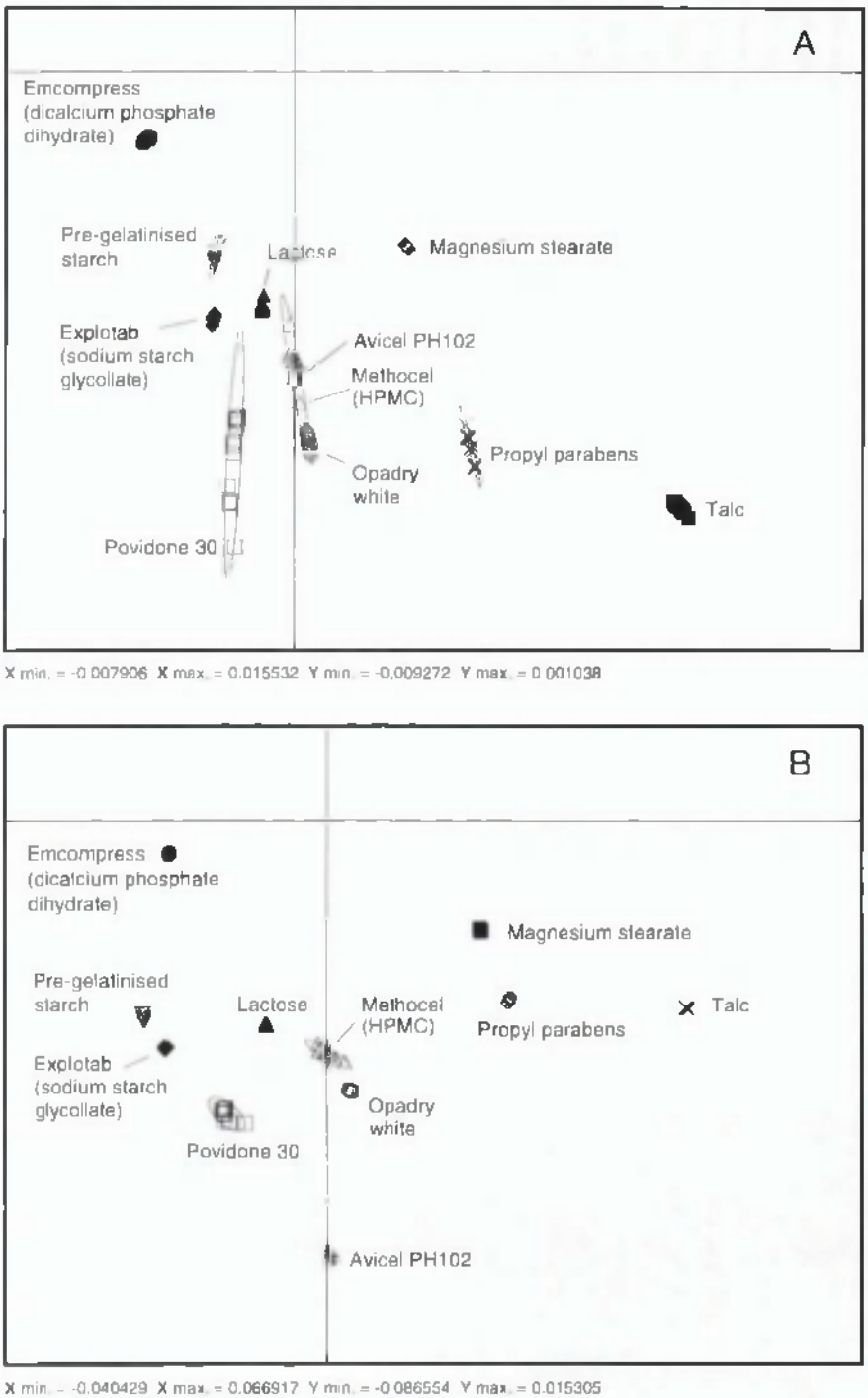


Fig 6.3 Centre of gravity plots for eleven different excipients measured using instrument 1. Spectral range 1100 - 2498 nm. (A) Second-derivative absorbance spectra, and (B) normalised second-derivative spectra.

variation and in **Fig. 6.3B** it can be seen that all the excipients can be easily distinguished from one another

6.3.2 Differentiation between subgroups of excipients

In this section, the ability of PQS to distinguish between the variants within an excipient group was examined. The examples chosen were three *para* hydroxbenzoates and two grades of lactose, for which a large number of different batches of materials were available.

Five different batches of methyl and propyl *para* hydroxybenzoates and six different batches of butyl *para* hydroxybenzoate were used. Twelve spectra for each batch were measured using instrument I with the sample bottle being shaken between each measurement to allow for the effects of re-packing. These three compounds have fairly similar NIR spectra, **Fig. 6.4**. Centre of gravity plots are shown in **Fig. 6.5** for three different data pre-treatments: second-derivative absorbance, second-derivative absorbance normalised to 1 and second-derivative absorbance normalised using SNV. An equal frequency ellipse is shown for the measurements of each different batch of material. The first two data pre-treatments were successful in separating the groups of compounds; normalisation helping to make the ellipses tighter by removing some of the effects of re-packing. SNV normalisation resulted in a slight overlap of the equal frequency ellipses for propyl and butyl *para* hydroxybenzoates. Improved separation using PQS can be obtained by careful selection of the spectral range used. **Fig. 6.6**, shows some examples of the effects of using restricted spectral ranges.

All three data pre-treatments used above were successful in separating the two grades of lactose investigated. Four batches of lactose monohydrate (*Tablettose*, contains approx. 5.0% water) and six batches of anhydrous lactose ($\leq$ 1.0% water) were measured twelve times each as in the case of the *para* hydroxybenzoates. Because of the different water contents these two grades of lactose are fairly easily distinguished by visual inspection of their NIR spectra, **Fig. 6.7**. The centre of gravity plots, **Fig. 6.8**, show that though the two types of lactose are well separated, there is considerable

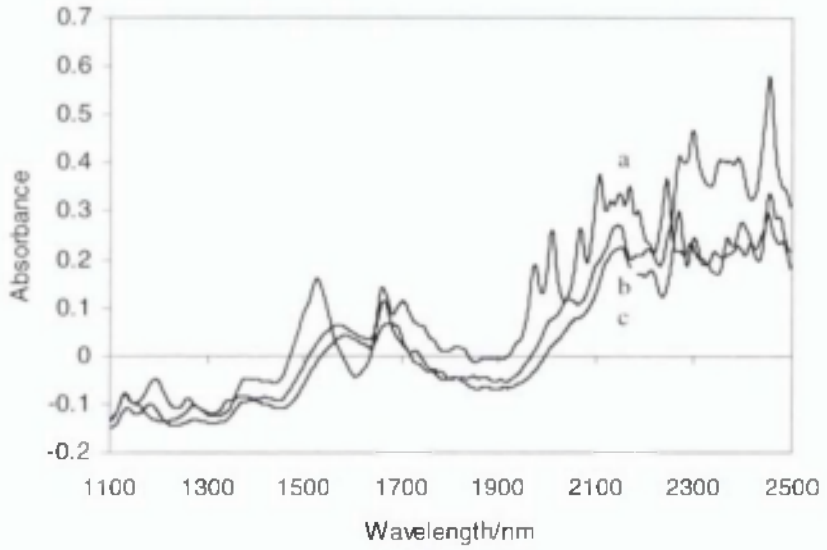


Fig. 6.4 Absorbance spectra of different types of *para* hydroxybenzoates measured on Instrument L a, butyl *para* hydroxybenzoate; b, methyl *para* hydroxybenzoate and c, propyl *para* hydroxybenzoate

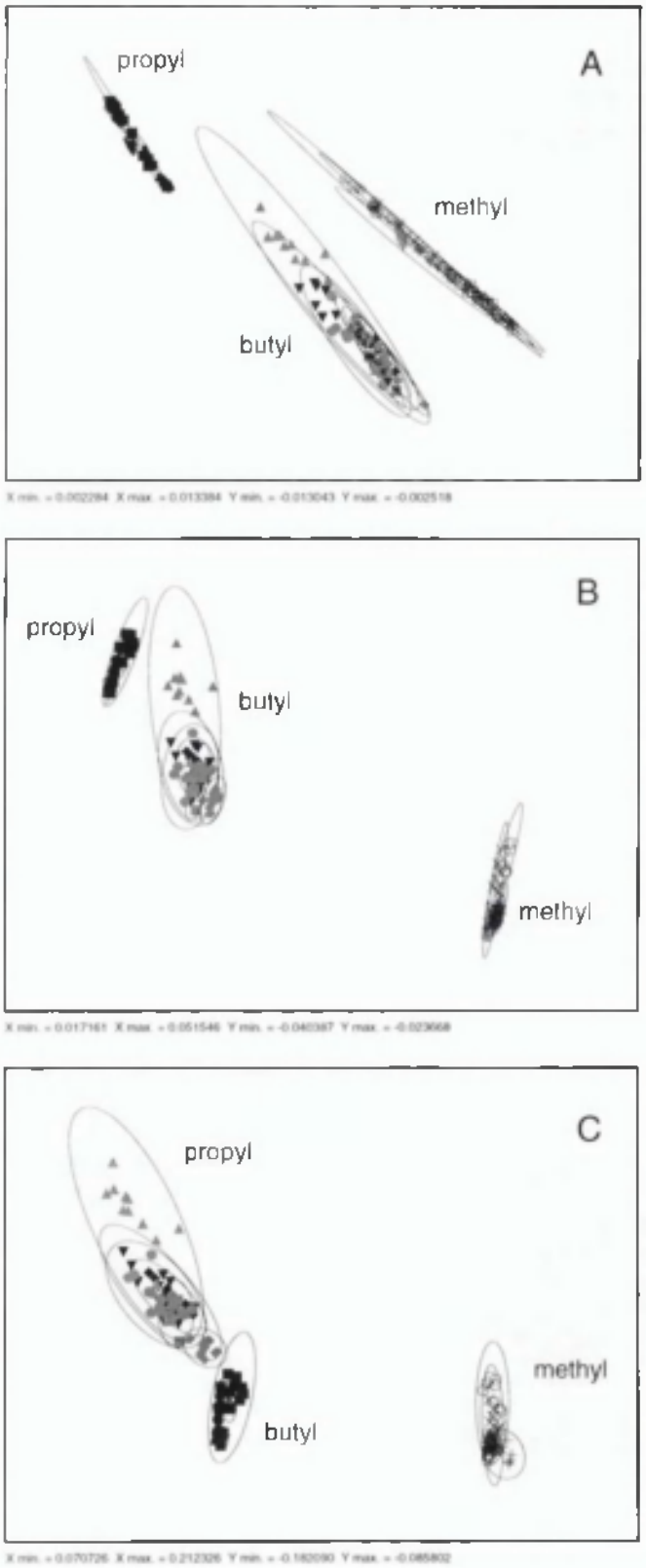
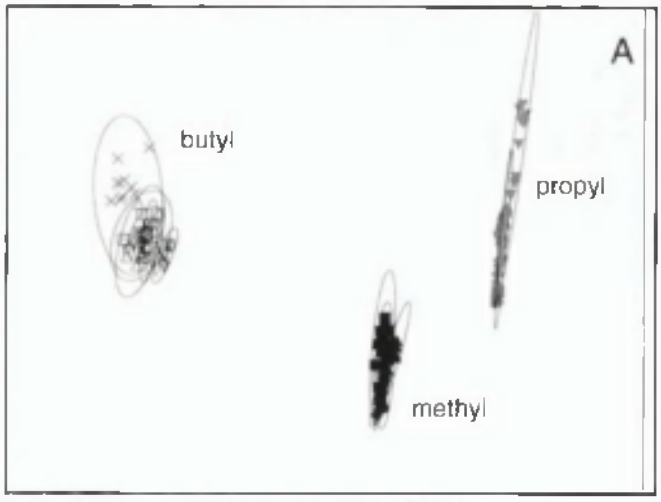
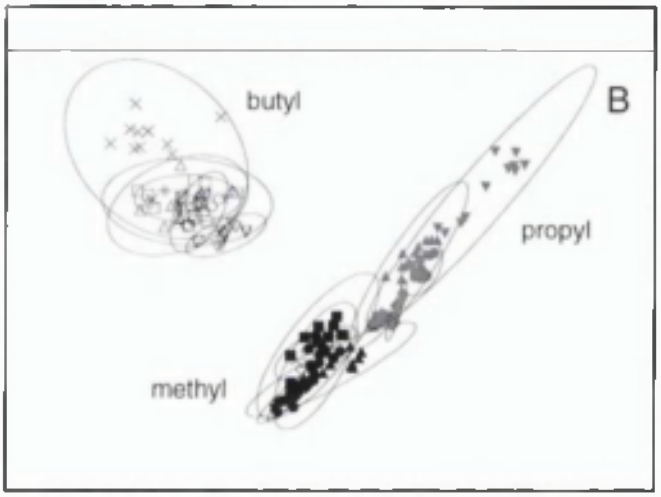


Fig. 6.5 Centre of gravity plots for methyl, propyl and butyl *para* hydroxybenzoates measured using instrument I. Spectral range 1100 - 2498 nm. (A) second derivative absorbance, (B) second derivative normalised, and (C) second derivative SNV normalised.





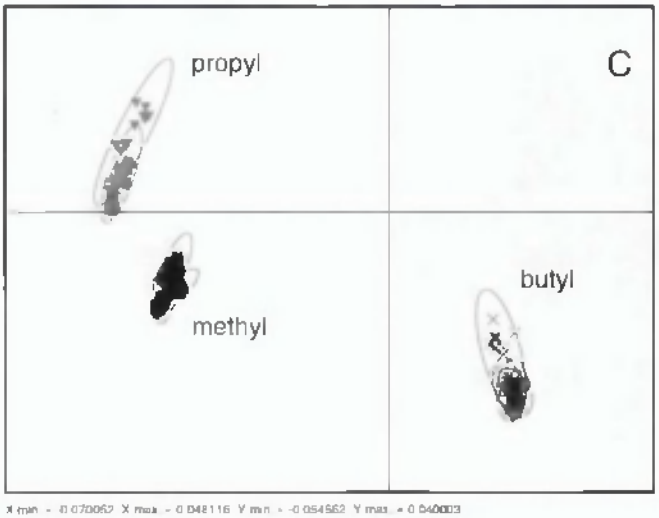


Fig. 6.6 Effect of spectral range on the centre of gravity plots for methyl, propyl and butyl *para* hydroxybenzoates. Second derivative SNV normalised spectra. (A) 1500 - 1900 nm, (B) 1500 - 2000 nm, and (C) 1500 - 2100 nm.

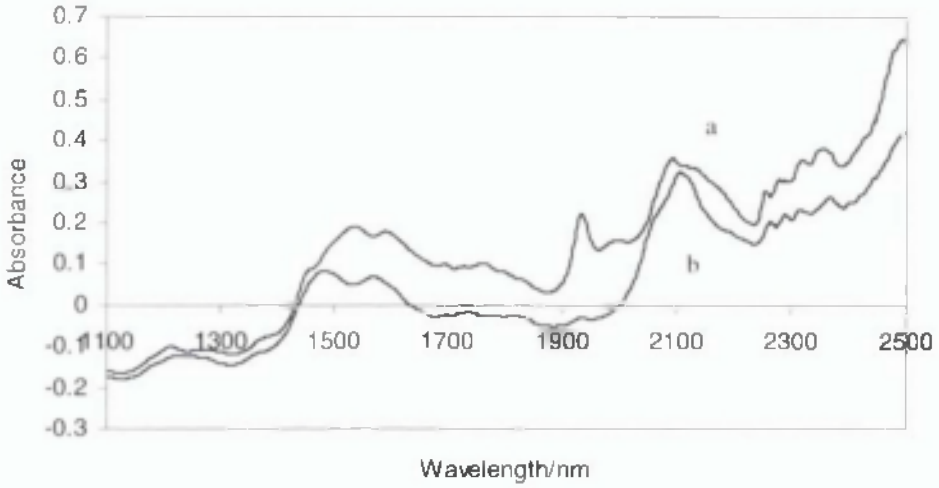


Fig. 6.7 Absorbance spectra of two different types of lactose
a. Anhydrous lactose (DMV International)
b. Lactose monohydrate (*Tablettose*, Meggle, GmbH)

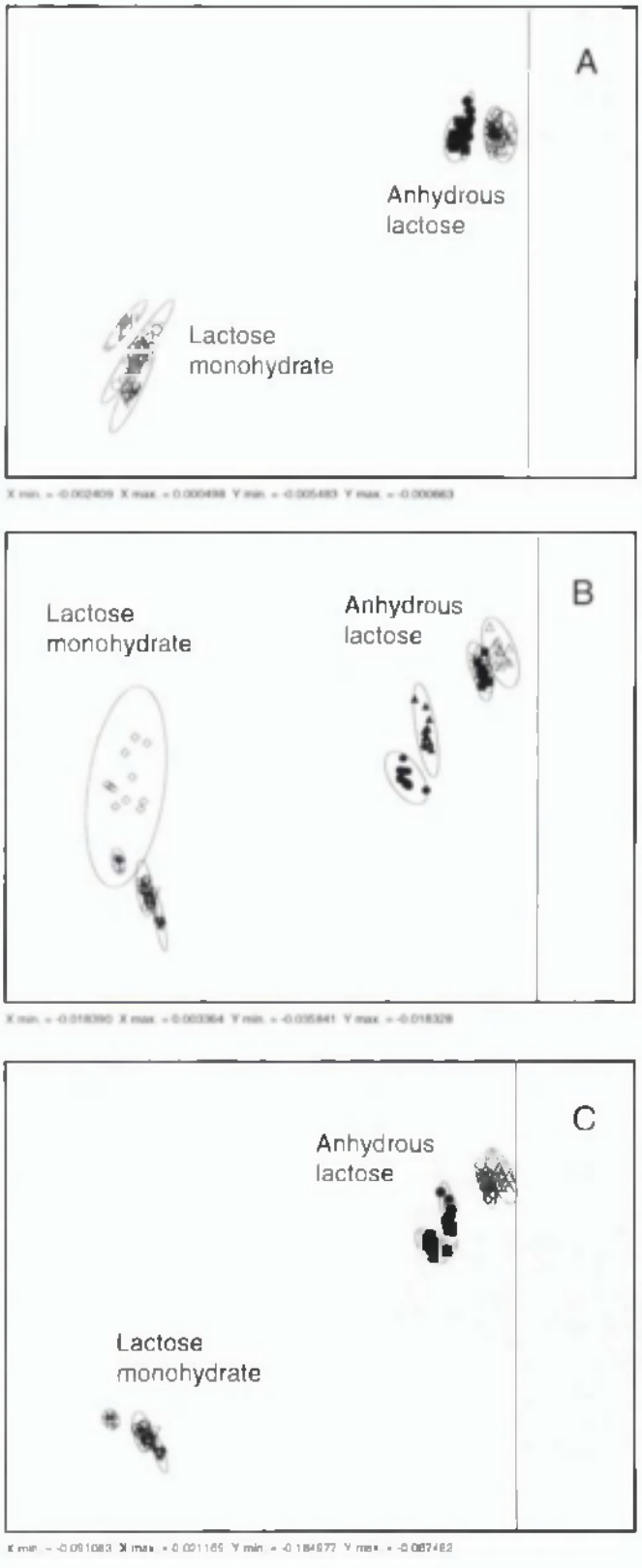


Fig. 6.8 Centre of gravity plots for anhydrous lactose and lactose monohydrate (*Tablettose*) measured on instrument 1. Full spectral range (1100 - 2498 nm). (A) second derivative absorbance spectra, (B) second derivative normalised spectra, and (C) second derivative SNV normalised spectra.

inter-batch variation with the equal frequency ellipses for certain batches of the same material were overlapping.

6.3.3 Transferability between different instruments

The effects of instrumental differences on the ability to identify the three different *para* hydroxybenzoates was studied by measuring all the batches on two further NIR instruments.

Unfortunately the spectra recorded on instrument II were distorted below about 1500 nm due to failure to correctly set the gain control, **Fig. 6.9**. Both instruments I and II used FOSS 6500 monochromators fitted with a RCA, however, one was fitted with an auto-gain amplifier and the other a manual-gain amplifier. **Fig. 6.10** shows the centre of gravity plots for the three *para* hydroxybenzoates as measured on the three instruments for a number of different spectral ranges. For propyl and butyl *para* hydroxybenzoates the centre of gravity points from instrument III were slightly separated from those using instruments I and II. Considering that the spectra from instrument III were originally measured on a linear wavenumber scale and had therefore required cubic spline interpolation to produce data points linear in wavelength, the agreement between the three instruments is remarkably good.

6.3.4 Effects of reference standard and stray radiation on spectra

Many possible causes of the apparent non-transferability of NIR spectral data are quoted in the literature, though often without quantitative justification. The commonly stated causes are: wavelength accuracy, band pass, photometric linearity, sample presentation, reflectance standard used, stray radiation, *etc.* (Mark and Workman, 1988; Miller, 1993; Knee and Deadman, 1998; Springsteen and Ricker, 1996). Modern instruments, however, have very tight specifications for many of these parameters and it is doubtful if many of them are really important. In the study for the transferability for a spectral library of common solvents (chapter 3) small changes in wavelength and band-pass within the manufactures tolerance were found to be of no consequence. Reference standards and stray radiation are two factors which are more

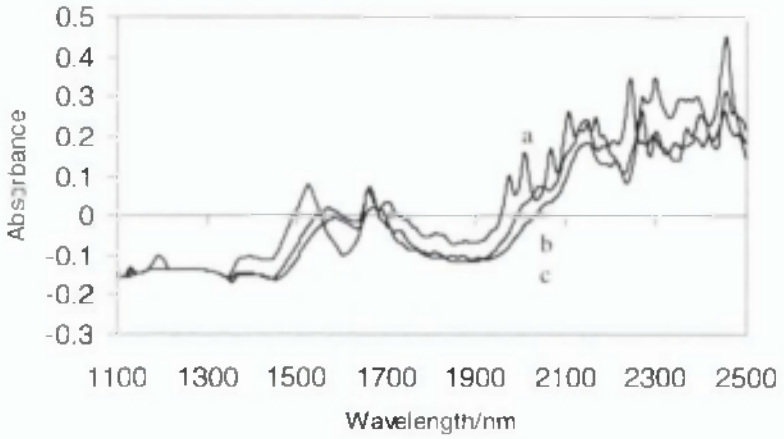


Fig. 6.9 Spectral anomaly occuring below the wavelength of 1500 nm observed with spectra measured on instrument II a, butyl *para* hydroxybenzoate; b, methyl *para* hydroxybenzoate and c, propyl *para* hydroxybenzoate.

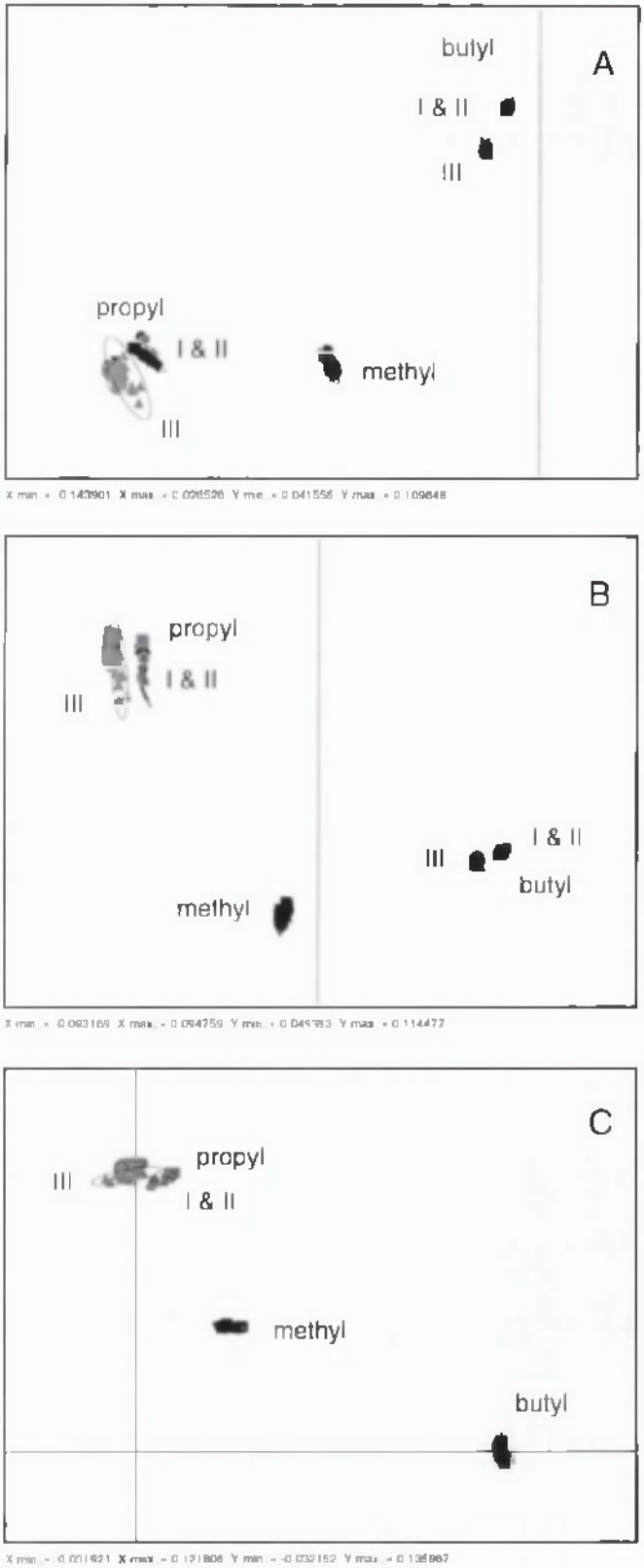


Fig. 6.10 Centre of gravity plots for different batches of methyl, propyl and butyl *para* hydroxybenzoates measured on instruments I, II and III. Second derivative normalised spectra. (A) 1500 - 1900 nm, (B) 1500 - 2000 nm, and (C) 1500 - 2100 nm.

difficult to control.

Relative reflectance measurements which are independent of the instrument or sample accessory used are required if true transferability of spectral data is to be achieved. The fact that different instrument manufacturers use different materials for their reflectance standards is only one of the problems.

Reflectance measurements made through the bottom of a transparent sample cup will always be subject to problems from stray radiation reflected from the various surfaces, **Fig. 6.11A**. The stray radiation will be made up of many components. There will be radiation reflected from the sample stage itself as well as from the sample cup. That from the sample stage will be instrument dependent, while that from the sample cup will of course depend upon the particular cup used. When a reference measurement is taken, **Fig. 6.11B**, again there will be stray radiation from the sample stage. Some reference standards have a covering window (e.g. FOSS NIRSystems) and there will therefore be a contribution of stray radiation from this. Whether radiation reflected from the reference window is considered as stray or part of the reference signal is a mute point. The radiation reflected by the sample in the cup (or reference standard) will similarly undergo reflections at all the interfaces as it makes its way to the detectors. For a given instrument/sample cup this will simply mean all reflected radiation will be decreased by a constant factor.

If there is any hope of obtaining truly transferable spectra it will be at least necessary to use a common reference standard and to eliminate stray radiation effects as much as possible. For the sample, a value for the stray radiation can be obtained by measuring the sample cup filled with a non-reflecting substance. Carbon black is a possible material. Alternatively, the empty sample cup could be measured providing it has no top or backing plate to reflect radiation. Note with an empty cup there is also the problem of radiation reflected by the instrument housing. The ideal reflectance standard is one that is readily available or can be prepared in the laboratory. A powder that can be measured in the sample cup and hence corrected for stray radiation would

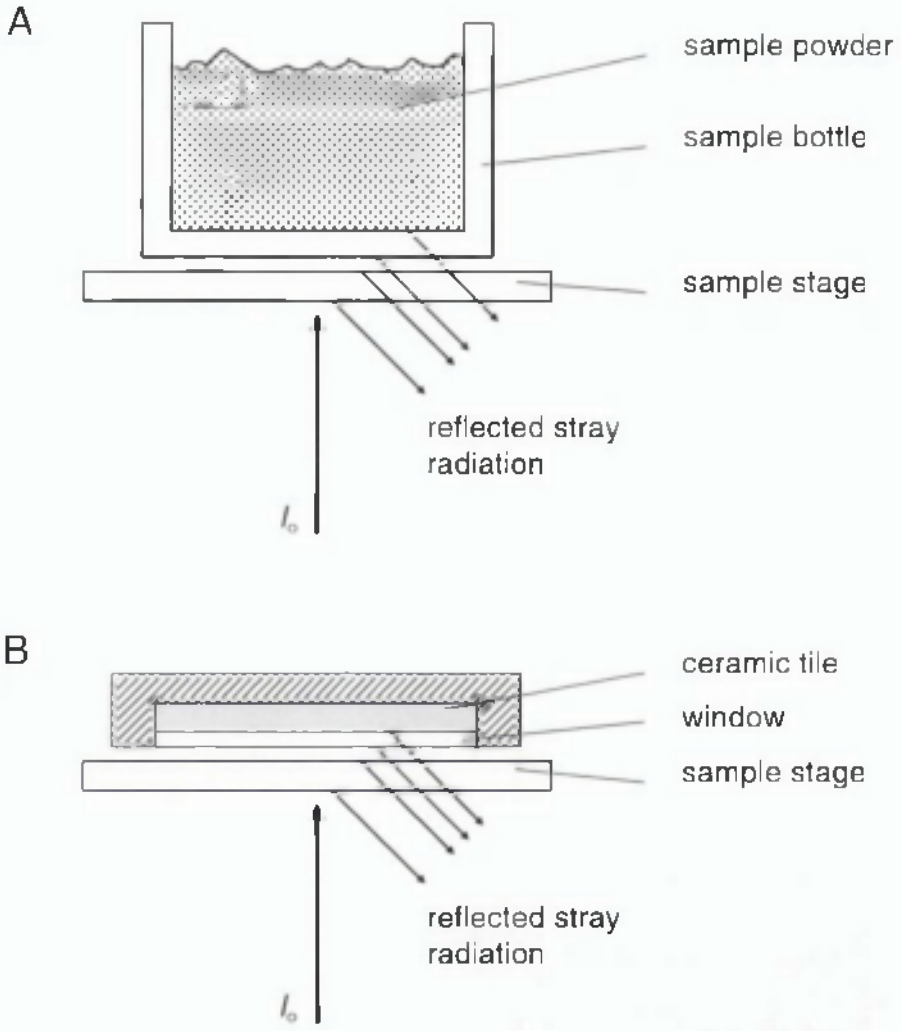


Fig. 6.11 Stray radiation reflected from various surfaces.
(A) sample measurement. (B) reference measurement.

be ideal. The powder would need to be unaffected by re-packing problems.

The theory developed below shows how reflectance measurements made using a sample cup may be corrected for stray radiation and then calculated with respect to a powdered reference standard measured in the same cup.

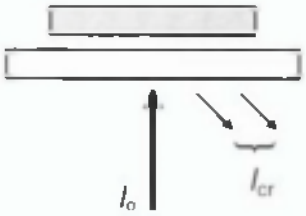
Fig. 6.12. A to **D** illustrate the measurements that are required. It will be assumed that all reflectance measurements are made with respect to the instrument manufacturers reference standard provided. This is represented by **Fig. 6.12A** and the measurement (I_{cr}) is normally held internally by the instrument software and not directly available to the user. The value of I_{cr} includes any stray radiation reflected by the stage *etc.* What is required is the reflectance of the sample with respect to the powdered reference standard and corrected for stray radiation, $R^{\circ}_{samp/ref}$.

$$R^{\circ}_{samp/ref} = \frac{I_{samp}}{I_{ref}} \qquad \textbf{(6.1)}$$

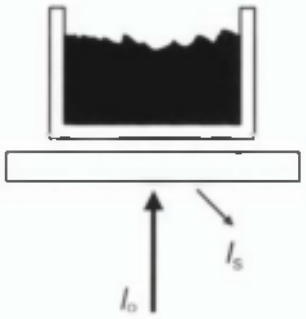
The relative reflectance of the sample with respect to the ceramic standard (**Fig. 6.12C**) and corrected for stray radiation measured using the sample cup filled with carbon black (**Fig. 6.12B**) can be calculated according to:

$$R^{\circ}_{samp/cr} = \frac{I_{samp}}{I_{cr}} = \frac{I_{samp} + I_s}{I_{cr}} - \frac{I_s}{I_{cr}}$$

$$\therefore R^{\circ}_{sample} = R_{sample} - R_{slcr} \qquad \textbf{(6.2)}$$

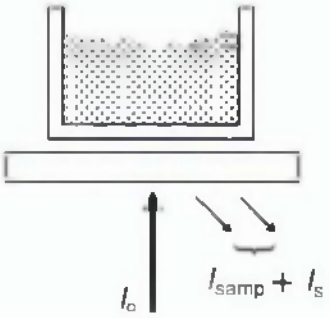


A Ceramic reference standard



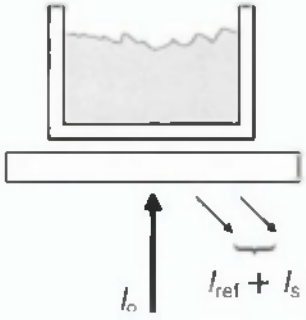
B Stray radiation (carbon black)

$$R_{sler} = \frac{I_s}{I_{cr}}$$



C Sample

$$R_{sampler} = \frac{I_{samp} + I_s}{I_{cr}}$$



D Powdered reference

$$R_{refler} = \frac{I_{ref} + I_s}{I_{cr}}$$

Fig. 6.12 Measurements for stray radiation correction.

Similarly the relative reflectance of the powdered reference with respect to the ceramic reference (**Fig. 6.12D**) and corrected for stray radiation is given by:

$$R^{\circ}_{ref/cr} = \frac{I_{ref}}{I_{cr}} = \frac{I_{ref} + I_s}{I_{cr}} - \frac{I_s}{I_{cr}}$$

$$\therefore R^{\circ}_{ref/cr} = R_{ref/cr} - R_{s/cr} \qquad \textbf{(6.3)}$$

The relative reflectance of the sample with respect to the powdered reference can then be calculated as follows:

$$R^{\circ}_{samp/ref} = \frac{I_{samp}}{I_{ref}} = \frac{\frac{I_{samp}}{I_{cr}}}{\frac{I_{ref}}{I_{cr}}} = \frac{R^{\circ}_{samp/cr}}{R^{\circ}_{ref/cr}}$$

Using $A = -\log_{10} R$, therefore, $A^{\circ}_{samp/ref} = A^{\circ}_{samp} - A^{\circ}_{ref}$ **(6.4)**

Program STRAY was written to perform these transformations automatically so that sample spectra could be corrected for stray radiation and the apparent absorbance with respect to a powdered reference calculated.

6.3.5 Experimental verification for stray radiation and reference correction

The difference between two instrument manufacturers reference standards can be clearly seen in **Fig. 6.13** which shows the Bran+Luebbe Spectralon® reference measured with respect to the FOSS ceramic standard. The negative absorbance values simply indicate that more radiation is reflected from the Spectralon® standard. A simple offset would only alter the magnitude of sample spectra measured using these two standards, however, the fact that the trace is not absolutely flat and/or horizontal means that sample spectra will be distorted in shape.

Even measurements all made on one instrument and with respect to the same standard depend upon the sample cup used. Measurements of PVC powder in a number of

different types of sample bottles were made on instrument I to see if it was possible to use the theory described in section 6.3.4 to correct for spectral differences. **Fig. 6.14** shows the spectra of the PVC powder measured in the various bottles. The smaller the diameter of the bottle the greater the absorbance offset from the baseline. Spectrum 'a' was obtained using a brown glass sample bottle and consequently shows distortions due to absorption from the bottle. The different bottles also reflect different amounts of stray radiation and this can be seen from the spectra obtained when the bottles were filled with carbon black, **Fig. 6.15**. The larger the bottle the greater the amount of stray radiation i.e. smaller absorbance. For example the Waters 1 ml bottles gave the equivalent of about 5% stray radiation at 1100 nm reducing to about 2% at 2500 nm as compared to the radiation reflected by the ceramic reference while the corresponding stray light for the Waters 4 ml bottles were 6.3% and 3.2%.

No investigations were carried out to find the best powdered reference substance, but barium sulphate was used to demonstrate the feasibility of the proposed correction scheme. **Fig. 6.16** shows the spectra of barium sulphate measured with respect to the ceramic reference as obtained in the various bottles. The PVC spectra were corrected for stray radiation and the apparent absorbance with respect to the barium sulphate standard calculated as described in section 6.3.4. Corrected spectra are shown in **Fig. 6.17** and clearly show that the spectra measured in the three clear glass bottles now agree well. The spectrum measured in the brown glass bottle is much improved, but does still differ a little. Note different bottles of each type were used for the sample, carbon black and barium sulphate. Ideally the same bottle should have been used for all three measurements, however, bottles of the same type can be assumed to have similar reflective properties. Centre of gravity plots for the uncorrected and corrected spectra of PVC powder in the various bottles are shown in **Fig. 6.18**. The corrected values are all much closer together than for the uncorrected spectra. The spectra for the three *para* hydroxybenzoates measured using the three different instruments were corrected for stray radiation and referenced to barium sulphate as above and displayed as a centre of gravity plot, **Fig. 6.19**. An improved overlap between the different batches of the same material measured on the three different instruments is observed (cf. **Fig. 6.10**). The spectral correction procedure while not perfect, does considerably

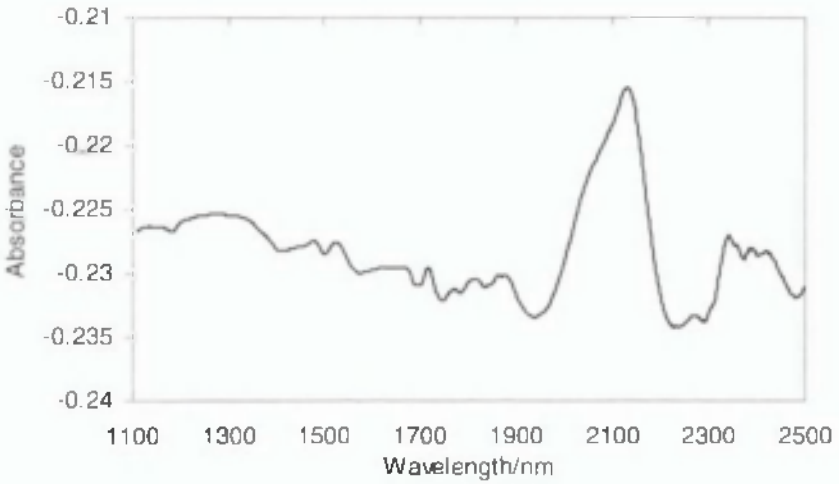


Fig. 6.13 Spectrum of Spectralon® (Bran+Luebbe reference standard) measured with respect to ceramic (FOSS reference standard).

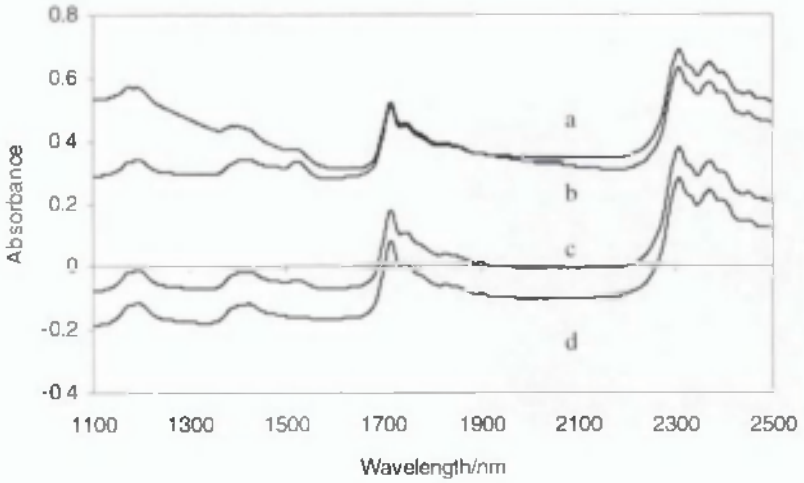


Fig. 6.14 Absorbance spectra of polyvinyl chloride measured in a, brown bottle; b, Waters 1 ml vial; c, Waters 4 ml vial and d, clear neutral glass.

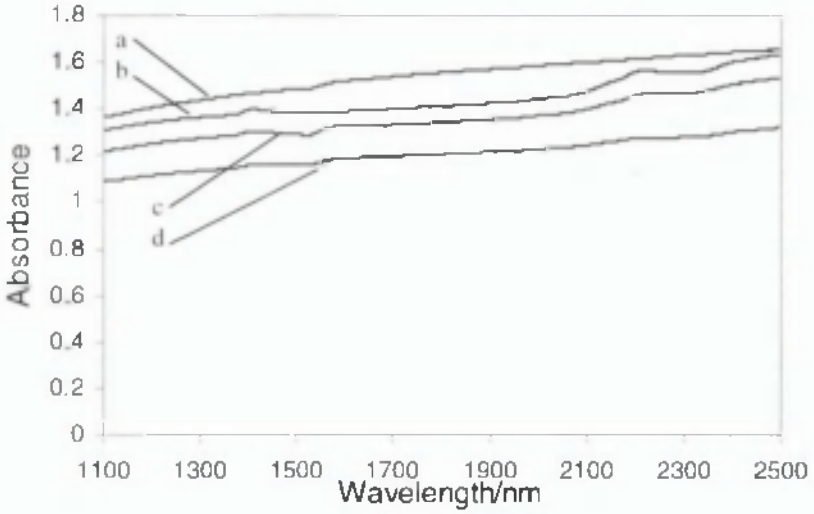


Fig. 6.15 Spectra of carbon black measured with respect to ceramic reference in various types of sample bottles.
a. brown bottle; b. Waters 1 ml; c. Waters 4 ml and d.clear neutral glass.

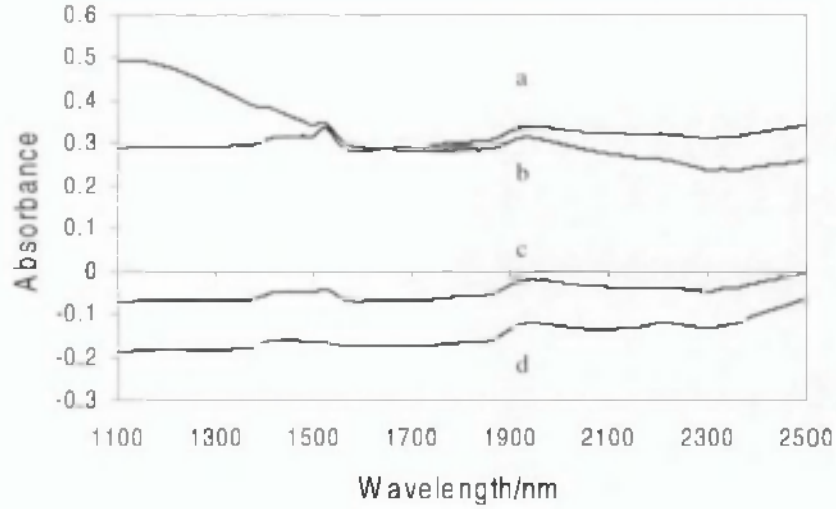


Fig. 6.16 Spectra of barium sulphate measured with respect to ceramic reference in various types of sample bottles.
a. brown bottle; b. Waters 1 ml; c. Waters 4 ml and d.clear neutral glass.

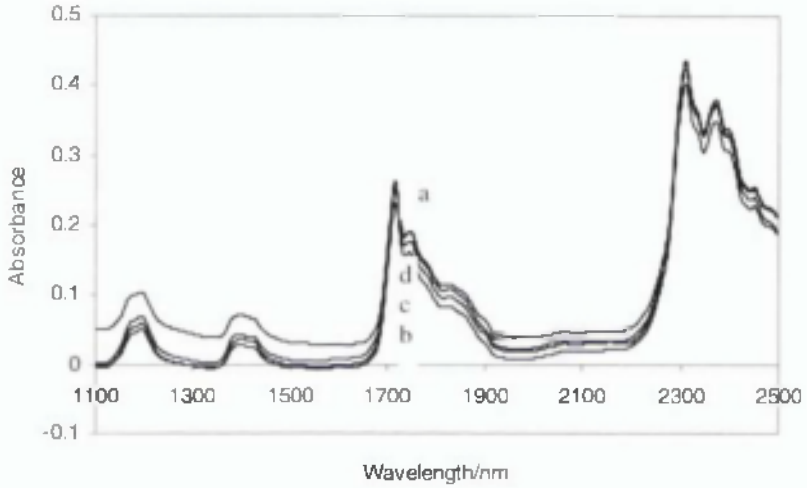


Fig. 6.17 Corrected absorbance spectra of polyvinyl chloride measured in a, brown bottle; b, Waters 1 ml vial; c, Waters 4 ml vial and d, clear neutral glass.

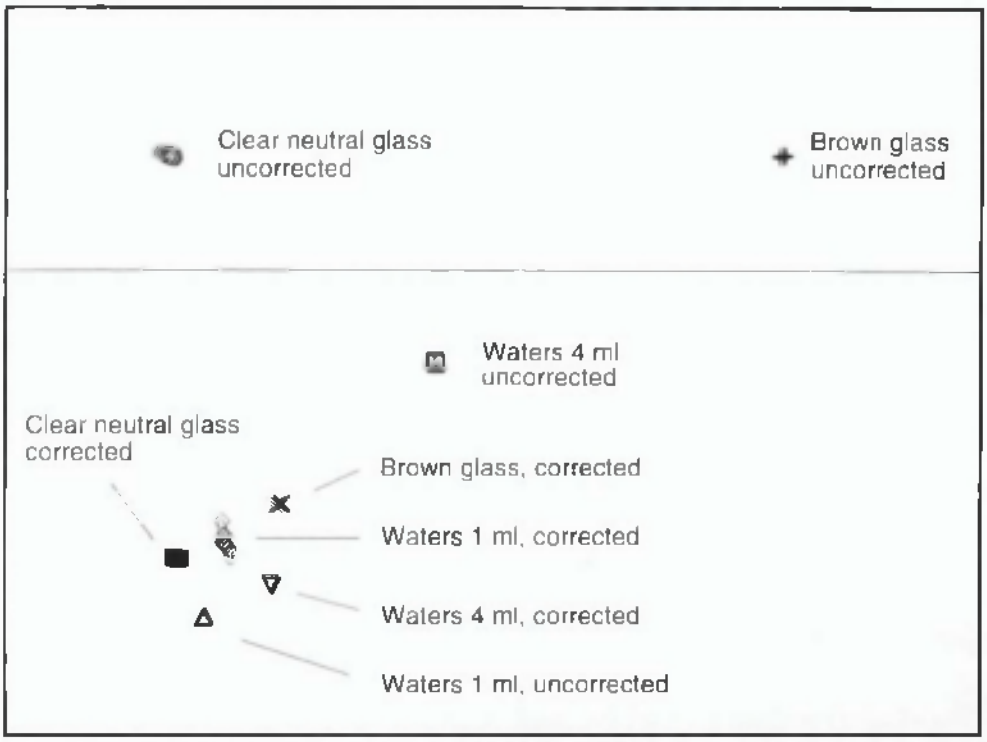


Fig. 6.18 Centre of gravity plots for polyvinyl chloride measured in different sample cups. Measured on instrument 1 over the wavelength range 1100 to 2498 nm. Plots for both corrected and original spectra shown.

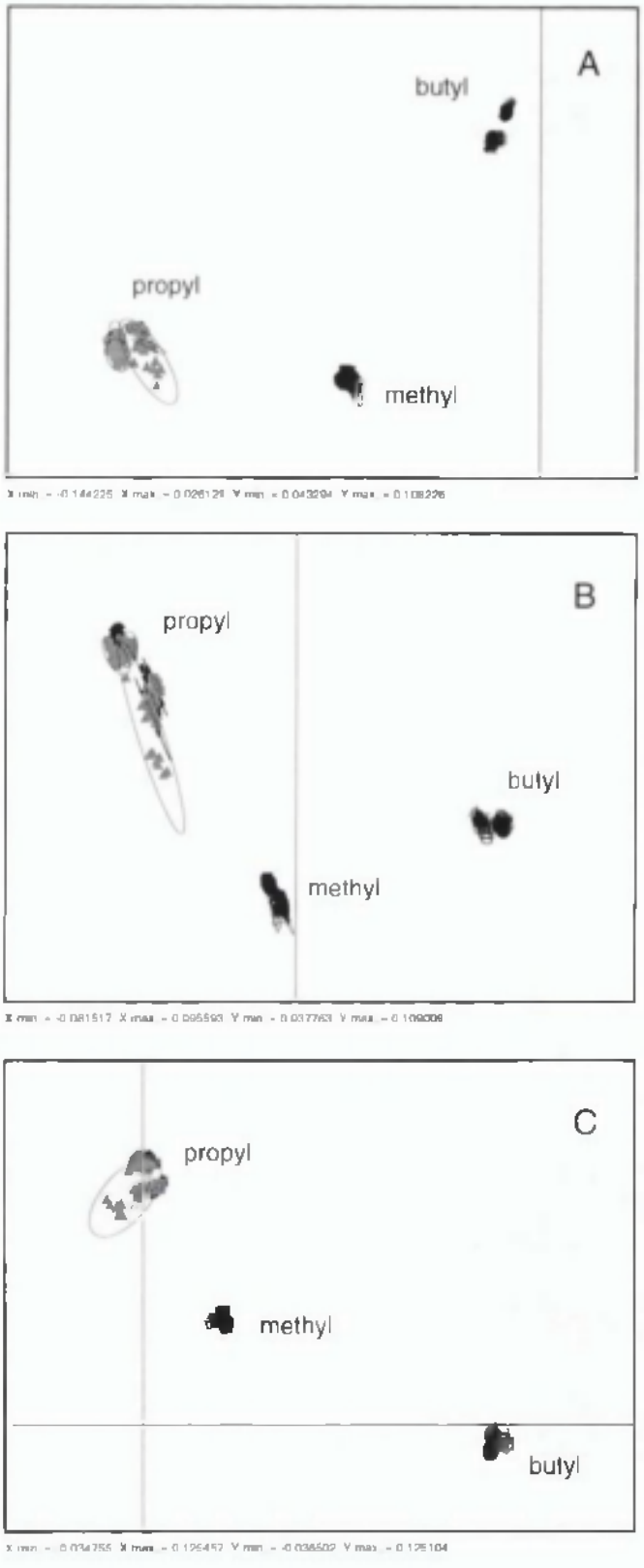


Fig. 6.19 Centre of gravity plots for *corrected* spectra. Different batches of methyl, propyl and butyl *para* hydroxybenzoates measured on instruments I, II and III. (A) 1500 - 1900 nm, (B) 1500 - 2000 nm, and (C) 1500 - 2100 nm.

improve transferability.

6.3.6 Effect of sample diameter

In chapter 5 the effects of sample diameter were investigated by placing a sample on an iris diaphragm and measuring spectra for different settings to give the spectra shown in **Fig. 5.2**. These results are at variance with the effects of different bottle diameters observed in section 6.3.5 above, where a reduction in diameter causes an increasing baseline offset but little change in magnitude of the size of the peaks. This difference is due undoubtedly to radiation reflected from the under side of the iris diaphragm giving a very high stray radiation component. The effect observed in **Fig. 6.20** can be simulated by introducing stray radiation into a spectrum according to:

$$R = \frac{d \times I + s \times I_{cr}}{I_{cr}}$$

Where d is a function of the iris diameter, s is the fraction of stray radiation, I the intensity of radiation reflected by the sample at 'infinite diameter' and I_{cr} the intensity of radiation from the ceramic reference. An example of a simulated set of spectra for lactose monohydrate are shown in **Fig. 6.20**. The values of d and s were chosen on an arbitrary basis, however, the agreement between **Fig. 5.2** and **6.20** suggests this is the correct explanation.

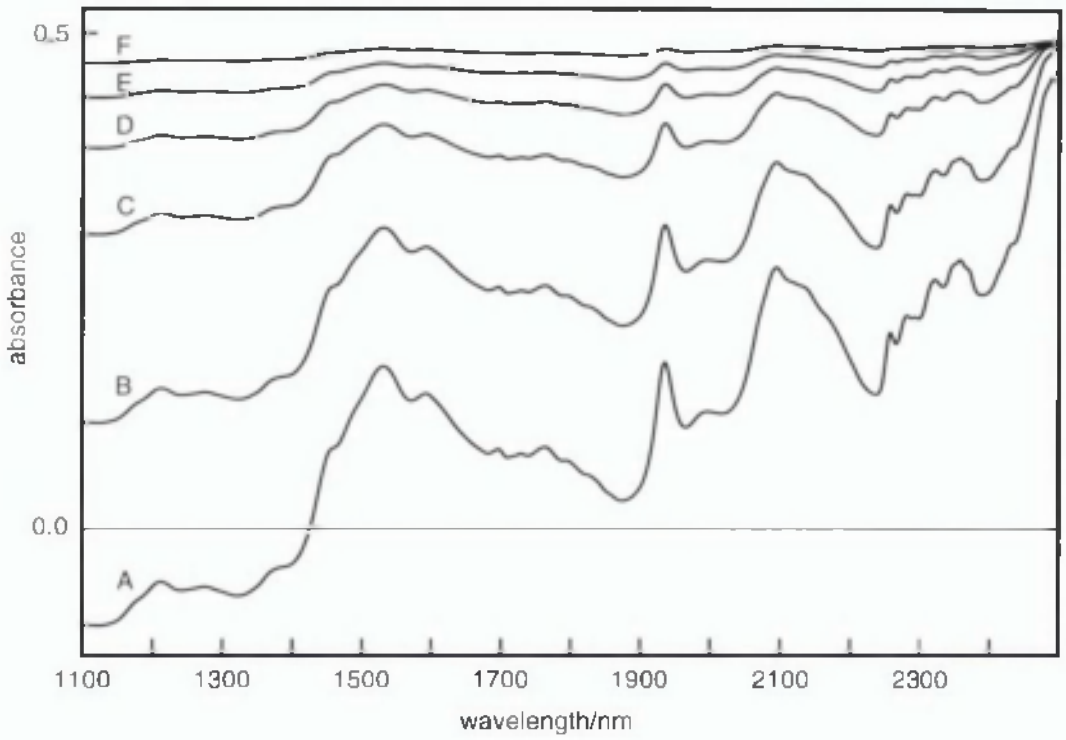


Fig. 6.20 Simulated spectra illustrating the effect of stray radiation and sample diameter according to equation 6.6.
(A) $d = 1$, $s = 0.02$, (B) $d = 0.5$, $s = 0.165$, (C) $d = 0.2$, $s = 0.259$, (D) $d = 0.1$, $s = 0.291$, (E) $d = 0.05$, $s = 0.307$, and (F) $d = 0.02$, $s = 0.316$.

6.4 Conclusion

PQS offers a sensitive and robust method for the identification of pharmaceutical excipients with the optimisation of data pretreatment. With its sensitivity, is also offers tremendous potential for qualification of material. Spectral differences arising from instrumental differences, although visible in polar plots, did not appear to affect the differentiation process between closely related compounds i.e. *para* hydroxybenzoates. Stray light arising from different instruments and/or sample bottles and also variations in reflectance standards are potential sources of errors. The suggested method for correction of stray light, by measuring a common reference sample and total absorber in the sample bottle, has been shown to improve the agreement of spectra measured under different conditions.

CHAPTER 7: GENERAL CONCLUSION

This systematic study has shown that identification applications based on NIR spectroscopy can be transferred. Chapter 3 and 4 shows that for solvents, which give rise to spectra which are relatively simple, a fully validated method based on Correlation Coefficient in Wavelength Space can be transferred provided that wavelength parameters such as data pre-treatment and wavelength range are optimised. The method was found to be robust to changes in ambient conditions (i.e. temperature and humidity changes, wavelength and band-pass errors between instruments, use of different instruments and also inter-laboratory variation). The success of this part of the work is particularly marked by the ability to directly transfer the method between instruments based on completely different optical principles i.e. diffraction grating monochromators and FT instruments, without the need to apply correctional methods which have been so heavily relied upon in previous studies. Apart from instrumental differences, chapter 4 highlights another issue to be dealt with to achieve the holistic transferability of NIR applications: uniformity of spectral file formats. File formats originating from the software packages of the three instrument manufacturers used were found to be different. To circumvent this issue, it was decided to use in-house written software. However, within the use of NIR spectroscopy in pharmaceutical analysis, greater efforts on the part of the instrument manufacturers would have be to be expended to harmonise the software specifications.

Transferring of solid samples proved more difficult, due to the contribution of two additional sources of variation: differences in reflectance standard and also variations in sample presentation (as established in chapter 5). Sample presentation factors such as cup diameter, sample thickness, cup material and packing method were found to be potential sources of errors for two methods of identification i.e. Correlation in Wavelength Space and Maximum Wavelength Distance and their effects could not be fully correctly by data pre-treatment alone. Therefore, optimisation of these factors should be built into any Standard Operating Procedure when constructing spectral libraries.

In Chapter 6, a new algorithm, the Polar Qualification System, was chosen to differentiate between pharmaceutical excipients. Centre of gravity plots showed that PQS is a sensitive method with its ability to differentiate between closely related compounds such as anhydrous lactose with lactose monohydrate and the different types of *para* hydroxybenzoates. Effects of instrumental differences, although visible on centre of gravity plots, were not sufficient to cause a mis-identification.

From the feasibility study on PQS method carried out in chapter 6, it is evident that this method presents enormous potential as a method of identification and even qualification for solid powdered excipients. To be able to fully establish its capability as a transferable algorithm, a more thorough investigation (as carried out in chapter 3) involving more samples (to give a better representation on the population of the data-set) would be recommended. In any eventual further investigation, differences in reflectance standards, stray light arising from different instrumental design and sample bottles would continue to present a challenge to transferability. For measurements using horizontal setups, one possible solution is to use a common reflectance standard such as barium sulphate and total absorber like carbon black to obtain a reflectance measurement independent of these sources of error. The last section of the this chapter demonstrated that this theory improves the agreement of spectra measured in different types of sample bottles and also using different instruments.

REFERENCES

A

Adams, K. J., **1995,** *Chemometrics in Analytical Spectroscopy*, The Royal Society of Chemistry, Cambridge, UK.

ASTM, Standard Definitions of Terms and Symbols relating to Molecular Spectroscopy, Vol. 14.01, Standard E131-90.

B

Banwell, C. N., **1966**, *Fundamentals of molecular spectroscopy*, McGraw-Hill Publishing Company Limited, Berkshire, England.

Barnes, R. J., Dhanoa, M. S. and Lister, S. J., **1989**, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, *Appl. Spectrosc.*, 4, 772-777.

Bjorsvik, H. R. and Martens, H., **1992**, Calibration of NIR instruments by PLS regression. In: *Handbook of Near-Infrared Analysis*, Burns, D. A. and Ciurczak, E. W. (eds.), Marcel Dekker, New York, USA.

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. and de la Pezuela, C., **1996,** Quantification of the active compound and major excipients in a pharmaceutical formulation by near-infrared diffuse reflectance spectroscopy with the fibre optic probe, *Anal. Chim. Acta*, 333, 147-156.

Blanco, M., Coello, J., Iturriaga, H., Maspoch S. and de la Pezuela, C., **1997,** Effect of data preprocessing methods in near-infrared diffuse reflectance

spectroscopy for the determination of the active compound in a pharmaceutical preparation, *Appl. Spectrosc.*, 51, 240-246.

Blanco, M., Coello, J., Iturriaga, H., Maspoch S. and de la Pezuela, C., **1997a,** Strategies for constructing the calibration set in the determination of active principles in pharmaceuticals by near infrared diffuse reflectance spectrometry, *Analyst,* 122, 761-765.

Blanco, M., Coello, J., Iturriaga, H., Maspoch S. and de la Pezuela, C., **1998,** Critical review: near infrared spectroscopy in the pharmaceutical industry, *Analyst,* 123, 135R-150R.

Bouveresse, E. and Massart, D. L., **1996,** Standardisation of near-infrared spectrometric instruments: a review, *Vib. Spectrosc.,* 11, 3-15.

Bouveresse, E., Hartmann, C., Massart, D. L., Last, I. R. and Prebble, K. A., **1996a,** Standardisation of near-infrared spectrometric instruments, *Anal. Chem.,* 68, 982-990.

Bouveresse, E., Rutan, S. C., Heyden, Y. V., Pennickx, W. and Massart, D. L., **1997,** Detection, interpretation and correction of changes in the instrumental response of near-infrared monochoromator instruments over time, *Anal.Chim. Acta,* 348, 283-301.

Bran+Luebbe, **1996,** *Operation Manual,* GmbH, Norderstedt, Germany.

British Pharmacopoeia, **1999,** H.M. Stationery Office, London, UK.

Brunner, S., Groh, A. and Fischer, M., **1999,** The use of near-spectroscopy in the pharmaceutical control for raw materials, *Proceedings to the 9th International Conference on near-infrared spectroscopy,* Verona, Italy, O.10 – 1.

Buchanan, B., **1992,** NIR Analysis of Petrochemicals, In: *Handbook of Near-Infrared Analysis*, Burns, D. A. and Ciurczak, E. W. (eds.), Marcel Dekker, New York, USA.

Bureau Central Internationale de l'Eclairage, **1970,** International Lighting Vocabulary, *CIE Publication 17,* Paris, France.

C

Candolfi, A., Wu, W., Massart, D. L . and Heuerding, S., **1998,** Comparison of classification approaches applied to NIR-spectra of clinical study lots, *J.Pharm. Biomed. Anal.*, 16, 1329-1347.

Choquette, S. J., Travis, J. C., Duewer, D. L. and Zhu, C., **1999,** SRM 2035: a rare earth oxide for wavelength calibration of near infrared spectrometers, *Proceedings to the 9th International Conference on Near-infrared Spectroscopy,* Verona, Italy, O.10 – 5.

Ciurczak, E. W. and Maldecker, T. A., **1986,** Identification of actives in multicomponent pharmaceutical dosage forms using near-infrared reflectance analysis, *Spectroscopy*, 1, 7, 36-39

Ciurczak, E. W, Torlini, R. P. and Demkowicz, M. P., **1986a,** Determination of particle size of pharmaceutical raw materials using near-infrared reflectance spectroscopy, *Spectroscopy,* 1, 1, 36-39.

Ciurczak, E. W. and Torlini, R. P., **1987,** Analysis of solid and liquid dosage forms using near-infrared reflectance spectroscopy, *Spectroscopy,* 2, 41-43.

Ciurczak, E. W., **1991,** Pharmaceutical mixing studies using near-infrared spectroscopy, *Pharm. Tech.*, 15, 140-145.

Corti, P., Dreassi, C., Murratzu, C., Corbini, G., Bullerini, L. and Gravina, S., **1989,** Applications of NIRS to the quantity control of pharmaceuticals ketoprofen assay in different pharmaceutical formulae, *Pharm. Acta. Helv.*, 4, 140-145.

Corti, P., Dreassi, E., Corbini, G., Lonardi, S., Viviani, R., Mosconi, L. and Bernuzzi, M., **1990,** Application of near infrared reflectance to the analytical control of pharmaceuticals. Assay of ranitidine and water content in tablets, *Pharm. Acta. Helv.*, 65, 28-32.

D

Davies, A. M. C. and Grant, A., **1987,** Air-conditioning-generated noise in a near-infrared spectrometer caused by fluctuations in atmospherix water vapour, *Appl. Spectrosc.*, 41, 1248-1250.

DeBraekeleer, K., Sanchez, F. C., Hailey, P. A., Sharp, D. C. A., Pettman, A. J. and Massart, D. L., **1998,** Influence and correction of temperature perturbations on NIR spectra during the monitoring of a polymorph conversion process prior to self-modelling mixture analysis, *J. Pharm. Biomed. Anal,* 17, 141-152.

Delwiche, S. R., Pierce, R. O., Chung, O. K. and Seabourn, B. W., **1998,** Protein content of wheat by near-infrared spectroscopy of whole grain: collaborative study, *J. Assoc. Off. Anal. Chem.*, 81, 587-603.

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.

Devaux, M. F., Dufour, N. N., Robert, P. and Bertrand, D., **1995,** Effects of particle aize on the near-infrared spectra of wheat and rape seed meal mixtures, *Appl. Spectrosc.*, 49, 84-91.

Dreassi, E., Ceramelli, G., Perruccio, P. L. and Corti, P., **1998,** Transfer of calibration in near-infrared reflectance spectrometry, *Analyst,* 123, 1259-1264.

E

Ellison, S. L. R., Greogory, S. and Hardcastle, W. A., **1998,** Quantifying uncertainty in qualitative analysis, *Analyst,* 123,1155-1161.

European Commission, **1997**, *The Rules Governing Medicinal Products in the European Union: Good Manufacturing Products for Human & Veterinary Use*, Strausbourg, 4, 103.

European Pharmacopoeia, **1997,** European Department for the Quality of Medicines within the Council of Europe, Strausbourg, 3rd edn, 43- 45.

F

FOSS NIRSystems, **1996,** *NSAS Training Manual,* 12101 Tech Road, Silver Spring, MD, USA.

FOSS NIRSystems, **1996a,** *InTact® Tablet Analyzer Installation and User Manual*, 12101 Tech Road, Silver Spring, MD, USA.

G

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

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

Gemperline, P. J., Webber, L. D. and Cox, F. O., **1989,** Raw materials testing using soft independent modelling of class analogy analysis of near-infrared reflectance spectra, *Anal. Chem.*, 61, 138-144.

Gerhausser, C. I. and Kovar, K. A., **1997**, Strategies for constructing near-infrared libraries for the identification of drug substances, *Appl. Spectrosc.*, 51, 1504-1510.

Ghosh, S.and Rodgers, J., **1992,** NIR Analysis of textiles, In: Handbook of Near-Infrared Analysis, Burns, D. A. and Ciurczak, E. W. (eds.), Marcel Dekker, New York, USA.

Gonzalez, F. and Pous, R., **1995,** Quality control in manufacturing process by near-infrared spectroscopy, *J. Pharm. Biomed. Anal.*, 13, 419-423.

Griffin, J., **1993,** A method by which diffuse reflectance measurements made on different spectrometers of the same family can be standardised, Technical Note, Bran+ Luebbe, GmbH, Norderstedt, Germany.

Grum, F. and Saltzman, M., **1976,** P-75-77 New white standard reflectance, *CIE Publication 36,* Paris, France, 91-98.

H

Hailey, P. A., Doherty, P., Tapsell, P., Oliver, T. and Aldridge, P. K., **1996,** Automated system for the on-line monitoring of powder blending processes using near-infrared spectroscopy Part I System development and control, *J. Pharm. Biomed. Anal.*, 14, 551-559.

Hammond, J., Kellam, B., Moffat A. C. and Jee, R. D., **1999,** Non-destructive real-time reaction monitoring in solid phase synthesis by near-infrared reflectance spectroscopy – esterification of a resin-bound alcohol, *Anal. Commun.*, 36, 127-129.

Hammond, S., **1996,** *Proceedings to the Perstorp Analytical/NIRSystems Pharmaceutical User Group Meeting*, Astra Hassle, Molndal, Sweden,

Han, S. M. and Faulkner, P. G., **1996,** Determination of SB216269-S during tablet production using near-infrared reflectance spectroscopy, *J. Pharm. Biomed. Anal.*, 14, 1681-1689.

Hindle, P. H. and Smith, C. R. R., **1996,** A comparison of calibration robustness relating to different data reatments of a standard set of spectra, *J. Near Infrared Spectrosc.*, 4, 119-128.

J

Jones, J. A., Last, I. R., MacDonald, B. F. and Prebble, K., **1993**, Development and transferability of near-infrared methods for determination of moisture in a freeze-dried injection product, *J. Pharm. Biomed. Anal,* 11, 1077-1085.

Jouan-Rimbaud, D., Khots, M. S., Massart, D.L., Last, I. R. and Prebble, K. A., **1995,** Calibration line adjustment to facilitate the use of synthetic calibration samples in near-infrared spectrometric analysis of pharmaceutical production samples, *Anal. Chim. Acta,* 315, 257-266.

K

Khan, P. R., Jee, R. D., Watt, R. A. and Moffat, A. C., **1997,** The identification of active drugs in tablets using near-infrared spectroscopy, *Pharm. Sci.*, 3, 447–453.

Knee, P. C. and Deadman, A. J., **1998**, Intercomparison of near infrared diffuse reflectance measurements, Report COEM 6, *National Physics Laboratory,* Teddington, England.

Kowalski, B. R., Callis, J. B. and Illman, D. L., **1987,** Process analytical chemistry, *Anal. Chem.* 59, 624A-635A.

Kradjel, C. and McDermott, L., **1992,** NIR Analysis of Polymers, In*: Handbook of Near-Infrared Analysis*, Burns, D. A. and Ciurczak, E. W. (eds.), Marcel Dekker, New York, USA.

L

Last, I. R. and Prebble, K. A., **1993,** Suitability of near-infrared methods for the determination of moisture in a freeze-dried injection product containing different amounts of the active ingredient, *J. Pharm. Biomed. Anal,* 11, 1671-1676.

Lieper, K., Vessman, J. and Van der Vlies, C., **1998,** Issues relating to establishment of validation guidelines for material identification using NIR, *Pharmeuropa,* 10, 468-470.

Lonardi, S., Viviani, R., Mosconi, L., Bernuzzi, M., Corti, P., Dreassi, E., Murratzu, C. and Corbini, G., **1989,** Drug analysis by near-infrared reflectance spectroscopy. Determination of the active ingredient and water content in antibiotic powders., *J. Pharm.Biomed. Anal.,* 7, 303-308.

M

McDonald, R. S. and Wilks P. A., JR., **1988,** JCAMP-DX: A standard form for exchange of infrared spectra in computer readable form, *Appl. Spectrocs.,* 42, 151-162.

Mark, H. L., and Tunnel, D., **1985,** Qualitative near-infrared reflectance analysis using Mahalanobis Distance, *Anal. Chem.,* 57, 1449-1456.

Mark, H., and Workman, J., **1988,** A new approach to generating transferable calibrations for quantitative near-infrared spectroscopy, *Spectroscopy,* 11, 28-37.

Mark, H., **1992,** Data analysis: multilinear regression and principal component analysis, In: *Handbook of Near-Infrared Analysis,* Burns, D. A. and Ciurczak, E. W. (eds.), Marcel Dekker, New York, USA.

Mark, H., **1992a,** Qualitative discriminant analysis, In: *Handbook of Near-Infrared Analysis,* Burns, D. A. and Ciurczak, E. W. (eds.), Marcel Dekker, New York, USA.

Mark. H. and Workman, J., **1986,** Effect of repack on calibrations produced for near-infrared reflectance analysis, *Anal. Chem.,* 58, 1454-1459.

McClure, W. F., **1994,** Near-infrared spectroscopy: the giant is running strong, *Anal. Chem.,* 66, 43A-53A.

Miller, C. E., **1993,** Sources of non-linearity in near infrared methods, *NIR News*, 6, 3–5.

Miller, J. C. and Miller, J. N., **1993,** in *Statistics for Analytical Chemistry*, 3rd ed., Ellis Horwood, Chichester, UK.

Moffat, A. C., **1998,** The case for the use of near-infrared spectroscopy as a primary method, *J.Pharm. Pharmacol.*, 50 (Supplement), 46.

Moffat, A. C., Jee, R. D. and Watt, R., **1997**, New Near-Infrared Centre of Excellence for Europe, *Eur. Pharm. Rev.*, 37-41.

Morisseau, K. M. and Rhodes C. T., **1997,** Near-infrared spectroscopy as non-destructive alternative to conventional tablet hardness testing, *Pharm. Res.*, 14, 108-111.

N

Næs, T. and Isaksson, T., **1994,** Some aspects of scatter-correction in calibration. Part I, *NIR News*, 5, 4-5.

Norris, K., **1962,** Instrumentation of infrared radiation, *Transactions of the American Society of Agricultural Engineers*, 1962, 17-20.

Norris, T., Aldridge, P. K. and Sekulic, S. S., **1997,** Determination of end-points for polymorph conversions of crystalline organic compounds using on-line near-infrared spectroscopy, *Analyst*, 122, 549-552.

O

O'Neil, A. J., Jee, R. D. and Moffat, A. C., **1998,** The application of multiple linear regression to the measurement of the median particle size of drugs and pharmaceutical excipients by near-infrared spectroscopy, *Analyst*, 123, 2297-2302.

O'Neil, A. J., Jee, R. D. and Moffat, A. C., **1999,** Measurement of the cumulative particle size distribution of microcrystalline cellulose using near-infrared reflectance spectroscopy, *Analyst,* 124, 33-36.

Olinger, J. M. and Griffiths, P. R., **1992,** Theory of Diffuse Reflectance in the NIR region, In: *Handbook of Near-Infrared Analysis*, Burns, D. A. and Ciurczak, E. W. (eds.), Marcel Dekker, New York, USA.

Olinger, 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- and mid-infrared. Part I – Single Analytes, *Appl. Spectrosc.*, 47, 687-694.

Osborne, B. G., Fearn. T. and Hindle, **1993**, P. H., *Practical NIR spectroscopy : with applications in food and beverage analysis,* 3rd ed., Longman Scientific and Technical, Essex, England.

P

Plugge, W. and Van der Vlies, C., **1993,** Near infrared spectroscopy as an alternative to asses compliance of ampicillin trihydrate with compendial specifications, *J. Pharm. Biomed. Anal.*, 11, 435-442.

Pouchet, C. J., **1985,** *The Aldrich Library of FT-IR Spectra,* Aldrich Chemical, Milwaukee, W1, USA.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P., **1992,** *Numerical Recipes in C,* 2nd edition, Cambridge University Press, Cambridge, UK.

S

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

Sekulic, S. S., Ward, H. W., Brannegan, D. R., Stanley, E. D., Evans, C. L., Scianvolino, S. T., Hailey, P. A. and Aldridge, P. K., **1996,** On-line monitoring of

powder blend homogeneity by near-infrared spectroscopy, *Anal. Chem.*, 68, 509-513.

Shenk, J. S., Westerhaus, M. O. and Templeton, W. C., **1991,** New standardization and calibration procedures for near-infrared spectroscopy analytical spectrum, *Crop Sci.*, 31, 1694-1696.

Sokal, R. R. and Rohlf, T. J., **1981**, *Biometry*, W. H. Freeman & Company, 2nd ed., New York, USA.

Springsteen, A. and Ricker, T., **1996,** Standards for the measurement of diffuse reflectance, *NIR News,* 7, 6-9.

Swierenga, H., Haanstra, W. G., de Weijer, A. P. and Buydens, L. M., **1998,** Comparison of two different approaches toward model transferability in NIR spectroscopy, *Appl. Spectrosc.*, 52, 7-16.

T

Taylor J.,**1983**, Standard Reference Materials, *Handbook for Standard Reference Materials Users,* NIST Publication 260, 16.

Thompson, M., Mertens, B., Kessler, M. and Fearn, T., **1993,** Efficacy of robust analysis of variance for the interpretation of data from collaborative trials, *Analyst*, 118, 235-240.

Trafford, A. D., Jee, R. D., Moffat, A. C. and Graham, P., **1998,** A rapid quantitative assay of intact paracetamol tablets by reflectance near-infrared spectroscopy, 124, 163-167.

V

Van der Vlies, C., **1996,** Near-infrared spectroscopy and the regulatory hurdles, *Eur. Pharm. Rev.*, February issue, 49-54

Van der Vlies, C., Kaffka, K. J., Plugge, W., **1995,** Qualifying pharmaceutical substances by fingerprinting with NIR spectroscopy and PQS, *Pharm. Technol. Eur.*, 7, 46-49.

Verril, J., **1994,** UVSG meeting on near infrared standards, *The National Physical Laboratory,* Teddington, UK.

W

Weidner, V. R. and Hsia, J. J., **1981,** Reflection properties of pressed polytetrafluoroethylene powder, *J. Opt. Soc. Am.*, 71, 856-861.

Weidner, V. R. and Hsia, J.J. and Adams, B., **1985,** Laboratory intercomparison study of pressed polytetrafluoroethlyene powder reflectance standards, *Appl. Opt.*, 24, 2225-2230.

Weidner, V. R., Barnes, P. Y. and Eckerie, K. L., **1986,** A wavelength standard for the near infrared based on the reflectance of rare-earth oxides, *J. Res. Natl. Bur. Stand.*, 91, 243-253.

Willard, H. H., Merrit, L. L., Jr., Dean, J. A. and Settle, F. A., Jr., **1998**, *Instrumental Methods of Analysis*, 7th edition, Wadsworth Publishing Company, Belmont, California, USA.

Williams, P. C., **1992,** Sample preparation and selection, In: *Handbook of Near-Infrared Analysis,* Burns, D. A. and Ciurczak, E. W. (eds.), Marcel Dekker, New York, USA.

Wood, R.W., **1902,** XLII. On a remarkable case of uneven distribution of light in a diffraction grating spectrum, *Philosophical Magazine*, 4, 396-402.

Wulfert, F., Kok, W. T. and Smilde, A. K., **1998**, Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models, *Anal. Chem.*, 70, 1761-1767.

Z

Zappala, A. F. and Post, A. J., **1977,** Rapid near IR spectrophotometric determination of meprobomate in pharmaceutical preparations, *J. Pharm. Sci.*, 66, 292-299.

APPENDIX I

PROTOCOL

Transferability of a Library for the Identification of Solvents

Background to the collaborative trial

The ability to transfer near infrared applications is an important issue in view of the high expenditure of time, labour and materials needed to construct spectral reference libraries.[1]. Various attempted have been made to transfer NIR quantitative calibrations, though none involved identification methods.

A preliminary study carried out at the Centre for Pharmaceutical Analysis, University of London has shown that a library containing 15 solvents was transferable between 6 Foss NIRSystems systems, of different monochromator models and types of sampling accessories (see Table 1). The algorithm applied was by correlating the measured spectrum to the reference spectrum in wavelength space , between a wavelength range of 1100 - 2100 nm. An identification threshold of 0.97 was used. The solvents used were either Laboratory or Analytical grade reagents, therefore having a specified purity of > 99 % and are listed in Table 2.

Table 1 Foss NIRSystems instrumental systems used during the preliminary study

Instrument system	Monochromator Model	Sampling accessory
1	6500	Interactance Immersion Probe
2	6500	Interactance Immersion Probe
3	6500	Smartprobe®
4	5000	Smartprobe®
5	6500	Interactance Reflectance Probe
6	6500	Rapid Content Analyzer

Table 2 Solvents used within the library

Solvents
Absolute alcohol
Acetone
Butan-1-ol
Butan-2-ol
Chloroform
Dichloromethane
Dimethylformamide
Ethyl acetate
Ethanol 96 %
Methanol
Propan-1-ol
Propan-2-ol
Tetrahydrofuran
Toluene

Objectives:

- to correctly identify solvents included in the library (as listed above), using any of the systems similar to that listed in Table 1
- to correctly exclude other solvents from the library
- to investigate the ruggedness of the ID threshold of 0.97 to sampling accessories not investigated during the preliminary study i.e. Direct Content Analyzer.

Your participation in this experiment would be deeply appreciated. Participants in the experiment will receive feedback of their particular measurements in a report. Names of operators and sites of measurements will remain confidential.

References

1. Blank, B.T., Sum, S. T., Brown, S.D. and Mofre, S. L. (1996), Transfer of Near Infrared Multivariate Calibrations Without Standards. Anal. Chem. 68: 2987- 2995

You are provided with :

- a diskette containing the library of the 15 solvents. Names of files to create library: master.nlb, master.dta, master.hda, master.sda, master. iqq and master.riq
- a detailed set of instructions to carry out the experiments

Instructions

Instrumentation

- a complete set of accessories to measure liquids as defined on the first page.
- please record details of instrument in **Part A** of the **Form I** provided

Samples

- solvents similar to the reference solvents are and also those not included in the library can be used
- minimum of 5 different solvents. We welcome as many solvents as possible.
- laboratory or analytical grade reagents (specified purity > 99 %)
- within expiry date and been opened < 1 year
- reagent bottles should have been stored as recommended by the manufacturer since being opened
- please record detail of solvents used in **Part B** of the **Form I** enclosed

Note : The spectra of hygroscopic solvents i.e. DMF should be measured as soon as they are removed from the bottle

Experimental

1. Installation of library into NSAS

a. If your computer is Windows 95-based, you can copy the library into NSAS within Windows Explorer

b. If your computer is DOS-based, then copy all files i.e. a:\master.* to c:\nirs

2. Setting up of instrument

a. For **probe** accessories- attach the probe tip to the end of the probe and set the *path distance* to 1 mm. Tighten the probe tip in place with an Allen key. This can be carried out accurately using a feeler gauge or the microscope slide provided. *Please note that pathlength = 2 x path distance.*

b. For **horizontal setup** accessories, path distance of gold reflector is already fixed at 0.5 mm

3. ***Log into NSAS programme.***

a. Carry out standard diagnostics tests and record the result in **Form II** or enclose print outs with the results

Wavelength linearization
Noise Test
Bandwidth test

b. To enter IQ^2 mode:

⇒ Return to Main Menu
⇒ Select <IQ^2 Analysis>
⇒ Select <IQ^2 >
⇒ Select file <master >
⇒ Select <Acquire Data>

4. ***Scan the reference measurement - air***

Please do not use ceramic for reference measurement.

Probe : measure reference spectrum with probe tip attaches

Horizontal setups : measure reference spectrum with gold reflector placed in liquid sampling cup.

5. ***Measurement of samples***

a. Pour a fresh sample of solvent into a suitable container and record spectra immediately.

Probe : Insert the probe in the container and ensure that the path distance is adequately immersed in the solvent.
Horizontal setups : pour solvent into a suitable container before inserting gold reflector *slowly* to prevent formation of air bubbles.

Set number of scans at 32 for each measurement

Obtain one representative measurement. If more than measurement for a single solvent is necessary, please indicate clearly in the form.

After sample is scanned. press <RETURN> for *Product* which registers the sample as an *Unknown*

b. Record identification result in **Part B** of the **Form I**

1) Correlation coefficient
2) Identity of solvent as indicated from the library i.e. methanol or solvent not

in library

c. Before recording the spectra for the next solvent, clean the probe with tissue. If necessary, remove probe tip

6 ***Save ALL 'master' files back to disk***

7 ***Return of experimental results***
Please ensure that you are enclosing the following back to the Centre:

Form I ☐

Form II/ Printouts of diagnostics test results ☐

Disk containing ALL the master files ☐

Address:

Weng Li Yoon
Centre for Pharmaceutical Analysis
The School of Pharmacy
University of London
29-39 Brunswick Square
London
WC1N 1 AX
Email : wyoon22@yahoo.com

APPENDIX II

Table showing the suppliers of solvents used by participants for the interlaboratory trial in chapter 4.

Solvent	Supplier	Grade (supplier code)
Acetone	Fisher Scientific UK	Certified analytical reagent HPLC
	BDH Analar	HPLC
Acetonitrile	Fisher Scientific UK	HPLC
	Rathburn Chemicals Ltd	HPLC
	Carlo Erba	HPLC
	Romil	Super purity
Butan-1-ol	Fisher Scientific UK	Specified laboratory reagent Certified analytical reagent High purity reagent
	Fisons Ltd	Analytical reagent
Butan-2-ol	Fisher Scientific UK	Certified analytical reagent
	BDH	Analytical reagent HPLC high purity
	Sigma Aldrich	Analytical reagent
Chloroform	Fisher Scientific UK	High purity Certified analytical reagent
	Fisons	HPLC high purity
Dichloromethane	Fisher Scientific UK	Analytical reagent HPLC high purity High purity reagent
	Pancreac	Chemically pure
	BDH	HPLC high purity
Dimethylformamide	Fisher Scientific UK	Analytical reagent
	Sigma Aldrich	HPLC
	BDH	HPLC high purity
	Fisons	Analytical reagent
	Prolabo	Analytical reagent

Ethanol 96 %	Fisher Scientific UK	Certified analytical reagent
	BDH	'USP Analar Grade'
	Hayman Limited	Analytical reagent
	Fluka	Puriss ACS
Absolute alcohol	Fisher Scientific UK	Certified analytical reagent
	Hayman Limited	Analytical reagent
	BDH	HPLC Analytical reagent
Ethyl acetate	Fisons Ltd	HPLC high purity
	Fisher Scientific UK	Laboratory grade reagent
	Sigma Aldrich	HPLC
Methanol	Fisher Scientific UK	HPLC high purity
	Romil Ltd	Super purity grade
	Fluka	HPLC standard
	Promochem	HPLC grade
Propan-1-ol	Fisons Ltd	Specified laboratory reagent HPLC high purity
	Romil Ltd	Super purity grade
	Fisher Scientific UK	Certified analytical reagent
Propan-2-ol	FSA	HPLC high purity
	Fisons Ltd	Specified laboratory reagent HPLC high purity
	Fisher Scientific UK	Certified analytical reagent
	Pancreac	Analytical reagent

Toluene	Fisher Scientific UK	Specified laboratory reagent Certified analytical reagent
	Sigma Adrich	HPLC grade Analytical reagent
	Chromacol Ltd	HPLC grade
	Panreac	Chemically pure
Tetrahydrafuran	Rathburn Chemicals Ltd	HPLC
	Fisher Scientific UK	Laboratory reagent grade High purity reagent Analytical reagent
	Romil Ltd	High purity grade
	Fluka Chemicals	Analytical reagent
	Panreac	Purissimum

Fisher Scientific UK, Leicestershire, UK
Romil Ltd, Cambridge, UK
Rathburns Chemicals Ltd, Peebleshire, UK
Chromacol Ltd, Hertfordshire, UK
Hayman Ltd, Essex, UK
Fluka Chemicals, Dorset, UK
BDH (Merck Ltd), Lutterworth, UK
Panreac, Montcada, Spain
Carlo Erba, Rodano, Italy
Prolabo, Paris, France
Sigma Aldrich, Company Ltd, Dorset, UK
Promochem, Wesel, Germany

APPENDIX III

PUBLICATIONS:

WL Yoon, RD Jee and AC Moffat
The Optimisation of Sample Presentation in the Analysis of Pharmaceutical Excipients using Near Infrared Spectroscopy
Journal of Pharmacy and Pharmacology,Vol 49, Supplement, 165.
(British Pharmaceutical Conference Science Proceedings, 134th Meeting, Scarborough, 15-18th September **1997**)
(Abstract for poster presentation)

W L Yoon, RD Jee and AC Moffat
The Optimisation of Sample Presentation for the Near-Infrared Spectra of Pharmaceutical Excipients
The Analyst, **1998**, May, 123, 1029-1034.

WL Yoon, RD Jee, AC Moffat, DC Lee, PD Blackler and K Yeung
The Transfer of Near Infrared Spectra of Solvents between Different Instruments.
Journal of Pharmacy and Pharmacology, Vol 50, Supplement, 43.
(British Pharmaceutical Conference Science Proceedings, 135th Meeting, Eastbourne, 8-11th September **1998**)
(Abstract for oral presentation).

G Buckton, E Yonemochi, WL Yoon and A C Moffat
Water Sorption and Near Infrared Spectroscopy to Study the Differences Between Microcrystalline Cellulose Before and After Wet Granulation.
International Journal of Pharmaceutics, 181, **1999,** 41 – 47.

WL Yoon, RD Jee, AC Moffat, PD Blackler, Ken Yeung and DC Lee
Construction and Transferabilty of a Spectral Library for the Identification of Common Solvents by Near Infrared Transflectance Spectroscopy.
The Analyst, 124, **1999,** 1197-1203.

WL Yoon, RD Jee and AC Moffat
An Interlaboratory Trial to study the Transferability of a Spectral Library for the Identification of Solvents using Near Infrared Spectroscopy.
Journal of Pharmacy and Pharmacology, Vol 51, Supplement, 71.
(British Pharmaceutical Conference Science Proceedings, 136th Meeting, Cardiff, 13-16th September **1999**)
(Abstract for poster presentation)

WL Yoon, NC North, RD Jee and AC Moffat
Application of Polar Qualification System to the Near Infrared Identification and Qualification of Raw Pharmaceutical Excipients.
Proceedings to the 9th International Conference on Near-Infrared Spectroscopy
Verona, Italy, **1999,** O. 5-3.
(Abstract for oral presentation).

WL Yoon, RD Jee and AC Moffat
A Transferable Method for the Identification of Solvents based on Near-infrared Spectroscopy.
AAPS Pharm. Sci., Vol. 1, No. 4, Abstract no. 2414,
(AAPS Meeting Abstracts, New Orleans, 14-18 Nov 1999)
(Abstract for poster presentation).

Optimisation of sample presentation for the near-infrared spectra of pharmaceutical excipients†

Weng Li Yoon*, Roger D. Jee and Anthony C. Moffat

Centre for Pharmaceutical Analysis, The School of Pharmacy, University of London, 29–39 Brunswick Square, London, UK WC1N 1AX

The effects of sample presentation on near-infrared (NIR) reflectance spectra were examined. Using a Foss NIRSystems Rapid Content Analyzer, which uses sample cups for sample presentation, four important parameters were identified: cup diameter, sample thickness, cup material and packing method. Below a critical diameter of 20 mm, which is dependent on the detector geometry, the spectra became increasingly distorted (*i.e.*, changes in spectral intensities and spectral shape, shifts in peak positions and occurrence of Wood's peak). The minimum sample thickness not to cause spectral distortion was dependent on the physical and chemical nature of the substance. A thickness ≥ 10 mm was found to be adequate for most pharmaceutical excipients. The method of packing was also important. Tapping a powdered sample sometimes caused significant changes ($P < 0.05$) in the spectral absorbance values compared with simply pouring the sample into the sample cup. Standard sample cups made from quartz were to be preferred owing to their lack of background absorptivity. However, the two commercially available flat based vials examined, which were made from soda glass and clear neutral glass, proved to be as suitable for all except applications of the most exacting nature. The spectral distortions resulting from variations in cup diameter, sample thickness and cup material were also shown to alter significantly the values of two commonly used identification algorithms, correlation coefficient (< 0.95) and maximum distance (> 3.0 standard deviation distance), sufficiently to cause misidentifications.

Keywords: *Near-infrared spectroscopy; pharmaceutical excipients; sample presentation; optimisation*

The application of near-infrared (NIR) spectroscopy in the pharmaceutical industry is facing continued growth because of its ease of use, speed of measurement and minimum of sample preparation. Common applications include the identification of raw materials, quality control and quantitative analysis.[1–6] Establishing such procedures involves the expenditure of considerable time and effort to create the necessary calibration sets and libraries of compounds. Nevertheless, it is generally recognised that calibrations for quantification purposes are not always transferable between different instruments, even of the same type and from the same manufacturer.[7–9] It is still to be established if libraries of reflectance spectra for identification purposes are transferable. Factors affecting transferability include instrument variability (ceramic reflectance reference, wavelength accuracy, detector linearity, stray light, radiation source) and possibly sample presentation.[10,11]

The most commonly used sample presentation methods in reflectance NIR use either a fibre optic probe or sample cups. Despite their wide use, there is still little information concerning the necessity to standardise parameters such as cup size, amount of sample, pressure when inserting the probe, *etc.*

Experience gained from the food and agricultural industry suggests that sample presentation is a variable which must not be overlooked. Williams[11] found that cell loading affected the precision of protein determination more than sample grinding. He also reported that variations in bulk density of the sample could lead to errors. Mark and Tunnel[12] reported that variations in packing affected the calibrations which they developed for the measurement of moisture, fat and protein in ground beef, mixed animal feed and breakfast cereal. It was necessary to make multiple measurements on the same sample to average out the variations due to sample presentation.

In this work, the effects of sample presentation when using a sample cup module on the reflectance spectra of some commonly used pharmaceutical excipients were systematically examined. By standardising and eliminating factors responsible for spectral variations, it is hoped that in the near future it will be possible to establish transferable libraries of spectra.

Experimental

Apparatus

A NIRSystems 6500 spectrophotometer (Foss) fitted with a Rapid Content Analyzer (RCA) or a Direct Content Analyzer (DCA) was used for the measurement of all reflectance spectra over the wavelength range 1100–2500 nm. Except where indicated otherwise, the RCA attachment was used for all investigations. Each recorded spectrum was the average of 32 scans.

Samples were measured in flat based cups :

(a) quartz (standard sample cup, 52 mm diameter, Foss catalogue number NR7072);
(b) Pyrex glass (reflectance vessel, 40 mm diameter, Foss catalogue number NR6544);
(c) clear neutral glass (Fbg-Anchor glass vials, 21 mm diameter, catalogue number BDH/Merck/215/0074/23);
(d) soda glass (Philip Harris specimen tubes, 23 mm diameter, catalogue number PH13 T82-528).

Materials

All excipients were of pharmaceutical grade: microcrystalline cellulose (Avicel PH101 and Avicel PH102, FMC, Philadelphia, PA, USA), sodium starch glycollate (Explotab, Edward Mendells), anhydrous dibasic calcium phosphate (A-TAB, Edward Mendells), dibasic calcium phosphate dihydrate (Emcompress, Edward Mendells), lactose monohydrate regular (Broculo Whey Products UK), hydroxypropylmethylcellulose (Methocel E5 Premium, Colorcon), purified talc (Luzenac Europe), propyl and butyl *p*-hydroxybenzoate (Nipa Laboratories) and Kollidon 25 and 30 (Povidone, BASF).

Data treatment

NSAS Version 3 software[13] was used for the calculation of first- and second-derivative spectra using a segment size of 20 data

† Presented at the British Pharmaceutical Conference 1997, 134th meeting, Scarborough, UK, September 15–18, 1997.

points and a gap size of zero data points. Standard normal variate transformations[13] of the absorbance spectra were carried out using an in-house program written in C-language.

The effects of sample presentation were quantified using correlation coefficients and maximum distance.[13] The correlation coefficient (r_{jk}) between the absorbances or mathematically transformed values, x, of two spectra j and k measured at p wavelengths was calculated according to the equation

$$r_{jk} = \frac{\sum_i x_{ij} x_{ik}}{\sqrt{\sum_i x_{ij}^2 \sum_i x_{ik}^2}} \quad (1)$$

Maximum distance (d_{jk}) was calculated using the equation

$$d_{jk} = \max\left[\left|\mathrm{abs}\left(\frac{x_{ip} - \bar{x}_{kp}}{s_{kp}}\right)\right|\right]_{\text{over all } p} \quad (2)$$

where s_{kp} is the inflated standard deviation for the n spectra in set k at wavelength p and given by eqn. 3 (note: the inflated standard deviation was used to allow for uncertainty in the value when n is small).

$$s_{kp} = \left[1 + \frac{1}{\sqrt{2(n-1)}}\right]\left[\frac{\sum_{i=1}^{n}(x_{ikp} - \bar{x}_{kp})^2}{n-1}\right]^{1/2} \quad (3)$$

Results and discussion

Reflectance spectra were found to be affected by sample cup diameter, sample thickness, cup material and packing method.

Cup diameter

This was investigated by placing a quartz sample cup on the adjustable iris diaphragm. This elevated the sample to 1 cm above the sample stage. Spectra were measured over the cup diameter range of 4–50 mm by adjusting the iris diaphragm. Increasing the sample diameter resulted in downward multiplicative shifts of the absorption spectra, with absorbance values at the shorter wavelengths decreasing much more rapidly than at longer wavelengths (*e.g.*, Fig. 1). Peaks also became increasingly well defined, with increasing diameter. All these effects stabilised towards an 'infinite diameter' which was dependent upon detector geometry. For the RCA, this 'infinite diameter' was also found to be a linear function of the elevation from the quartz window. Whilst the value was approximately 28 mm at 1 cm above the sample stage, it was corrected to 20 mm at zero elevation (normal sample position). For the DCA, which has a different detector geometry, the 'infinite diameter' was 35 mm.

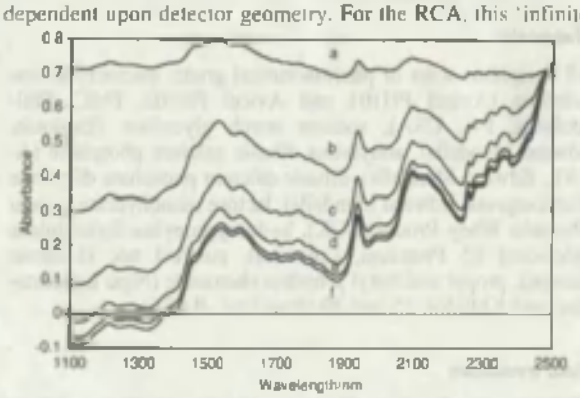


Fig. 1 NIR spectra of lactose monohydrate using different sample cup diameters: a, 4; b, 8; c, 12; d, 16; e, 20; and f, 50 mm.

Below the 'infinite diameter', spectral shape was found to be a function of sample diameter and this was most noticeable with second derivative spectra. Peak amplitudes relative to the most intense peak in the spectrum (normalisation) varied with sample diameter (Fig. 2). For peaks at wavelengths greater than the position of the most intense peak, the ratio generally increased with diameter, while peaks at shorter wavelengths showed the opposite effect. The cause of this effect is not clear, but was observed for all the excipients investigated. As the sample diameter is reduced below the 'infinite diameter', Wood's peak[14] (1520 nm in second derivative absorbance spectra)

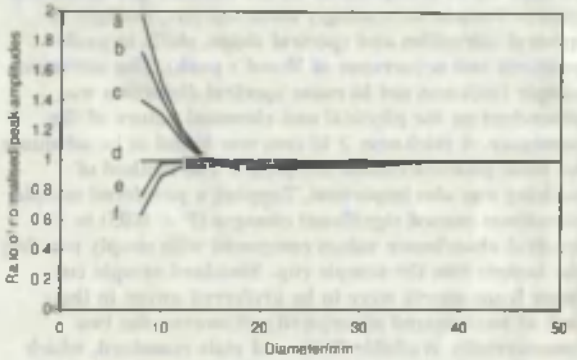


Fig. 2 Effects of varying the sample diameter on the relative peak amplitudes for the second-derivative absorbance spectra of Kollidon 25 at different wavelengths: a, 1372; b, 1430; c, 1695; d, 2274; e, 2374; and f, 2465 nm.

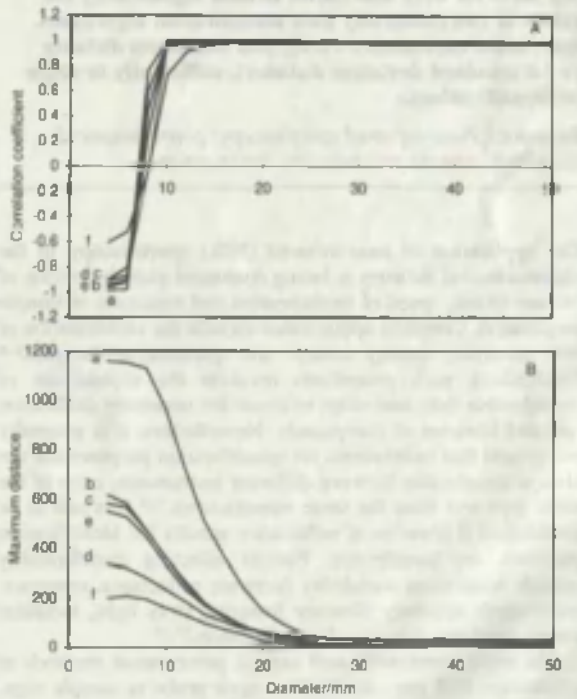


Fig. 3 Effects of varying the sample diameter on A, correlation coefficient and B, maximum distance values, compared with reference spectra measured using a diameter of 50 mm. Excipients : a, Emcompress; b, A-TAB; c, Avicel PH102; d, lactose monohydrate; e, Methocel E5 Premium; and f, purified talc.

appears and becomes increasingly prominent with decreasing diameter. Greater peak position shifts were also observed with spectra obtained with smaller cup diameters. At 8 mm where most peaks become sufficiently defined to be discernible on the second derivative spectra, peak position shifts of up to 5.6 nm were observed, proving to be significant as the wavelength accuracy of the instrument was to be within 0.3 nm. All the spectral distortions mentioned above (*i.e.*, shifts in absorbance values, changes in spectral shape, occurrence of Wood's peak and peak position shifts) could not be compensated for with mathematical treatments such as derivatisation or standard normal variate transformations.

Changes in sample diameter were found to have pronounced effects on the values of the identification algorithms, correlation coefficient and maximum distance, which are commonly used for the identification of excipients in the pharmaceutical industry. This is illustrated in Fig. 3, which shows how r_{jk} and d_{jk} vary with sample diameter using spectra recorded using a sample diameter of 50 mm as reference. Larger diameter cups provide more consistent and well defined spectral information which can enhance the ability to distinguish between closely related substances. Table 1 shows the effects of diameter on the correlation coefficient and maximum distance parameters for three closely related pairs of compounds. Avicel PH101 and 102 differ only in their nominal mean particle size (50 and 100 μm, respectively) and are just distinguishable when using the large diameter cup by the maximum distance parameter. A value of $d_{jk} > 3$ is generally considered to indicate a significant difference.[13] Propyl and butyl *p*-hydroxybenzoate are clearly distinguishable by both r_{jk} and d_{jk} parameters using the larger diameter cups. A value of $r_{jk} < 0.95$ is considered significant.[13] Kollidon 25 and 30 differ in relative molecular mass (30 000 and 50 000, respectively) and again are distinguishable using d_{jk}.

Table 1 Effect of sample diameter on the ability to distinguish between closely related substances using NIR absorbance spectra

Pairs of closely related substance	Mean correlation coefficient[a]		Mean wavelength distance[a]	
	6[b]	52[b]	6[b]	52[b]
Avicel PH101and Avicel PH102	1.000	1.000	2.809	7.934
Propyl and butyl *p*-hydroxy-benzoate	0.996	0.948	8.08	39.338
Povidone 25 and Povidone 30	0.999	0.994	8.368	14.625

[a] The mean spectrum of the first excipient named was used as the reference spectrum. [b] Sample diameter/mm.

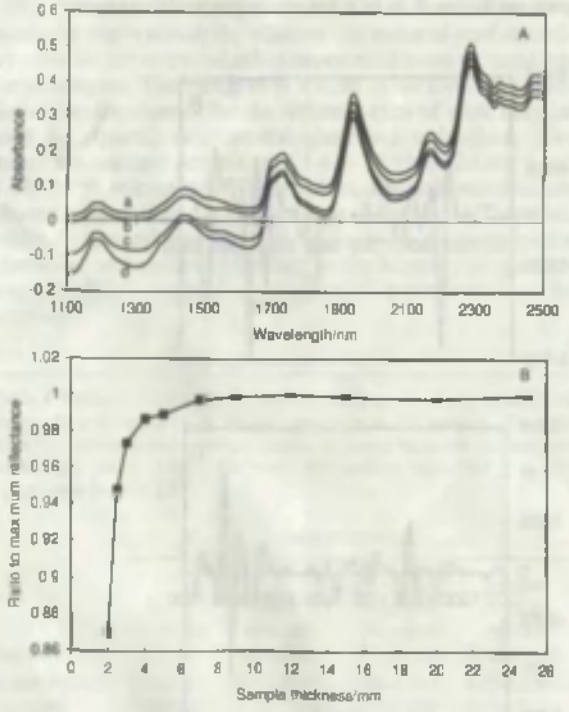


Fig. 4 A, Spectra for Kollidon 25 at different sample thicknesses: **A**, a, 1; b, 2; c, 3; and d, 25 mm. B, dependence of reflectance values at 1100 nm on sample thickness for Kollidon 25. Reference spectra recorded at 25 mm thickness.

Sample thickness

The effects of changing sample thickness were investigated over the range 1–25 mm by weighing increasing masses of sample into a sample cup (clear neutral glass, 21 mm diameter),

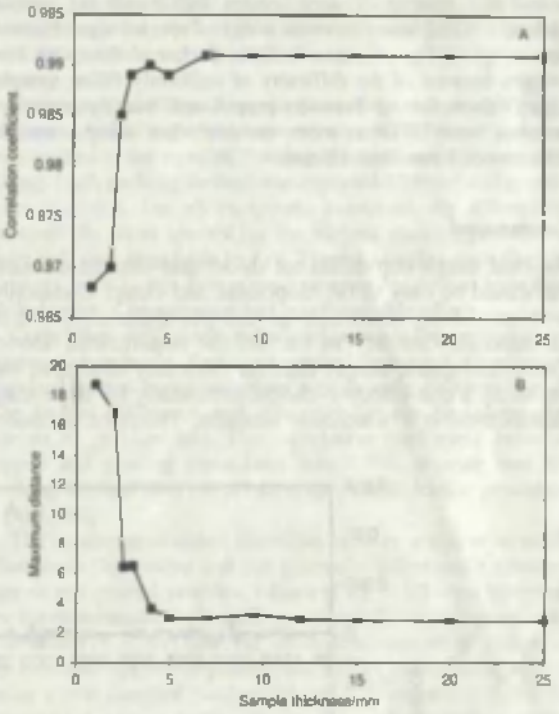


Fig. 5 Effects of sample thickness on A, correlation coefficient and B, maximum distance for Kollidon 25 using absorbance spectra, with a 25 mm thick sample used as the reference spectra.

Table 2 Effect of sample thickness on the ability to distinguish between closely related substances using NIR absorbance spectra

Pairs of closely related substance	Mean correlation coefficient[a]		Mean wavelength distance[a]	
	1[b]	10[b]	1[b]	10[b]
Avicel PH101and Avicel PH102	1.000	1.000	0.576	9.818
Propyl and butyl *p*-hydroxy-benzoate	0.943	0.954	9.037	46.67
Povidone 25 and Povidone 30	0.997	0.996	3.599	7.608

[a] The mean spectrum of the first excipient named was used as the reference spectrum. [b] Sample thickness/mm.

measuring the sample thickness and recording the spectra. The spectral features became increasingly well defined and the absorbance baseline shifted downwards with increasing sample thickness, although the effect was less pronounced than seen with changes of sample diameter [Fig. 4(A)]. Reflectance values became independent of sample thickness above a certain value, the 'infinite thickness' [Fig. 4(B)]. However, unlike the position with sample diameter, the 'infinite thickness' was dependent on the sample material. Both identification algorithms were found to be sensitive to the effects of sample thickness. Fig. 5 shows this for Kollidon 25.

The existence of an 'infinite thickness' has long been recognised and known to be affected by a sample's physical characteristics (*i.e.*, particle size, distribution, shape, bulk density, *etc.*) and chemical nature, *i.e.*, absorptivity.[11,16,17] Attempts to predict the infinite thickness of a sample directly from any measurable physical properties such as bulk density or mean particle size were not successful in this work.

Spectra measured using thicker samples improved the discrimination of closely related excipients (Table 2). Greater differences in spectra were observed for both absorbances and second-derivative absorbance spectra with increasing sample thickness. Also, better reproducibility of spectra was obtained with samples of greater than 'infinite thickness' than with thin samples because of the difficulty of uniformly filling sample cups. Values for r_{jk} between propyl and butyl *p*-hydroxybenzoate were 10 times more variable when using a sample thickness of 1 mm than 10 mm.

Cup material

The ideal sample cup should not absorb near-infrared radiation and should be easy to fill, disposable and cheap. Commonly used materials are quartz and various types of glass, but none of the materials currently in use fits the requirements above. Customized quartz sample cups are minimally absorptive but are hardly a cost-effective choice, particularly for large scale identification in a warehouse situation. Therefore, commercially available glass vials were examined to determine their fitness for use in NIR applications.

Fig. 6 shows the second-derivative absorbance spectra for a range of cups as measured by transflectance (*i.e.*, by placing the ceramic reflectance reference over an empty cup). Quartz showed the least absorbance, followed by soda glass, clear

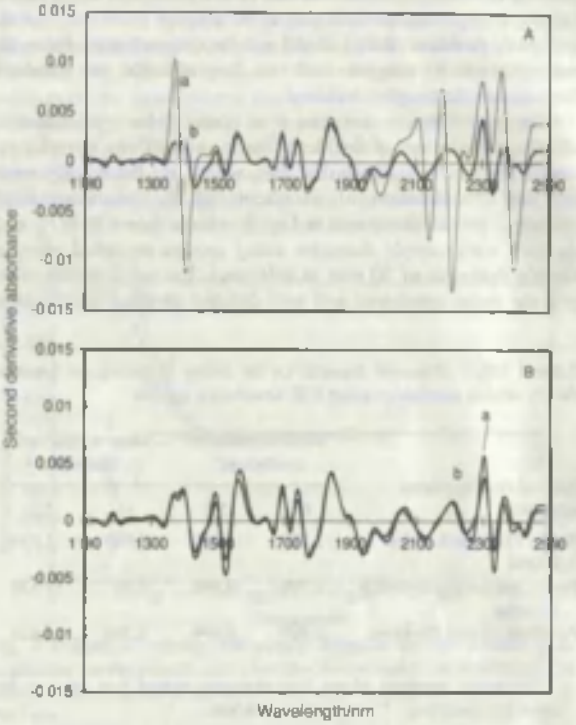


Fig. 7 A, Second-derivative spectra of A-TAB in a, Pyrex glass and b, quartz sample cup. B, as in A, but after subtraction of cup spectra.

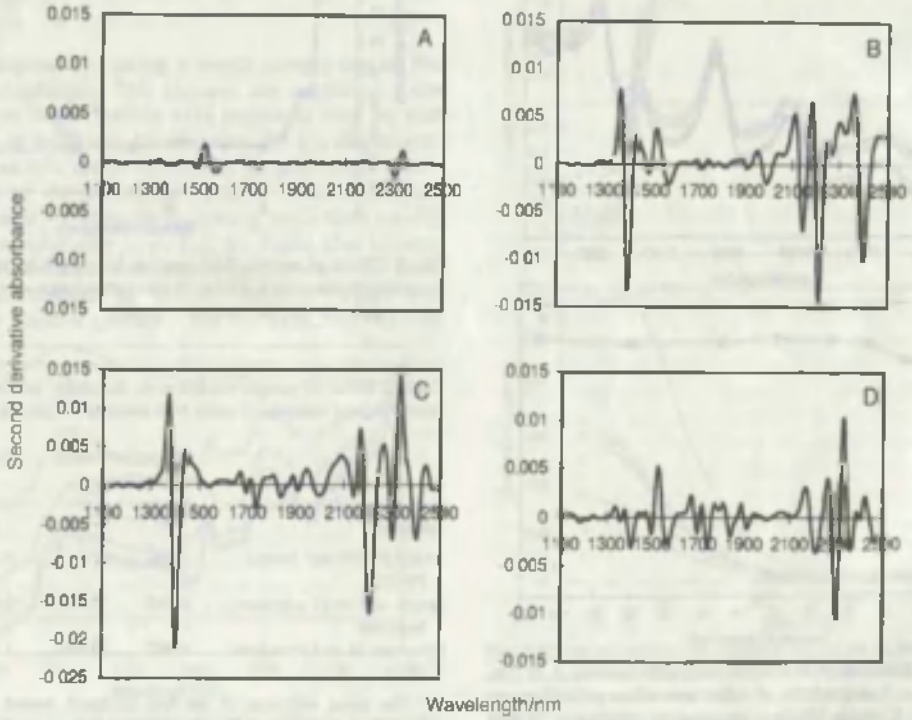


Fig. 6 Second-derivative absorbance *versus* wavelength for A, quartz, B, Pyrex glass, C, clear neutral glass and D, soda glass.

Table 3 Effect of cup material on identification parameters, r_μ and d_μ. Spectra measured in the quartz cup taken as the reference. Values are means of six spectra

	A-TAB				Purified talc			
	Quartz	Pyrex glass	CNG*	Soda glass	Quartz	Pyrex glass	CNG*	Soda glass
Absorbance spectra—								
r_μ	1	0.995	0.999	0.998	1	0.963	0.992	0.995
d_μ	0.870	4.881	3.236	2.037	0.963	44.637	22.251	13.058
Second-derivative absorbance spectra—								
r_μ	1	0.416	0.836	0.939	10.999	1	1	
d_μ	1.494	332.019	86.873	35.462	1.485	112.802	48.112	30.423

* Clear neutral glass

neutral glass and Pyrex glass, in that order. The peaks in glass at approximately 1400 and 2200 nm can be assigned to the O–H first overtone bands from the SiOH and also C=O forming combination bands, possibly from the carbonates of calcium and sodium.[13]

The spectrum of the cup material was found to be additively superimposed upon the spectrum of the sample and can therefore cause serious distortion of the sample spectrum for poorly absorbing materials. Fig. 7(A) clearly shows this for the spectrum of A-TAB measured in quartz and Pyrex glass cups. Table 3 illustrates the effect of such distortion on correlation coefficients and maximum distances. It is also obvious that more strongly absorbing excipients are affected to a lesser extent.

Although it was possible to compensate for the cup spectrum by subtraction [Fig. 7(B)], the success was very much dependent upon obtaining a representative spectrum for the empty cup. This is difficult as the spectrum obtained by transflectance is dependent upon the height at which the ceramic reflectance reference is placed above the cup base. Generally, this is set by the physical dimensions of the cup and it cannot be placed in direct contact with the cup base as required.

Apart from establishing the necessity to standardise on a particular cup material, the effect of cup material reproducibility could be just as crucial as it is impractical to use the same cup for all samples. The background spectra in the form of second-derivative absorbance, for six different cups of each material used *i.e.*, quartz, clear neutral glass and soda glass, gave maximum standard deviations of 1.4×10^{-4}, 8.92×10^{-4} and 8.48×10^{-4}, respectively. These values, although minimal, are above the noise level of the instrument (2.0×10^{-5}). Therefore, whilst the spectral reproducibility of glass vials is acceptable for identification purposes, depending on the accuracy of quantitative measurements required, the use of quartz cups may be necessary.

Table 4 Student's t values for the two-sampled t-test for the comparison of mean values of tapped ($n = 10$) and poured ($n = 10$) samples. The table gives the maximum and minimum values of t observed across the complete wavelength range (1100–2500 nm). The critical value for t at 5% significance level is 2.1

	Data treatment			
Excipient	Absorbance spectra	First-derivative absorbance	Second-derivative absorbance	Standard normal variate absorbance
Avicel PH102	0.02–4.87	0.67–10.55	0.03–11.43	0.03–12.86
Explotab	3.09–4.86	0–15.76	0–25.14	0.01–11.45
A-TAB	0–2.84	0.04–11.18	0–11.57	0.11–10.22
Emcompress	1.35–3.04	0.08–8.07	0–5.67	0–5.94
Purified talc	0.08–8.26	0.6–327.65	0–282.13	0–100.63

Packing method

Powder packing has been recognised as a source of random variation which can result in small spectral shifts.[12] However, it is questionable whether the packing method affects the spectra systematically. For example, tapping a powdered sample can cause stratification of the sample, giving a greater density at the bottom of the sample and hence a greater reflectance. Recognising that such effects were important, three sample packing methods were examined : tapping, compression and pouring.

Tapping entailed knocking the base of the sample cup gently 10 times after filling. For compression, a 5000 N m^{-2} pressure was applied to the powder. Pouring involved no treatment after filling. Each packing method was repeated 10 times and spectra were recorded. For all excipients examined, the differences between the mean spectra for the various packing procedures were only just detectable by eye. Tapped samples gave slightly stronger reflectances than poured samples because of increased bulk density. Compression had no observable effect.

Correlation coefficients calculated between the various mean spectra (absorbance, first- and second-derivative absorbance, standard normal variate) were not significantly different from 1. The largest difference was observed for the first-derivative spectra of purified talc. The correlation coefficient between tapped and pouring procedures was 0.994, proving that the packing method does not affect simple identification processes significantly.

The maximum distance algorithm is more sensitive to small changes in the spectra and can generally differentiate between tapped and poured samples. Values of $d_\mu > 3.0$ were observed for the mathematically treated spectra of all samples examined. The values of d_μ are, however, dependent upon what is taken as the reference (tapped or poured) and a more general comparison using a two sampled Student's t-test is presented in Table 4. Student's t values were calculated at corresponding wavelengths across the whole wavelength range. Significant differences ($P < 0.05$) between the mean values for pouring and tapping were observed across the majority of the spectra with absorbance and mathematically treated spectra. Purified talc showed gross differences across most of the spectrum. Purified talc has plate-like-shaped particles and tapping can cause considerable reorientation of the particles, resulting in differences in light–particle interactions.

Conclusion

The work in this paper has clearly shown that sample presentation can have a significant effect on the near-infrared reflectance spectrum of a substance. As the requirement on the accuracy and precision of a spectrum is dependent on a particular application, it is important to evaluate the sample presentation effects as part of any methodology development. In all applications, sample diameter, thickness and cup material are important parameters and can affect even the simplest NIR

applications such as sample identification. Ideally, sample diameter and thickness should exceed their 'infinite values' or at least be standardised. Spectral differences arising from packing variations are generally less significant, however, in more exacting applications such as qualification and quantitative analysis, such subtle differences can be crucial.

The authors thank SmithKline Beecham Pharmaceuticals for financial support and providing the samples, and Foss NIRSystems for the loan of the NIRSystems 6500 spectrophotometer. They also thank Nigel North of SmithKline Beecham Pharmaceuticals for suggestions for the project and Sheelagh Halsey of Foss NIRSystems for helpful discussions.

References

1 Dubois, P., Martinez, J., and Levillain, P., *Analyst*, 1987, **112**, 1675.
2 Corti, P., Dreassi, E., Corbini, G., Montecchi, L., and Paggi, J., *Analusis*, 1990, **18**, 117.
3 Plugge, W., and Van Der Vlies, C., *J. Pharm. Biomed. Anal.*, 1992, **10**, 793.
4 Gemperline, P. J., and Boyer, N. R., *Anal. Chem.*, 1995, **67**, 160.
5 Dreassi, E., Cerimelli, G., Savini, L., Corti, P., Perruccio, P. L., and Lonardi, S., *Analyst*, 1995, **120**, 319.
6 Kirsch, J., and Drennen, J., *Pharm. Res.*, 1996, **13**, 234.
7 Blank, T., Sum, S., Brown, S., and Monfre, S., *Anal. Chem.*, 1996, **68**, 2987.
8 Bouveresse, E., Hartman, C., Massart, D. L., Last, I. R., and Prebble, K. A., *Anal. Chem.*, 1996, **68**, 982.
9 Wang, Y., and Kowalski, B., *Appl. Spectrosc.*, 1992, **46**, 764.
10 Dardenne, P., Biston, R., and Sinnaeve, G., in *Near Infra-red Spectroscopy: Bridging the Gap between Data Analysis and Near Infra-red Spectroscopy Applications*, ed. Hildrum, K. I., Isaksson, T., Naes, T., and Tandberg, A., Ellis Horwood, Chichester, 1992, pp. 453–458.
11 Williams, P. C., in *Handbook of Near-Infrared Analysis*, ed. Burns, D., and Ciurczak, E., Marcel Dekker, NewYork, 1992, pp. 307–311.
12 Mark, H. L., and Tunnel, D., *Anal. Chem.*, 1985, **57**, 1449.
13 *NSAS Training Manual*, NIRSystems, Silver Spring, MD.
14 Wood, R. W., *Philos. Mag.*, 1902, **4**, 396.
15 Barnes, R. J., Dhanoa, M. S., and Lister, S. J., *Appl. Spectrosc.*, 1989, **43**, 772.
16 Olinger, J. M., and Griffiths, P. R., *Appl. Spectrosc.*, 1993, **47**, 687.
17 Olinger, J. M., and Griffiths, P. R., *Appl. Spectrosc.*, 1993, **47**, 695.

Paper 8/00358K
Received January 13, 1998
Accepted March 3, 1998

Construction and transferability of a spectral library for the identification of common solvents by near-infrared transflectance spectroscopy

Weng Li Yoon,*[a] Roger D. Jee,[a] Anthony C. Moffat,[a] Paul D. Blackler,[b] Ken Yeung[b] and David C. Lee[b]

[a] *Centre for Pharmaceutical Analysis, The School of Pharmacy, University of London, 29–39 Brunswick Square, London, UK WC1N 1AX*
[b] *SmithKline Beecham Pharmaceuticals, New Frontiers Science Park (North), Harlow, Essex, UK CM19 5AW*

Received 28th April 1999, Accepted 8th June 1999

A transflectance near-infrared spectroscopic method has been developed for the identification of 15 common solvents using correlation in wavelength space as the identification algorithm. Second-derivative absorbance spectra over the wavelength range 1136–2000 nm were found to give the optimum conditions for distinguishing between the solvents. The spectral library was tested on eight instrumental setups and found to be directly transferable between different instruments not only from the same manufacturer, but also between grating and Fourier transform systems. Identification was not affected by small changes in temperature or optical path length. The presence of water could easily be detected by visual insepection of the solvent spectra. However, small traces of water did not normally interfere with the identification process. Correlation coefficient values between a given solvent from different batches and/or between different instruments were generally >0.99. Values <0.99 invariably indicated the presence of impurities. Mixtures of similar solvents such as ethanol–methanol could not always be reliably differentiated from that of the pure solvent at the highest concentration.

Introduction

Checking the identity and quality of raw materials in the pharmaceutical industry is an everyday analytical problem and is subject to the requirements of Good Manufacturing Practice.[1] Near-infrared (NIR) reflectance spectroscopy can offer a simple solution to this problem because of its ease of use, speed of measurement and non-destructive nature.

The use of NIR reflectance spectroscopy for identification purposes requires the setting up of a library of spectra containing all the materials of interest and preferably those of other closely related substances. The setting up and evaluation of such a library are time consuming and costly procedures. It is therefore highly desirable that a library, once established, should be transferable between different instruments. Nevertheless, it is generally recognised that calibrations for quantification purposes are not always transferable between different instruments, even of the same type and from the same manufacturer.[2–6] Differences in reflectance spectra between instruments can arise from many factors; wavelength accuracy,[7] detector linearity,[8] bandpass, stray light, sample presentation[9] and, most importantly, the reflectance standard used.[10,11] Varying degrees of success in transferring data have been reported. Jones *et al.*[12] successfully transferred a calibration for moisture determination between instruments of the same model using a simple slope/bias correction. However, in most other cases reported in the literature, various more complex standardisation approaches have been developed to correct for instrumental differences.[2–6] It is still to be established if libraries of reflectance spectra for identification purposes are transferable.

The problems with differences in reflectance standards are largely avoided in transflectance measurements on liquid samples as spectra may be referenced to air. In a transflectance measurement the radiation passes through the sample to a reflector and back (Fig. 1), the optical pathlength being approximately twice the distance from the point of entry of the radiation to the reflector. Typically the reflector is made from stainless steel, gold plate, PTFE or another inert reflecting surface.

In this work, a library of spectra of 15 commonly used solvents was constructed and the optimum wavelength range, mathematical processing, *etc.*, were evaluated for distinguishing between the solvents. The transferability of the spectral library was tested using eight different instrumental setups.

To compare spectra, correlation in wavelength space was used. This simple method has the advantage of depending only on the shape of the spectra and not on the absolute magnitude of the response and consequently should be insensitive to

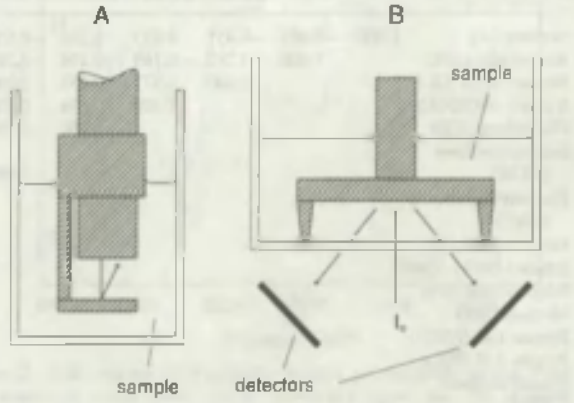


Fig. 1 Schematic optical diagram for transflectance measurements using (A) a fibre-optic probe and (B) a sample cup and reflector (not to scale).

pathlength changes. The dot product correlation coefficient, r is given by

$$r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} \quad (1)$$

where x_i and y_i are the ordinate values of the two spectra being compared at wavelength i, the summations being performed over the whole or a selected part of the spectra. An r value of 1 represents a perfect match; however, in practice r is always less than 1 and it is desirable to find a value for r above which correct identification occurs. The correlation values for pairs of different solvents should fall below this value. When trying to identify substances there will always be some probability that false positive and false negative identifications will occur.[13] By optimising the spectral wavelength range used and the mathematical pre-treatment of the spectra, it should be possible to minimise these errors.

Experimental

Materials

All solvents were of 99% or greater purity and used as obtained directly from the suppliers. The following solvents were obtained from Fisher Scientific UK (Loughborough, Leicestershire, UK): acetone (specific laboratory reagent), butan-1-ol (analytical reagent), butan-2-ol (analytical reagent), chloroform, dichloromethane, absolute ethanol, ethanol (96%), ethyl acetate, ethyl methyl ketone, industrial methylated spirit (64 OP and 74 OP), methanol, propan-1-ol (specific laboratory reagent), propan-2-ol, tetrahydrofuran (specific laboratory reagent) and toluene (analytical reagent). Additional samples and other solvents were obtained as follows: acetonitrile and propan-2-ol (Romil Ltd., Cambridge, UK, super purity grade), chloroform (Spectrosol grade), dichloroethane, heptane, isopentane and pentan-1-ol (BDH Merck Ltd., Lutterworth, UK, Spectrosol), benzyl alcohol (puriss), dichloromethane and dimethylformamide (ACS grade for UV spectroscopy) (Fluka Chemicals, Gillingham, Dorset, UK), absolute ethanol (Hayman Ltd., Witham, Essex, UK), octane and nonane (Avocado Research Chemicals, Heysham, Lancashire, UK), decane, pentane and undecane (Aldrich Chemicals, Gillingham, Dorset, UK) and hexane (HPLC grade) (Rathburn Chemicals Ltd., Walkerburn, Peebleshire, Scotland, UK).

Instrumentation

Eight different near-infrared instrumental setups were used from four different laboratories. Setups 1–6 were all based on FOSS NIRSystems (Silver Springs, MD, USA) instrumentation.

Setup 1: 6500 scanning spectrophotometer fitted with an Interactance immersion probe (NR 6770). Setup 2: 6500 scanning spectrophotometer fitted with an interactance immersion probe. Setup 3: 6500 scanning spectrophotometer fitted with a Smartprobe (NR 6770). Setup 4: 5000 scanning spectrophotometer fitted with a Smartprobe. Setup 5: 6500 scanning spectrophotometer (same monochromator as used in setup 3) fitted with a Rapid Content Analyser (AP 1300) and liquid sampling kit (NR 6544) comprising a Pyrex glass cell (40 mm diameter) and gold reflector (2 × 0.5 mm optical pathlength). Setup 6: 6500 scanning spectrophotometer fitted with an interactance reflectance probe (NR 6775). Setup 7: Bran + Luebbe (Norderstedt, Germany) Infraprover II Fourier transform polarisation interferometer (124-A020-01) fitted with a diffuse reflectance fibre-optic probe with transflectance cap (124-B603-02). Setup 8: Buhler (Uzwil, Switzerland) FT-NIR NIRVIS spectrophotometer (Model 100.1) fitted with a Buhler fibre-optic probe (Model 1359) and transflectance cap. All FOSS probes were fitted with a transflectance probe tip.

Measurement of spectra

All spectra were measured by transflectance with air as reference. For all probes the distance between the probe tip and the reflector was set using a 1 mm gauge. Solvents were poured into small sample bottles and the probe with transflectance probe tip was immersed. The probe and reflector were dismantled and cleaned between each measurement. With setups 1–6 each recorded spectrum was the average of 32 scans and measured over the range 1100–2498 nm (2 nm intervals, 700 data points). For setup 7 the recorded spectra were the average of six scans and measured over the range 4500–9996 cm^{-1} (12 cm^{-1} intervals, 459 data points). Spectra for setup 8 were measured over the range 4000–9996 cm^{-1} (12 cm^{-1} intervals, 500 data points) and each recorded spectrum was the average of six scans.

Data treatment

All spectra were exported from the manufacturers' software used to run the instruments as JCAMP files. Dot product

Table 1 Correlation coefficients for spectral library measured using setup 1: second-derivative spectra (nine data point block size), wavelength range 1136–2000 nm

Solvent	A	AN	B1OL	B2OL	CH	DCM	DMF	E	E96	EA	M	P1OL	P2OL	THF	TOL
Acetone (A)	1.000	−0.001	−0.451	0.075	0.292	−0.057	0.009	0.403	0.330	0.573	0.171	−0.238	0.372	0.227	0.573
Acetonitrile (AN)		1.000	0.212	−0.165	−0.494	−0.389	−0.142	−0.178	−0.168	0.611	0.039	−0.034	−0.278	0.213	−0.068
Butan-1-ol (B1OL)			1.000	0.577	−0.053	0.290	0.500	0.406	0.343	−0.183	0.606	0.894	0.199	0.565	−0.424
Butan-2-ol (B2OL)				1.000	0.674	0.750	0.807	0.890	0.742	0.089	0.811	0.839	0.894	0.494	0.284
Chloroform (CH)					1.000	0.794	0.583	0.651	0.506	0.037	0.473	0.329	0.774	0.036	0.617
Dichloromethane (DCM)						1.000	0.645	0.608	0.501	−0.163	0.497	0.546	0.684	0.180	0.255
Dimethylformamide (DMF)							1.000	0.701	0.667	0.071	0.680	0.752	0.716	0.493	0.343
Ethanol (abs.) (E)								1.000	0.853	0.258	0.853	0.699	0.903	0.580	0.346
Ethanol (96%) (E96)									1.000	0.251	0.651	0.591	0.747	0.457	0.288
Ethyl acetate (EA)										1.000	0.146	−0.143	0.237	0.213	0.496
Methanol (M)											1.000	0.796	0.705	0.586	0.189
Propan-1-ol (P1OL)												1.000	0.560	0.596	−0.109
Propan-2-ol (P2OL)													1.000	0.406	0.514
Tetrahydrofuran (THF)														1.000	−0.204
Toluene (TOL)															1.000

correlation coefficients [eqn. (1)] were calculated using a simple in-house written computer program.

Derivative spectra were calculated by simple difference procedures using an in-house written computer program. Equal sized blocks of data points before and after the data point at which the derivative is to be calculated were averaged and the derivative was taken as the simple difference in means. The process was repeated across the whole spectrum a number of times until the required derivative was obtained. With derivative spectra some data points are lost at each end of the spectrum and this was equal to the block size times the derivative used.

Cubic spline interpolation was used to convert spectra measured in wavenumbers to equally spaced data points in terms of wavelength. The wavenumber *vs.* transmittance spectra were first converted to absorbance spectra and then absorbance values at 2 nm intervals over the range 1100–2498 nm to give 700 calculated values. Out of range values were set to zero. Cubic spline interpolation was performed using an in-house written computer program based on the functions *spline* and *splint*.[14] Simulations of bandpass effects were performed using an in-house computer program based on the function *convlv*.[14]

Results and discussion

The spectra of all 15 solvents (Table 1) were obtained for all eight setups with the exception of setup 4, for which the ethanol (96%) spectrum was not available. All spectra were measured with respect to air as the reference. In general, different batches of solvents were used with each setup. The correct identity of the solvents used with setup 1 were also confirmed by recording their mid-IR spectra and comparing them with reference spectra.[15]

The instrumental setups varied widely: 1–6 were grating instruments and 7 and 8 were FT instruments. Setups 1 and 2 used an Interactance probe designed for liquids, setup 6 used an Interactance reflectance probe designed for highly scattering solutions and slurries, setups 3 and 4 used Smartprobes designed for raw materials (powders) and setup 5 used a sample cup and reflector. Fig. 2 shows the spectra of dimethylformamide (DMF) as measured on all eight setups. DMF, like most of the solvents studied, exhibited strong absorption bands above about 2200 nm. Useful spectral information above this wavelength was not generally obtained when using a fibre-optic probe. The probe used with the Bran+Luebbe instrument restricted the wavenumber range to above 4500 cm^{-1} (*i.e.* < 2222 nm). Original spectra from both the FT instruments were in terms of transmittance *vs.* wavenumber. After conversion to absorbance, cubic spline interpolation was used to produce equally spaced wavelength data so as to match the spectra measured on the other instrumental setups. The original wavenumber and converted spectra are shown in Fig. 2(B) and (A), respectively. Any transferable library clearly needs to take these considerations into account. It was decided to use the spectra from setup 1 as the reference set and use these to find the optimum conditions for distinguishing between the solvents.

Three parameters were investigated to find the optimum conditions: (1) wavelength range, (2) derivative and (3) data smoothing. Optimum parameters were taken as those which showed the best discrimination between the solvents as measured by the dot product correlation coefficient. Correlation coefficient values were calculated for all the combinations of different pairs of solvents in the library. Data smoothing and taking derivatives while decreasing the information content of a spectrum can, however, increase the ability to distinguish between different spectra.

The correlation coefficients between the original absorbance spectra of different solvents (*e.g.*, 0.997 for butan-1-ol–propan-1-ol) were often very high and not suitable for identification purposes. A considerable improvement was obtained using first-derivative spectra where the highest r value between different solvents was 0.917. The best results were obtained using second-derivative spectra calculated with a block size of nine data points over the wavelength range 1136–2000 nm. The full correlation matrix is shown in Table 1, from which it can be seen that the highest correlation between spectra of different solvents was 0.903. Third- and fourth-derivative spectra gave almost equivalent results, with slightly lower r values, but with the r values concentrated more towards higher values than when using the second-derivative spectra. The same optimum conditions were obtained using the spectra from setup 5, which used a sample cup and reflector and gave usable spectra over the full wavelength range. Information in the 2000–2498 nm range, far from enhancing the ability to distinguish between solvents, actually decreased it.

A comparison of all different possible pairs of the 119 spectra measured on all eight setups was carried out. This gives 7021 different possible pairs of which 413 are between the same solvents but measured on different instruments. The correlation coefficients fall into two distinct groups with a clear gap between them (Table 2). There was no overlap between mismatches and correct matches. The highest r value for a mismatch was 0.931 between ethanol (setup 6) and ethanol 96%

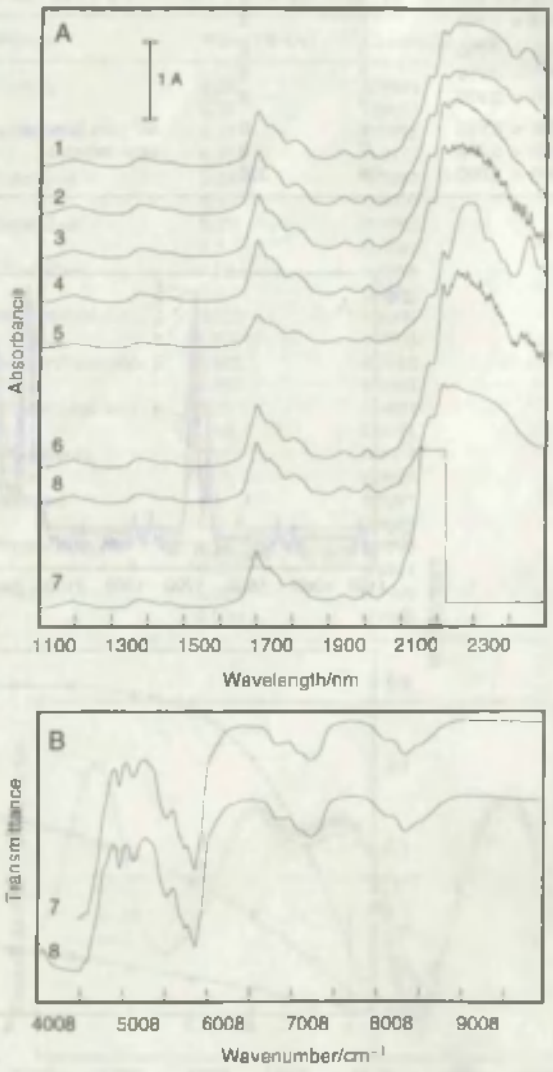


Fig. 2 NIR spectra of dimethylformamide as measured on the eight instrumental setups. (A) Linear wavelength scale and (B) original wavenumber scale spectra for setups 7 and 8. Numbers indicate instrumental setup. Spectra have been vertically displaced for clarity.

(setup 8) and the lowest r for a correct match was 0.978 between dichloromethane samples on setups 2 and 7. Restricting the spectra to only those measured on the FOSS setups gave the highest r value between different solvents as 0.917 between ethanol (setup 6) and 96% ethanol (setup 3) and the lowest r value between the same solvent as 0.991 (ethanol, setups 6 and 4). This improvement is hardly surprising since the spectra from the FT instruments had had to undergo wavenumber to wavelength conversion. There is no indication that spectra differed in quality between any of the setups. Clearly an r value of >0.97 strongly suggests a positive identification although any value below 0.99 may indicate a difference in purity and should be treated with caution.

It should be noted that using the original wavenumber-scaled spectra from setups 7 and 8, that is, the two FT NIR instruments, the optimum conditions for distinguishing between the solvents are different, namely first-derivative absorbance spectra calculated using a one data point block size and over the range 5000–9996 cm^{-1} (2000–1000 nm). The highest r value between different solvents was 0.887 (butan-2-ol–propan-2-ol) and is lower than for the wavelength-scaled data.

The effect of water vapour on the air reference spectrum was investigated by recording the spectrum of air dried with silica gel and air saturated with water. The effects were very small and unlikely to cause the r value to change by more than 0.0005. Humidity changes are therefore unlikely to be important.

Table 2 Distribution of correlation coefficients calculated between all different pairs of spectra (library solvents) measured on the eight setups: second-derivative absorbance spectra (nine data point block size), wavelength range 1136–2000 nm

Correlation coefficient range	Number of pairs		Notes
	Setups 1–8	Setups 1–6	
<0.900	6535	3653	All pairs between different solvents
0.900 < 0.910	41	37	
0.910 < 0.920	25	6	
0.920 < 0.930	6	0	
0.930 < 0.940	1	0	
0.940 < 0.950	0	0	
0.950 < 0.960	0	0	
0.960 < 0.970	0	0	
0.970 < 0.980	1	0	All pairs between same solvents
0.980 < 0.990	7	0	
0.990 < 1.000	405	220	

Table 3 Effect of wavelength errors on the correlation coefficient. Whole spectrum shifted to lower wavelengths by specified amount. Second-derivative absorbance spectra (nine data point block size), wavelength range 1136–2000 nm. All r values with respect to unshifted spectrum

Solvent	Wavelength shift/nm	Correlation coefficient, r
Butan-1-ol	0.5	0.999
	1.0	0.997
	2.0	0.987
	3.0	0.970
Dichloromethane	0.5	0.998
	1.0	0.994
	2.0	0.976
	3.0	0.946
Toluene	0.5	0.999
	1.0	0.996
	2.0	0.984
	3.0	0.964

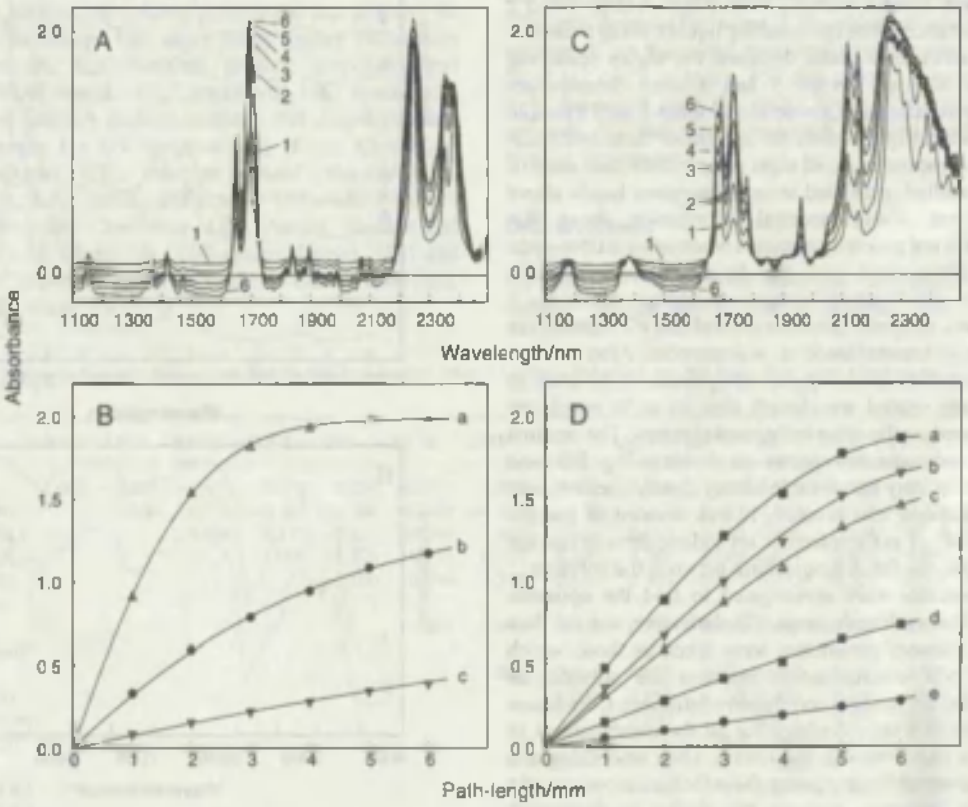


Fig. 3 Effect of pathlength. (A) Absorbance spectra for dichloromethane. (B) Absorbance *vs.* pathlength plots: (a) 1692, (b) 1648 and (c) 1840 nm. Absorbances measured with respect to baseline at 1578 nm. (C) Absorbance spectra for acetone. (D) Absorbance *vs.* pathlength plots: (a) 2118, (b) 1732, (c) 1692, (d) 1962 and (e) 1378 nm. Absorbances measured with respect to baseline at 1268 nm. Optical pathlengths: (1) 1, (2) 2, (3) 3, (4) 4, (5) 5 and (6) 6 mm. All spectra measured using setup 3.

Effect of wavelength accuracy and bandpass

Errors in wavelength calibration between different instruments will cause the r value to decrease for a given solvent. The effect of wavelength errors was simulated for three solvents by shifting the whole spectrum by small amounts using cubic spline interpolation. Solvent examples were chosen so as to cover both broad and sharp spectra. The results are shown in Table 3, from which it can be seen that wavelength shifts up to 1 nm cause little problem, the r values remaining >0.99. This is within the tolerance specified for the FOSS monochromators of ±0.3 nm and also the FT instruments (±0.2 nm at 1000 nm to ±1.2 nm at 2500 nm). Examining the six most intense negative peaks in the second-derivative spectra over all six FOSS setups and 15 solvents indicated a mean maximum difference of 0.65 nm. Over all eight setups the mean maximum difference was 1.1 nm. The largest difference over all setups was 4.73 nm for the small poorly defined water (impurity) peak at −1940 nm in DMF. Clearly the effects of wavelength errors are small.

Differences in bandpass between monochromators can also be expected to affect r values. Simulation again indicated that small changes in bandpass of ±2 nm would not cause a problem with r remaining >0.99. This is within the tolerance set by FOSS for their monochromators of −1 to +2 nm.

Effect of pathlength

If the Beer–Lambert law holds, identification using correlation in wavelength space should not be affected by pathlength changes. This was investigated by measuring the spectra of acetone, dichloromethane and methanol over the optical pathlength range 1–6 mm using the Smartprobe (setup 3). Fig. 3 shows spectra and absorbance *vs.* pathlength plots for acetone and dichloromethane. Although the air reference spectrum was measured at the corresponding pathlengths, considerable baseline shifts are observed and this presumably arises because of refractive index differences between the solvent and air. Only for small peaks were reasonably straight-line plots of absorbance *vs.* pathlength obtained. Second-derivative absorbance *vs.* pathlength plots showed similar curvature. With too large a pathlength it is easy to exceed the dynamic range of the instruments, which varied considerably over the eight setups. The maximum measurable absorbance was approximately 1.9

Table 3 Effect of pathlength on the correlation coefficient. Second-derivative absorbance spectra (nine data point block size), wavelength range 1136–2000 nm. All r values with reference to the 2 mm optical pathlength spectra. Measured on setup 3

	Optical pathlength/mm					
Solvent	1	2	3	4	5	6
Acetone	0.9998	1.0000	0.9998	0.9994	0.9985	0.9970
Acetone (air ref. fixed at 2 mm)	0.9998	1.0000	0.9998	0.9995	0.9982	0.9969
Dichloromethane	0.9969	1.0000	0.9954	0.9799	0.9530	0.9283
Methanol	0.9999	1.0000	0.9991	0.9950	0.9464	0.8796

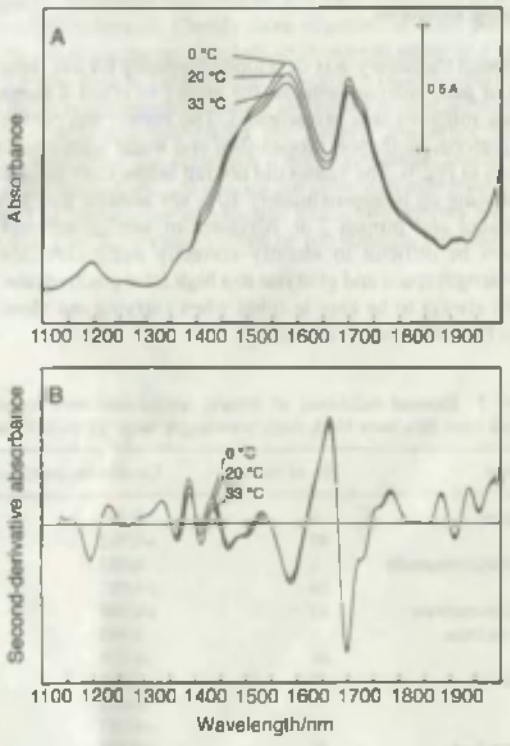


Fig. 4 Effect of temperature on the NIR spectrum of methanol. (A) Absorbance spectra and (B) second-derivative spectra (nine data point block size).

Table 5 Effect of water on the correlation coefficient. Solvent before adding water taken as reference. Second-derivative absorbance spectra (nine data point block size), wavelength range 1136–2000 nm

Solvent	Water (% v/v)	Correlation coefficient, r
Acetone	0.25	0.9904
	0.50	0.9602
Acetonitrile	0.15	0.9978
	0.30	0.9913
Butan-1-ol	0.25	0.9997
	0.50	0.9990
Butan-2-ol	0.25	0.9995
	0.5	0.9982
Chloroform	0.05	0.9998
	0.1	0.9996
Dichloromethane	0.025	0.9994
	0.050	0.9995
Dimethylformamide	0.025	0.9980
	0.050	0.9947
Ethanol (absolute)	0.25	0.9995
	0.50	0.9981
Ethyl acetate	0.01	0.9964
	0.02	0.9923
Methanol	0.1	0.9997
	0.2	0.9980
Propan-1-ol	0.25	0.9996
	0.50	0.9981
Toluene	0.025	0.9999
	0.050	0.9999

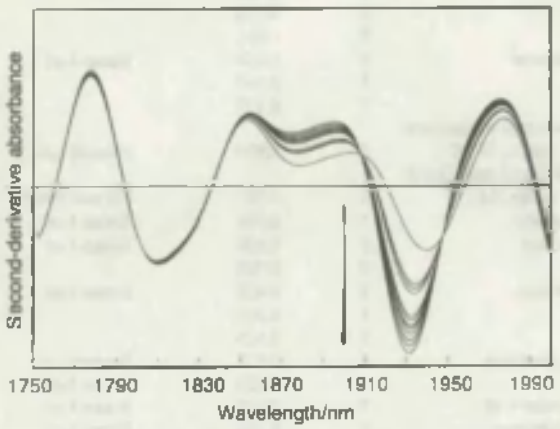


Fig. 5 Uptake of water by dimethylformamide over a period of about 10 min. The direction of the arrow indicates increasing time.

for setups 1, 2, 4 and 5, 2.4 for setup 3 and 2.9 for setup 6. Both FT instruments were software limited to absorbance values of ≤2.0. Even where the dynamic range of the instrument was not exceeded, non-linearity is clearly seen. Small changes in pathlength can be tolerated provided that the solvent absorbance is small (Table 4). Similar values of r for acetone to those shown in Table 4 were obtained for measurements made on the Buhler instrument.

Small differences between the pathlengths used to measure the sample and reference were found not to be important. This is shown in Table 4 for acetone measured with respect to a fixed 2 mm air reference.

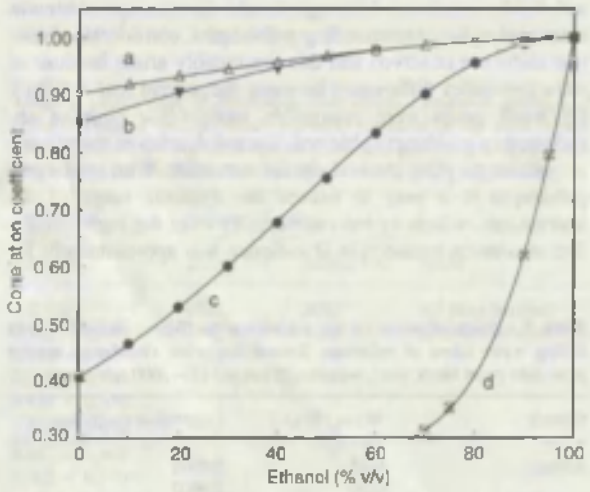


Fig. 6 Change of correlation coefficient with solvent composition: (a) ethanol–propan-2-ol; (b) ethanol–methanol; (c) ethanol–acetone; and (d) ethanol–water mixtures. Correlation coefficient values with respect to ethanol. Second-derivative absorbance spectra (nine data point block size), wavelength range 1136–2000 nm. Setup 3.

Table 6 Correlation coefficients for solvents not in library: second-derivative absorbance spectra (nine data point block size), wavelength range 1136–2000 nm

Sample	Setup	Correlation coefficient, r	Closest match solvent
Benzyl alcohol	5	0.889	Toluene
Decane	5	0.744	Butan-1-ol
Dichloroethane	5	0.723	Acetonitrile
	7	0.716	
Dodecane	5	0.677	Butan-1-ol
Ethyl methyl ketone	3	0.819	Acetone
	5	0.823	
	7	0.830	
Hexane	1	0.933	Butan-1-ol
	4	0.907	
	5	0.908	
	7	0.891	
Heptane	3	0.856	Butan-1-ol
	5	0.867	
	7	0.837	
Industrial methylated spirits, 74 OP	5	0.994	Ethanol (abs.)
Industrial methylated spirits, 64 OP	5	0.981	Ethanol (96%)
Nonane	5	0.784	Butan-1-ol
Octane	5	0.824	Butan-1-ol
	7	0.783	
Pentane	1	0.929	Butan-1-ol
	4	0.929	
	5	0.925	
Isopentane	5	0.827	Propan-1-ol
	7	0.823	Butan-2-ol
Pentan-1-ol	5	0.938	Butan-1-ol
Undecane	5	0.710	Butan-1-ol
Water	5	0.097	Ethanol (96%)

Effect of temperature

NIR spectra are known to be sensitive to temperature changes.[16,17] The effect of temperature was investigated using setup 1 by measuring spectra of all the solvents except 96% ethanol at approximately 0, 20 and 33 °C. These spectra were then compared against the reference library (second-derivative absorbance, nine data point block size). No significant effects on the correlation coefficients were observed and in no case did the r value fall below 0.992. Visually the most affected solvents were those in which hydrogen bonding occurred and absorbance changes in the OH regions could be easily seen. Fig. 4 shows the effect of temperature on the spectrum of methanol.

Effect of water

Water is a common impurity in most solvents and its effect on the ability to identify the solvents correctly was investigated. Samples of solvents were spiked with small amounts of water and the spectra compared with the original sample spectrum. The amount of water added corresponded approximately to 0.5 times and the same maximum water content as specified by the manufacturer for the solvent. The effects on r were small and caused no problems in correctly identifying the solvents (Table 5).

Very small amounts of water could easily be detected by visual inspection of the spectra. DMF is very hygroscopic and the uptake of water on exposure to air over a period of only 10 min is easily observed (Fig. 5). The uptake of water by acetone and methanol, although much less, could also be easily seen over a similar time period. The importance of visually comparing spectra and not simply relying on the r value cannot be too strongly stressed.

Solvent mixtures

Although the library was developed primarily for the identification of pure solvent samples, the ability to reject a number of binary mixtures was investigated. The correlation coefficients for acetone, methanol, propan-2-ol and water with ethanol are shown in Fig. 6. The values did not fall below 0.99 for mixtures containing up to approximately 10% v/v acetone and 20% v/v methanol and propan-2-ol. Mixtures of similar solvents will always be difficult to identify correctly using correlation in wavelength space and give rise to a high false positive rate. This needs always to be kept in mind when carrying out identifications by the proposed method.

Table 7 External validation of library: second-derivative absorbance spectra (nine data point block size), wavelength range 1136–2000 nm

Solvent	No. of batches	Correlation coefficient, r
Acetone	1	0.982
	40	>0.992
Dimethylformamide	1	0.983
	18	>0.991
Dichloromethane	42	>0.998
Ethyl acetate	1	0.992
	68	>0.996
Methanol	1	0.973
	1	0.990
	125	>0.993
Propan-2-ol	31	>0.998
Tetrahydrofuran	19	>0.998
Toluene	57	>0.998

Solvents not in library

There is always the possibility that a solvent external to the reference library will give a good match to a library solvent and be wrongly identified. Table 6 shows the correlation coefficients for a range of common solvents which were not included in the library and apart from industrial methylated spirits (IMS) no mis-identifications occurred. Not surprisingly, IMS 74 OP (~95% ethanol, ~4% methanol, ~1% water) matched absolute ethanol and IMS 64 OP (~86% ethanol, ~4% methanol, ~10% water) gave a fairly good match to 96% ethanol (r = 0.981). IMS represents a difficult problem that cannot be solved using the simple correlation in wavelength space algorithm used in this paper for identification.

External validation

To test the robustness of the library, it was used to identify the spectra from 405 different batches of solvents collected over a period of about 3 years (setup 4). Table 7 summarises the results, showing that all batches of solvents were correctly identified. In all cases where the r value was lower than normal (< 0.99), visual inspection of the spectra clearly showed the presence of traces of water in the solvents.

Conclusion

NIR transflectance spectra measured with respect to an air reference have been shown to be transferable, without any correction, between different instruments not only from the same manufacturer, but also between grating and FT based systems. A spectral library of 15 commonly used pure solvents constructed on one instrument could be used to identify reliably solvent spectra measured on other instruments using correlation in wavelength space. A correlation coefficient of > 0.97 was a clear indication of a positive identification, although r values below 0.99 were indicative of the presence of impurities. Small changes in pathlength, temperature and trace impurities of water could all be tolerated. Clearly false negative or false positive identifications can never be ruled out: however, using an r value of > 0.99 the data in Tables 2 and 7 suggest a false negative rate of 1–2%. It is important to include all closely related substances to those of direct interest in the library and other likely substances so that the optimum wavelength range, mathematical pre-treatment, *etc.*, of spectra are chosen to minimise these errors. The importance of visually inspecting spectra and not just relying on the r value cannot be too strongly stressed. The results of an inter-laboratory trial to test the robustness of this library and procedure will be reported shortly.

Acknowledgements

The authors thank Bran+Luebbe, Buhler and FOSS NIRSystems for the loan of NIR instruments. They also thank Nigel North, Kevin Smith and Lesly Senior of SmithKline Beecham Pharmaceuticals and Paul Frake of Glaxo Wellcome for the use of NIR instruments. Weng Li Yoon is grateful to SmithKline Beecham Pharmaceuticals for the funding of a research studentship.

References

1 *Good Manufacturing Practice: Medicinal Products for Human and Veterinary Use*, European Commission, Brussels 1997, vol. 4, p. 103.
2 Y. Wang, D. Veltkamp and B. R. Kowalski, *Anal. Chem.*, 1991, **63**, 2750.
3 E. Bourevesse, C. Sierra, J. L. Linossier and D. L. Massart, *Analusis*, 1996, **24**, 394.
4 E. Bourevesse, C. Hartmann, D. L. Massart, I. R. Last and K. A. Prebble, *Anal. Chem.*, 1996, **68**, 982.
5 E. Dreassi, G. Ceramelli, P. L. Perrucio and P. Corti, *Analyst*, 1998, **123**, 1259.
6 T. B. Blank, S. T. Sun, S. D. Brown and S. L. Monfre, *Anal. Chem.*, 1996, **68**, 2897.
7 H. Mark and J. Workman, *Spectroscopy*, 1998, **11**, 28.
8 C. E. Miller, *NIR News*, 1993, **6**, 3.
9 W. L. Yoon, R. D. Jee and A. C. Moffat, *Analyst*, 1998, **123**, 1029.
10 P. C. Knee and A. J. Deadman, *Intercomparison of Near Infrared Diffuse Reflectance Measurements*, Report COEM 6, National Physical Laboratory, Teddington, 1998.
11 A. Springsteen and T. Ricker, *NIR News*, 1996, **7**, 6.
12 J. A. Jones, I. R. Last, B. F. MacDonald and K. A. Prebble, *J. Pharm. Biomed. Anal.*, 1993, **11**, 1227.
13 S. L. R. Ellison, S. Gregory and W. A. Hardcastle, *Analyst*, 1998, **123**, 1155.
14 W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, Cambridge, 2nd edn. 1992.
15 C. J. Pouchet, *The Aldrich Library of FT-IR Spectra*, Aldrich Chemical, Milwaukee, WI, 1985.
16 K. DeBrackeleer, F. C. Sanchez, P. A. Hailey, D. C. A. Sharp, A. J. Pettman and D. L. Massart, *J. Pharm. Biomed. Anal.*, 1998, **17**, 141.
17 F. Wulfert, W. T. Kok and A. K. Smilde, *Anal. Chem.*, 1998, **70**, 1761.

Paper 9/03398I


TDK
CD-R74
CD-RECORDABLE
COMPACT
disc
Recordable
650MB