Time-Resolved Spectroscopy for Pharmaceutical Applications

Christoffer Abrahamsson

Department of Physics
2005

Lund Institute of Technology
Lund University

Doctoral Thesis
Atomic Physics Division
Department of Physics
Lund Institute of Technology
P.O. Box 118
SE–221 00 Lund
Sweden

Lund Reports on Atomic Physics, LRAP-348
ISSN 0281-2762
ISBN 91-628-6613-3

Printed by KFS, Lund, Sweden

Till Jenny

Abstract

The work presented in this thesis aims at improving spectroscopic techniques for analysis of pharmaceutical solids. This improvement is achieved by the combining and development of instrumentations and data evaluation tools from two research fields: the fields of near infrared spectroscopy and biomedical optics.

The thesis includes the construction and evaluation of different variable selection techniques. Variable selection is an important tool used to improve the evaluation of spectroscopic data. Variable selection was applied to near infrared data from pharmaceutical tablets and mid infrared data from atmospheric gases.

A novel instrumentation, using a photonic crystal fibre for light generation and a streak camera for detection, was developed, tested and used for time-resolved measurements. The system covers a wavelength range from 500 to 1200 nm and the time-resolution was measured to be 30 ps. The system proved to be very versatile and was used both for reflectance and transmission measurements.

New evaluation schemes for time-resolved data were developed. A new algorithm, based on diffusion theory was evaluated using time-resolved data required on apples. The same data was used to study the performance of a data evaluation scheme based on diffusion theory combined with least square support vector machines. Both evaluation schemes tested showed results comparable with results computed by the conventional evaluation scheme based on diffusion theory.

Time-resolved measurements were conducted on pharmaceutical solids. Quantitative analysis of intact tablets using time-resolved data was superior to analysis made using conventional near infrared data. This was especially true when there was a big difference in physical properties between the measured samples. An analysis method combining time-resolved spectroscopy and conventional near infrared spectroscopy was also developed. The scheme allows evaluation over a larger wavelength range than the one covered by the time-resolved system. The maximum range is only limited to the range covered by the near infrared spectroscopic instrument. The work also takes the first step toward the construction of a bench-top system, showing that the data from a simplified time-resolved system would still give valuable results.

Sammanfattning

Arbetet som denna avhandling baseras på syftar till att utveckla spektroskopiska tekniker för analys av farmaceutiska prover. Detta görs genom att kombinera och utveckla instrument och datautvärderingstekniker från två olika forskningsfält: spektroskopi i det nära infraröda (NIR) våglängdsområdet och biomedicinsk optik.

Avhandlingen behandlar utveckling och utvärdering av olika variabelselektions metoder. Metoderna är viktiga, både för att förbättra kvalitén på redan uppmätta spektroskopiska data och vid utveckling av nya instrument. I detta arbete användes variabelselektion både vid utvärdering av NIR spektra uppmätta på tabletter samt vid utvärdering av infraröda spektra uppmätta på gasblandningar.

En ny försöksuppställning för tidsupplöst spektroskopi utvecklades och utvärderades. Systemet använder sig av en speciell typ av optisk fiber (photonic crystal fibre) för generering av ljuspulser med en bred våglängdsfördelning och en streak-kamera för detektion av ljuspulserna efter att de passerat provet. Systemet kan göra tidsupplösta mätningar med en tidsupplösning på 30 ps inom ett våglängdsområde som sträcker sig från 500 till 1200 nm.

Två nya utvärderingsmetoder för att analysera tidsupplösta data utvecklades. Båda metoderna baseras på diffusionsteori. Den ena använder en anpassningsalgoritm för att beräkna de optiska storheterna hos proverna, medan den andra bestämmer de optiska storheterna hos proverna från en icke-linjär kalibreringsmodell (least squares support vector machines) baserad på teoretiska data. Båda metoderna visade jämförbara resultat med den vanligast förekommande metoden för att utvärdera tidsupplösta data.

Vid kvantitativ analys av tabletter visade sig tidsupplöst spektroskopi ge bättre resultat än konventionell NIR spektroskopi. Skillnaderna var som störst när de fysiska egenskaperna hos de studerade tabletterna varierade mycket. Även en analysmetod som kombinerar data från tidsupplösta mätningar med data från konventionell NIR spektroskopi utvecklades. Metoden har den fördelen att utvärderingar kan göras i ett stort våglängdsintervall, som bara begränsas av det konventionella NIR instrumentet. Metoden är också det första steget i utvecklingen av ett förenklat system, som skulle vara mer lämpat för praktisk användning på ett laboratorium.

Contents

List of papers

1. **Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets**
   Abrahamsson C, Johansson J, Sparen A, Lindgren F
   Chemometrics and Intelligent Laboratory Systems 69 (1-2): 3-12, 2003

2. **Multi-component chemical analysis of gas mixtures using a continuously tuneable lidar system**
   Weibring P, Abrahamsson C, Sjöholm M, Smith JN, Edner H, Svanberg S
   Applied Physics B-Lasers and Optics 79 (4): 525-530, 2004

3. **Time and wavelength resolved spectroscopy of turbid media using light continuum generated in a crystal fiber**
   Abrahamsson C, Svensson T, Svanberg S, Andersson-Engels S, Johansson J, Folestad S
   Optics Express 12 (17): 4103-4112, 2004

4. **MADSTRESS: A linear approach for evaluating scattering and absorption coefficients of samples measured using time-resolved spectroscopy in reflection**
   Chauchard F, Roger J-M, Bellon-Maurel V, Abrahamsson C, Andersson-Engels S, Svanberg S
   Applied Spectroscopy: Accepted for publication

5. **Least Squares-Support Vector Machines modelisation for Time Resolved Spectroscopy**
   Chauchard F, Roussel S, Roger J-M, Bellon-Maurel V, Abrahamsson C, Svensson T, Andersson-Engels S, Svanberg S.
   Applied Optics: Accepted for publication

6. **Time-resolved NIR/Vis spectroscopy for analysis of solids: Pharmaceutical tablets**
   Johansson J, Folestad S, Josefson M, Sparen A, Abrahamsson C, Andersson-Engels S, Svanberg S
   Applied Spectroscopy 56 (6): 725-731, 2002

7. **Time-resolved NIR spectroscopy for quantitative analysis of intact pharmaceutical tablets**
   Abrahamsson C, Johansson J, Andersson-Engels S, Svanberg S, Folestad S
   Analytical Chemistry 77 (4): 1055-1059, 2005

8. **Scatter correction of transmission NIR spectra by photon migration data - quantitative analysis of solids**
   Abrahamsson C, Löwgren A, Strömdahl B, Andersson-Engels S, Johansson J, Folestad S
   Applied Spectroscopy: Accepted for publication

Chapter 1

Introduction

The work presented in this thesis aims at improving spectroscopic techniques for analysis of pharmaceutical solids. This improvement is achieved by combining and developing instrumentations and data evaluation tools from two research fields: the fields of near infrared spectroscopy and biomedical optics.

NIR spectroscopy is an important tool for analysis of pharmaceutical samples. There are several reasons for this, where the most prominent are the robustness of the instruments, the fast measurement procedure and the possibility to perform on-line measurements through the use of fibre optics. Spectral measurements in the NIR region is though somewhat difficult to evaluate using standard univariate evaluation schemes. The absorption features are heavily overlapped and light scattering effects are substantial in turbid pharmaceutical solids. The use of multivariate data analysis techniques is therefore a necessity to extract the information from recorded spectral data. Multivariate data analysis and NIR spectroscopy has proven to be a successful combination, used in numerous studies for spectroscopic analysis in different disciplines.

The main limitation when using NIR spectroscopy is the fact that the measured spectra contain a mixture of scattering and absorption information. The analysis of NIR data can never be used to completely separate the absorption effects from the scattering effects in the recorded spectroscopic data. A multivariate calibration model can only handle a modest difference in scattering properties amongst the measured samples. As soon as the changes in optical properties are big enough, absorption and scattering effects get mixed up, leading to erroneous evaluations. Changes in scattering can be introduced by changes in the quality

of raw materials or other changes in the physical properties of the samples, for example sample density or particle size distribution. Within the NIR community many mathematical techniques have been developed in order to correct for the scattering effects, but no universal solution to the scattering problem has yet been found.

Similar problems are present within the field of biomedical optics, where spectroscopic techniques are used for analysis of tissue. To characterise tissue the absorption of different tissue constituents must be quantified. As light scattering by tissue is not constant, measurement techniques and theoretical models have been developed in order to separate the absorption effects from the scattering effects of the recorded data. These developments include time-resolved, spatially resolved and frequency-resolved instrumentations as well as models for light transport in turbid media. These different techniques allow the measurement of the absorption without any interference of scattering changes in the sample.

The main idea presented in this thesis is to develop tools for pharmaceutical analysis inspired by solutions from the field of biomedical optics. Consequently a time-resolved set-up was developed that are well adapted for measurements of the optical properties of pharmaceutical solids. The data from the time-resolved measurements was used directly for quantitative analysis of the measured samples, or combined with data required with a conventional NIR instrument to further enhance the results.

This thesis is divided into two parts. The first part contains some background information that serves as an introduction to the scientific work presented in the eight original papers forming the second part. Papers 1 and 2 contain work based on variable selection techniques, techniques important when working with multivariate data analysis and instrument development. Variable selection was applied to near infrared spectra measured on pharmaceutical tablets and to mid infrared spectra measured on atmospheric gases. Paper 3 describes the set-up and capabilities of the novel time-resolved system developed in this work. The system includes a photonic crystal fibre for light generation and a streak camera for detection. The system was used in Papers 4 to 8. Papers 4 and 5 describe new techniques for the evaluation of time-resolved data. Both tested evaluation schemes showed results comparable with results computed by the conventional evaluation scheme based on diffusion theory. Papers 6 to 8 describes time-resolved measurements on intact pharmaceutical tablets. Paper 6 demonstrates some basic capabilities of time-resolved spectroscopy while Paper 7 compares quantitative assessments made by time-resolved spectroscopy and conventional near infrared spectroscopy.

The comparison reveal that the analysis using time-resolved spectroscopy is superior to the analysis using conventional near infrared spectroscopy, especially when the difference in physical properties between the measured samples were large. Paper 8 shows the full potential of the technique when time-resolved measurements are complemented with conventional near infrared measurements.

Chapter 2

Interaction of light with turbid media

This chapter will introduce some basic properties of light and turbid media, and discuss the basic principles of interactions between light and matter. These properties and principles are important for the understanding of this thesis.

2.1 Basic properties of light

Electromagnetic radiation can be described in two complementary ways. It can be seen as an electromagnetic wave or as a stream of particles, photons. Both descriptions are valid, but are used in different applications. The particle model is well suited for light-matter interactions on a microscopic level, such as energy transitions in a molecule, while the wave model is more appropriate when describing interference phenomena. For some applications it might be difficult to judge which model to use, and often it is suitable to use a combination of the two.

Since the two models describe the same physical phenomenon, they have some measures in common. For example, light described as a photon will have the same velocity as the corresponding wave and the energy of a photon is proportional to the wavelength of the electromagnetic wave. The correlation is given by:

$$E = \frac{hc}{\lambda} \tag{2.1}$$

where E denotes the energy, h Planck's constant, c the speed of light and λ the wavelength. The electromagnetic spectrum can be divided into different regions. Fig. 2.1 shows the regions ranging from X-rays to radio waves. The term light is generally associated only with the visible region of the electromagnetic spectrum (400-770 nm), but will in this thesis be used for all

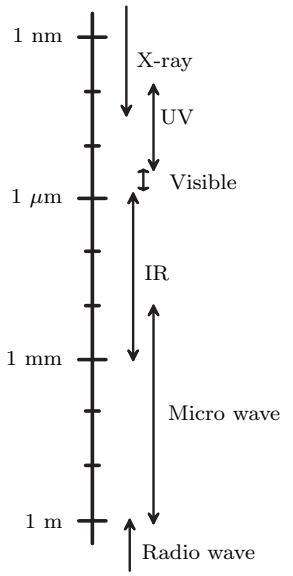


Figure 2.1. *The different regions of the electromagnetic spectrum as a function of wavelength.*

electromagnetic radiation, irrespective of its wavelength.

The wavelength region of most interest in this thesis is the near infrared (NIR) region, which is a part of the infrared (IR) region. The large IR range is usually divided into three sub-ranges: near IR (770-2500 nm), mid IR (2.5-25 μm) and far IR (25μm-1 mm)[1].

2.2 Basic properties of atoms and molecules

2.2.1 Energy levels

Every atom or molecule contains at least one electron. The electrons can be seen as negative particles contained in the field caused by the positive nucleus. The energy levels in the field that the electrons can occupy are quantified. For simple systems, like the hydrogen atom, these electronic energy levels (E_e) can be accurately calculated by means of quantum mechanics. For more complicated atoms and molecules, it is almost impossible to calculate the energy levels. However, by introducing simplifications and approximations it is still possible to the get approximate values of the electronic energy levels.

Molecules contain several atoms, allowing them to vibrate and rotate. The vibrations and rotations give rise to quantified vibrational and rotational energy levels.

The most basic molecule is a diatomic molecule, like NO or CO. One way to describe the vibrational energy levels (E_v) in a diatomic molecule, is to use the harmonic oscillator approximation, given by:

$$E_v = h\nu\left(v + \frac{1}{2}\right) \tag{2.2}$$

where v is the vibrational quantum number and ν is the classical vibrational frequency that is related to the reduced mass (μ) and the force constant (k) by:

$$\nu = \frac{1}{2\pi}\left(\frac{k}{\mu}\right)^{1/2} \tag{2.3}$$

In the harmonic oscillator approximation the energy levels are evenly spaced, but due to electronic repulsion and Van der Waals forces in the molecule, Eq. 2.2 has to be modified to:

$$E_v = h\nu\left(v + \frac{1}{2}\right) - x_e h\nu\left(v + \frac{1}{2}\right)^2 \tag{2.4}$$

where x_e is an anharmonicity constant. By introducing anharmonicity, closer lying energy levels are obtained, developing into

a continuum when E_v approaches the dissociation energy (E_d) of the molecular bond, see Fig. 2.2. Calculating vibrational energy levels for larger molecules is not as straightforward because the vibrational modes become more complicated than the simple stretching in the diatomic molecules. A non-symmetric molecule with n atoms will have $3n-6$ vibrational modes. Some examples of vibrational modes found in larger molecules are rocking, twisting and wagging[1].

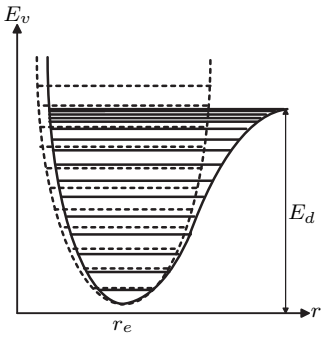


Figure 2.2. *Potential energy curve and energy levels for a diatomic molecule behaving as an anharmonic oscillator compared with those for a harmonic oscillator (dashed line). r is the radius of the molecule, with r_e being the equilibrium bond length.*

The rotational energy (E_r) for a diatomic molecule can be calculated by approximating the molecule with a rigid rotor:

$$E_r = \frac{h^2}{8\pi^2 I} J\,(J+1) \tag{2.5}$$

where J is the rotational quantum number and I is the moment of inertia defined as:

$$I = \mu r^2 \tag{2.6}$$

where μ is the reduced mass and r is the distance between the atoms. Eq. 2.5 implies that the energy distance between rotational energy levels become larger at higher lying levels, see Fig. 2.3. This is the opposite behavior compared to the energy distance between vibrational energy levels. It is almost impossible to calculate the rotational energy levels for larger, non linear molecules. The rotational quantum number, J, will though remain a good quantum number that can be used in approximative calculations.

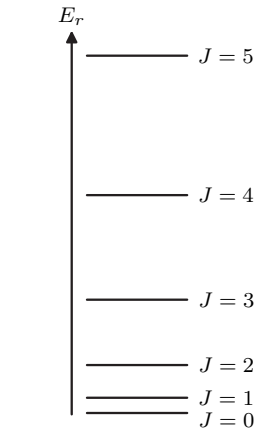


Figure 2.3. *Rotational energy levels in a diatomic molecule.*

2.2.2 Transitions

Transitions between energy levels are determined by selection rules that describe which transitions are allowed and which are not[2].

Transitions between electronic levels in atoms and molecules depend on several selection rules that will not be described in this thesis. The typical energy for a transition between two electronic energy levels corresponds to light in the UV or visible wavelength regions.

If the vibrations of a molecule could be described as a harmonic oscillator the selection rule would be $\Delta v = \pm 1$, but due to the anharmonicities described in section 2.2.1, the selection rule is modified to become:

$$\Delta v = \pm 1, \pm 2, \pm 3 \ldots \tag{2.7}$$

where transitions with $\Delta v = \pm 2, \pm 3 \ldots$ are called overtone transitions and are usually weak compared to the fundamental

transitions with $\Delta v = \pm 1$. The most anharmonic, and therefore the most probable overtone transitions, are those in vibrational modes including light atoms, such as hydrogen[3]. The energy of most of the fundamental vibrational transitions corresponds to light in the mid IR wavelength region, while the energy of many of the overtone transitions corresponds to light in the NIR region.

Polyatomic molecules that possess several fundamental vibrational modes, may exhibit a simultaneous energy change in two or more of those modes. This results in weak transitions called combination and subtraction transitions, due to the fact that the energy of the transitions corresponds to frequencies that are the sum of ($f_1 + f_2$, $2f_1 + f_2$, etc.) or the difference ($f_1 - f_2$, $2f_1 - f_2$, etc.) between the individual vibrational frequencies.

Rotational transitions obey the selection rule $\Delta J = \pm 1$ and corresponds to light energies in the far IR and micro wave regions.

The transitions between energy levels are not restricted to one of the modes. Instead the molecule can change electronic, vibrational and rotational levels simultaneously.

2.3 Optical properties of turbid media

Turbid media is usually used as a term for media that are highly scattering, but exhibit a low absorption. Some examples of turbid media are pharmaceutical tablets and powders, tissue and milk.

2.3.1 Index of refraction

The index of refraction (n) is a fundamental property of any medium. In the linear regime the index of refraction is given by a linear complex function, but the imaginary part, treating the attenuation of the wave, is omitted in many applications. The real part is defined in terms of the velocity of light in the medium (c_m):

$$c_m(\lambda) = \frac{c}{n(\lambda)} \tag{2.8}$$

where c is the speed of light in vacuum. The index of refraction is wavelength dependent, as can be seen in Fig. 2.4.

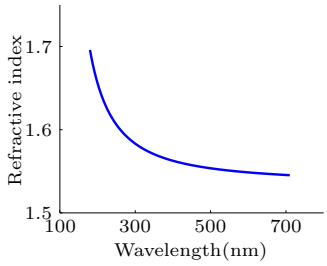


Figure 2.4. *Refractive index for α-quartz as a function of wavelength*[4].

2.3.2 Absorption

Light with a wavelength matching the energy difference between two energy levels in an atom or molecule, can be absorbed by a sample containing that particular atom/molecule. The absorbed energy will transfer the atom/molecule to an excited state. The

probability of a photon to be absorbed is given by the absorption coefficient, μ_a $[cm^{-1}]$. The absorption is strongly wavelength dependent, since different atoms and molecules absorb light at different wavelengths as described in section 2.2. The measured wavelength dependent absorption can be used as an identifier, or fingerprint of a substance.

2.3.3 Scattering

When light hits a particle it will be scattered. The scattering property of a medium is described by the scattering coefficient, μ_s $[cm^{-1}]$, that is the measure of the mean distance between scattering events.

There are two main scattering regimes, the Rayleigh and Mie regimes. If the particle is much smaller than the wavelength of the light, the scattering will be dominated by Rayleigh scattering. Rayleigh scattering is thus mainly due to molecules. The most important implication of the small size-to-wavelength ratio is that the scatterer sees a spatially uniform electric field that varies with time. This field induces an oscillating dipole moment in the particle, with the same frequency as the incoming light. The oscillating dipole will act as a transmitter re-emitting radiation with the same frequency as the incoming light. The distribution of the re-emitted light can be seen as isotropic, i.e. the light is equally distributed in all directions. The Rayleigh scattering is wavelength dependent. The wavelength dependence can be modeled by $\mu_s \propto \lambda^{-4}$. A closely related, but inelastic process is Raman scattering, in which the oscillating dipole interacts with a vibrational mode in the molecule. The radiation emitted from the molecule will then be of a different wavelength than the incoming radiation.

The other main scattering regime is the Mie regime. Mie theory describes scattering by spherical particles[5], and can in principle be applied to any size-to-wavelength ratio. Mie theory is primarily used in cases where the scattering particles are larger or of the same size as the wavelength, but small enough for geometrical optics not to be valid. The wavelength dependence as well as the angular dependence of Mie scattering are affected by the size of the scattering particles and the refractive indices of the particles and the medium surrounding the particles. The wavelength dependence can normally be assumed to be a function of the form $\mu_s \propto a\lambda^{-b}$, where a and b are constants and where b is smaller than 4. A typical fit to measured scattering coefficients can be seen in Fig. 2.5.

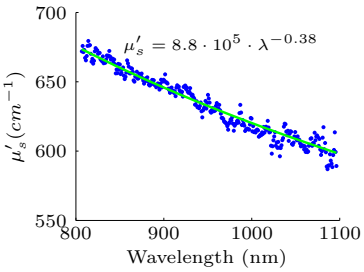


Figure 2.5. *Measured scattering coefficients (dots) and fit according to Mie theory (line). The measured values are contained from a time-resolved measurement on a pharmaceutical tablet.*

The direction of the light after a scattering event is defined by two angles, the deflection angle, θ, and the azimuthal angle, ψ.

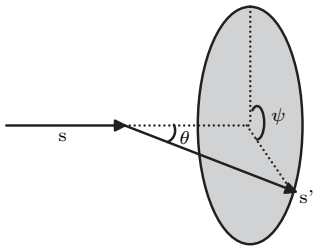


Figure 2.6. *Picture of the scattering angles θ and ψ.*

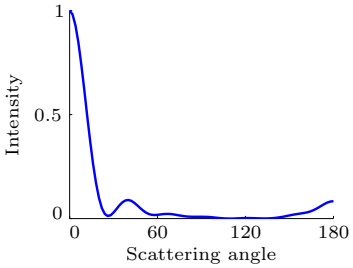


Figure 2.7. *Calculated Mie phase function for light with λ =1000 nm scattered by a 2 μm sized particle.*

The scattering in the azimuthal direction is often seen as isotropic and will therefore not vary with the size and shape of the scatterers. The definitions of the angles are seen in Fig. 2.6. The phase function, or directional distribution of the scattered light with regards to the deflection angle (θ), can be calculated by Mie theory, see Fig. 2.7 for an example, but simplified expressions like the Henyey-Greenstein phase function[6]:

$$p\left(cos\left(\theta\right)\right) = \frac{1-g^2}{2\left(1+g^2-2gcos\left(\theta\right)\right)^{3/2}} \tag{2.9}$$

is often applicable. By using the Henyey-Greenstein equation the directional distribution of the light after scattering is calculated by means of the anisotropy factor, g, which is the only variable in Eq. 2.9. The anisotropy factor is calculated as the expectation value for the cosine of the scattering angle, $cos\left(\theta\right)$, and can therefore only be a number between -1 and 1. A value near 1 means an almost totally forward scattering of the light, a value near -1 indicates backward scattering, while a value of 0 indicates isotropic scattering. Values of the anisotropy for pharmaceutical powders have, to my knowledge, not been measured but is thought to be approximately 0.8, a value comparable with the measured value of powdered TiO[7].

In some applications, such as when conducting diffusion calculations, the anisotropy factor is not regarded but the reduced scattering coefficient, μ_s' $[cm^{-1}]$, is used and defined as:

$$\mu_s' = \left(1-g\right)\mu_s \tag{2.10}$$

The inverse of μ_s' is a measure of the mean distance between artificial isotropic scattering events. The use of the reduced scattering coefficient is thus only valid after many scattering events, i.e. in the diffuse regime.

Another measure, that is often used, is the coefficient of transport, μ_{tr} $[cm^{-1}]$, defined as:

$$\mu_{tr} = \mu_a + \mu_s' \tag{2.11}$$

Chapter 3

Near infrared spectroscopy

Although the near infrared (NIR) region was discovered already in 1800[8], it was not used for spectroscopy to any larger extent until the 1950ies. In the last 50 years the number of applications of NIR spectroscopy has increased tremendously. This chapter will focus on instrumentation and applications for measurements on solid samples in general and pharmaceutical samples in particular.

3.1 Instrumentation

All spectroscopic systems have the same basic components, see Fig. 3.1. The distinguishable parts are the light source, the wavelength selection device, the sample interface optics and the detector. The wavelength selection device may be placed in front or behind the sample.

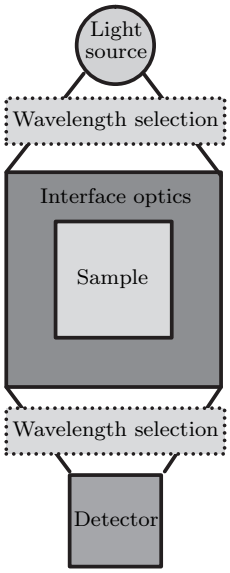


Figure 3.1. *Basic components of a spectroscopic system.*

3.1.1 Light sources

In most commercial systems a tungsten halogen lamp is used as light source. The lamp has a continuous spectrum covering the entire NIR region and has the benefit of being cheap and robust. One of the drawbacks with tungsten halogen lamps is that their emission is temperature dependent. The drift is typically 0.1 % per degree Kelvin in the NIR range[9]. The temperature effects can be minimized by using temperature stabilisation.

An alternative to tungsten halogen lamps is the use of one or more light emitting diodes (LED). LEDs have a limited spectral bandwidth of approximately 30-50 nm, and LEDs are mainly found in spectroscopic applications where their small size and cool operation is beneficial. Examples of applications are measurements of blood oxygenation and blood volume[10–13] and

humidity[14].

Tunable lasers, used in many applications in other wavelength regions, have not been used to any large extent in the NIR region, except for applications where the samples are gaseous and therefore have narrow absorption bands, e.g. for gas absorption in scattering media[15;16].

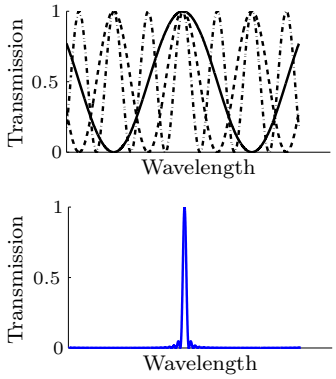


Figure 3.2. *Top figure shows the transmission through each of the three first layers of a Lyot filter. The bottom figure shows the total transmission through a five layer Lyot filter.*

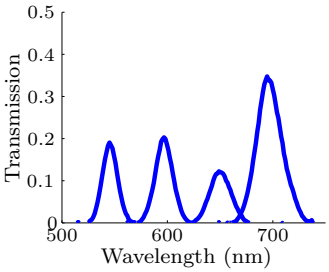


Figure 3.3. *Measured transmission through a liquid crystal tunable filter for four different central wavelengths.*

3.1.2 Wavelength selection

Filter based systems

The use of band pass filters as wavelength selection devices have the limitation that only one wavelength can be measured at a time. The filters usually have a bandwidth of 10 nm, and several filters can be combined by using a mechanically switched filter wheel, that mechanically change filters between measurements. Therefore band pass filters are mainly used in dedicated applications, where a few wavelengths are enough to obtain the requested information[17–20].

An alternative to multiple bandpass filters is the liquid crystal tunable filter (LCTF). An LCTF works basically as a multiple Lyot filter made of several stages, where each stage consists of a birefringent layer placed between two linear polarizers. Each succeeding stage has twice the retardation as the previous stage, effectively halving the width of the previous transmission peak, see Fig. 3.2. The resulting transmission through all layers becomes narrow, in the range of 10 nm. By inserting a layer of liquid crystal inside each stage, the retardation and therefore the maximum transmission wavelength can be changed[21], see Fig. 3.3. LCTFs have been used in different applications, for example hyperspectral imaging of pharmaceuticals[22] and within the field of biomedical optics[23–25].

Another type of tunable filter is the acousto-optic tunable filter (AOTF)[26]. Like the liquid crystal tunable filters the AOTF does not include any moving parts. The AOTF consists of a birefringent crystal (often TiO_2 in the NIR region) and a piezoelectric transducer. The piezoelectric transducer, driven by a radio frequency source, introduces a periodic variation of the refractive index within the crystal. This variation acts as a grating to the incoming polychromatic light, and the incoming beam will be divided into two beams with monochromatic light and one beam with the remainder of the polychromatic light. By changing the radio frequency, which drives the piezoelectric transducer, the wavelength of the monochromatic beams can be

controlled.

The different filter systems are suitable for different applications. Interference filters are best suited for systems where just a few wavelengths are needed, but where robustness and ruggedness is important. Both AOTFs and LCTFs are good when many wavelengths are to be measured. The AOTF gives the best wavelength resolution, while the LCTF can change wavelength faster and is better suited for imaging application due to less image distortions.

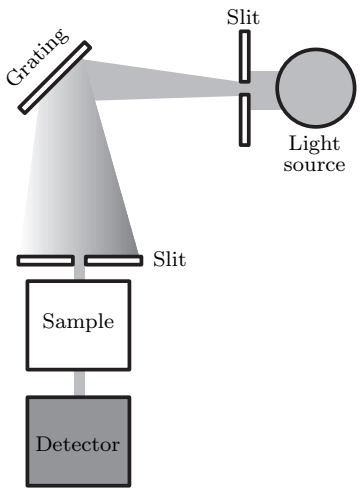


Figure 3.4. *Schematic picture of a grating based system for transmission measurements.*

Dispersive systems

Dispersive systems are probably the most commonly used type of instruments for NIR spectroscopy. In most dispersive systems the polychromatic light from the light source is divided into different wavelengths by a grating[27], see Fig. 3.4. The wavelength dispersion achieved from a grating is based on the number of grooves that are illuminated. In order to obtain a high resolution a highly dispersive grating, i.e. a grating with many grooves per unit length, in combination with a narrow entrance slit should be used. A narrow entrance slit leads to large diffraction of the light which means that a large portion of the grating will be illuminated. A long distance between the slit and the grating will also lead to the illumination of a large area of the grating, and by that a large number of illuminated grooves. These factors give rise to a trade off between signal intensity and resolution. In the NIR region, where the absorption features are wide, a low resolution grating, with a resolution in the range of 1-5 nm usually gives the best results.

Most commercial systems are based on concave holographic gratings, either mounted in a Rowland circle set-up or used as a moving grating in a scanning monochromator. The scanning grating systems can cover a wavelength range from 400 to 2500 nm and are relatively cheap, but quite slow and the moving parts of the systems deteriorate with time. The systems must therefore rely on an internal wavelength reference in order to correct for wavelength drift and response deterioration.

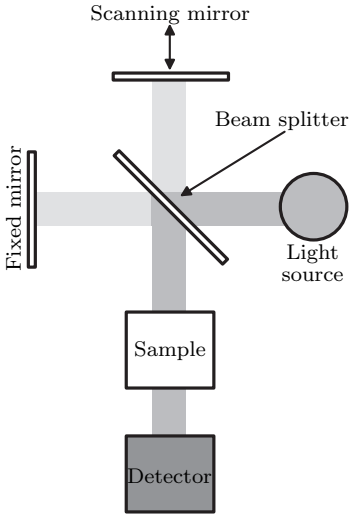


Figure 3.5. *Basic components of a Michelson interferometer system.*

Interferometric systems

The Michelson interferometer is the basis of most of the commercial Fourier transform (FT) systems. A schematic picture of a Michelson interferometer is seen in Fig. 3.5. The beam splitter divides the light from the source into two beams. One is reflected by a fixed mirror and one by a scanning mirror. The reflected beams are then recombined by the beam splitter and sent to the sample and the detector. By moving the scanning

mirror an intensity modulated response will be recorded by the detector. Applying a FT algorithm to the intensity modulated signal will recover the spectral information, and hence allow the reconstruction of the light intensity as a function of wavelength.

FT based systems have a couple of advantages compared to dispersive systems[1;2;28]. The Flegetts multiplex advantage comes from the fact that all wavelengths are measured simultaneously by the interferometer, while a dispersive system measures one wavelength at the time. Another important advantage is the Jacquinots throughput advantage, arising because no slit is needed in the instrument to gain a high resolution. The resolution of an interferometer is mainly restricted by the movement of the scanning mirror. In order to control the movement of the scanning mirror, a reference laser beam is used. This internal reference gives the FT instruments the advantage of an outstanding wavelength accuracy compared to dispersive systems.

3.1.3 Sample interface

The most common measurement modes used in NIR spectroscopy of solid samples are diffuse reflectance and diffuse transmission. Normally these set-ups use some basic collimation optics, such as a parabolic mirror at the light source, but none at the detector side. Instead the detectors are often placed close to the sample surface. This makes the set-up simple but it can introduce problems with stray light and boundary effects, especially when measuring on small samples[26]. Two alternatives to decrease the effect of the detector placement and ensure the measurement of only the diffusely scattered light are the introduction of an integrating sphere at the sample surface or the inclusion of some optics to put the detectors off axis from the sample[29].

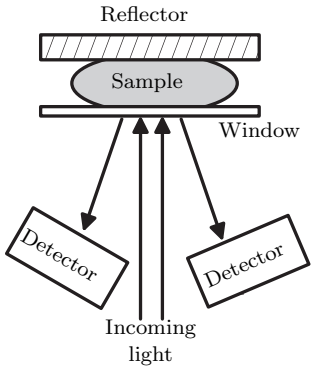


Figure 3.6. *Measurement set-up for transflectance measurements.*

The transflectance geometry is a third measurement geometry, only used in special cases where the samples do not allow transmission or reflectance measurements. When measuring in transflectance mode, a set-up similar to the set-up used when measuring in reflectance mode is used, but with the difference that a reflector is used, see Fig. 3.6. The reflector is placed on top of the sample, reflecting the part of the light that should otherwise have been transmitted. The transflectance geometry has been used in e.g. gel measurements[30].

One big advantage of using light in the NIR region compared to using the mid- or far-IR regions is the possibility to use standard optics, e.g. quarts and glass optics. The losses in standard optics are very high for wavelengths above 2000 nm. This enables the usage of optical fibres and different sampling probes. The num-

ber of applications using fibre optical probes is increasing due to the development of probes for different sample geometries and the possibility to separate the sampling area from the spectrometer, i.e. making on-line measurements possible. One of the drawbacks when using fibre optics is the absorption from OH groups in the glass material at 1450 nm[31].

3.1.4 Detectors

The detectors used in the NIR region are different types of semiconductor detectors. All detector materials generate holes and electrons when exposed to light, but the detection schemes may differ between detectors. Some detectors measure the resistance changes in the material, while others measure the electrical current introduced by the light.

Different semiconductor materials are suitable for different wavelength regions. For systems measuring below 1600 nm InGaAs detectors are often used. At longer wavelengths, up to 2500 nm, PbS is the most common choice. Both these detectors work at room temperature, which makes the instrumentation simple, but cooling the detectors to 233 K with a Peltier element will decrease the thermal noise and therefore extend the useful wavelength range of the detectors. For longer wavelengths materials like InAs or InSb can be used[9].

3.2 Spectral information in the NIR region

The overtone and combination absorption bands in the NIR range are wide and most often superimposed. Therefore a multivariate approach for quantification and classification is often needed. Multivariate data analysis will be described in chapter 4. In this section other, and mainly univariate, approaches for the analysis of NIR data will be treated.

The data evaluation tools for quantitative and qualitative analysis of IR or visible spectra are difficult to apply to NIR spectra. Since a measured NIR spectrum is a measurement of the diffuse light being transmitted/reflected from the sample, the intensity is severely affected by the scattering properties of the sample. A typical spectrum from a NIR measurement on a pharmaceutical tablet is shown in Fig. 3.7. When using, for instance IR spectroscopy, a quantitative identification of the sample can be made by band assignments. It is possible to assign different bands in the NIR region as well, but a complete structure elucidation is almost impossible to perform, due to the broad and overlapping absorption features. However, NIR spectroscopy can,

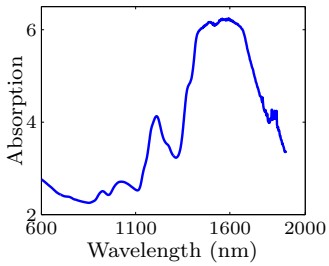


Figure 3.7. *Typical NIR absorption spectrum measured in transmission mode on a pharmaceutical tablet*

in some cases, be used for identification by using a multivariate calibration procedure.

The fact that it is multiple scattered light that forms the measured spectrum makes it impossible to use the Beer-Lambertian law to get any direct qualitative information from the spectrum, because the length of the light path through the sample is unknown. The Beer-Lambertian law is given by:

$$I = I_0 \times 10^{c\epsilon l} \tag{3.1}$$

where c is the concentration of the absorber, ϵ is the absorptivity and l is the optical path length. When using diffuse reflectance measurements the Kubelka-Munk function is a better alternative to extract quantitative information:

$$f(R) = \frac{(1-R)^2}{2R} = \frac{c}{a} \tag{3.2}$$

where $a = s/2.303\epsilon$, s is the sample dispersion coefficient and R is the diffuse reflectance. This equation gives a linear relationship between $f(R)$ and the concentration of the analyte. This relation only holds as long as the sample matrix does not absorb, the absorption band of the analyte is weak and the concentration variations are small[32].

3.3 Pharmaceutical applications

The pharmaceutical applications using NIR spectroscopy are many and very diverse. This section is intended to give some examples of the use of NIR spectroscopy in the pharmaceutical industry. For a more comprehensive overview several review articles on this topic are available[3;33;34].

3.3.1 Qualitative applications

Many papers report the use of NIR spectroscopy for identification of raw materials. Three examples of excipient classification are Svensson et al.[35] that classified 11 different sorts of cellulose, Krämer et al.[36] that report a reflectance method to separate 8 types of cellulose and Candolfi et al.[37] that classify 10 different excipients also using the reflectance mode. Identification of active drugs has been demonstrated by Gärhausser et al.[38], separating 17 different benzodiazepines. Another important application is quality control of raw materials. Andre[39] constructed a scheme for quality control of an active pharmaceutical ingredient using NIR spectroscopy. The method required only 2% of the workload

needed for a similar evaluation using conventional analysis methods.

Qualitative analysis is not only performed on raw materials, but also on intact tablets. One example is Dempster et al.[40] that developed a method for the identification of blister packed tablets using reflectance NIR spectroscopy.

3.3.2 Quantitative applications

There are quantitative applications for almost any form of pharmaceutical formulation, e.g. tablets[41–44], low dose tablets[45] and films[46;47].

When conducting quantitative measurements on intact tablets the choice of measurement geometry is more crucial than when a qualitative analysis is sufficient. Several studies[48–51] have compared quantitative measurements made using transmission and reflectance mode. All groups conclude that the transmission measurements are better when conducting a quantitative analysis of the active ingredient in intact tablets, mainly due to the larger sampling volume compared to the sampling volume obtained by a reflectance measurement. However, transmission measurements can be difficult to conduct on thick tablets, where reflectance measurements normally are recommended, but Ramirez et al.[52] showed that optical dense tablets can be compressed in order to increase the transmitted signal.

Clarce et al.[53] measured the depth from where information was collected in reflectance measurements, and found that the depth varied, both with sample type and wavelength, but that the information depth was at least 200 μm, but could reach 500 μm for shorter wavelengths (in cellulose samples). The sampling size was also studied by Berntsson et al.[54], that measured the sampling size in diffuse reflectance measurements on pharmaceutical powders to be in the range between 15 and 70 mg/cm^2.

Blanco et al.[55] points out one of the major drawbacks with NIR spectroscopy for quantitative measurements, namely that tablets from a different batch than the ones used in the calibration model are almost impossible to quantify correctly, even though several pretreatment techniques are used. This imply that a strong limitation in all the presented spectroscopic applications is that the optical path length in the measured samples may vary due to variations in their scattering properties or due to changes of the measurement geometry.

3.3.3 On-line applications

The development and use of fibre optical probes have made on-line NIR monitoring of almost all different steps of the tablet production feasible. One basic problem when measuring on moving samples, which is often the case for on-line application, is the spectral artifacts introduced by the movement. This problem has been examined by Andersson et al.[56](for a dispersive system) and Berntsson et al.[57](for a FT system). Both studies conclude that measurements on moving samples are possible when a good design of the spectrometer and appropriate treatment of data is used to reduce the impact of the spectral artifacts.

Many pharmaceutical manufacturing schemes start with some kind of powder mixing. Berntsson et al.[58] monitored the mixing process at rather moderate mixing speeds in both small and production scale blenders with good results. Other groups like El-Hagrasy et al.[59] use a measurement scheme where the blender is stopped during the measurement. Both sampling techniques are promising, but have some drawbacks. Stopping the blender results in problems with sub-sampling, while a moving sample might introduce artifacts in the spectra.

It is important to measure the moisture content in all pharmaceutical formulations, since that is an important factor for the quality of the end product. Rantanen et al.[25] measured the moisture content on-line in a granulation vessel with errors less than 0.2 % while Berntsson et al.[60] measured the water content in gelatine capsules, with an at-line system, claiming to have an accuracy of 0.1 %.

Granulation is a complex process that has been studied with on-line NIR systems to improve the control of the process. Rantanen et. al.[61] used a NIR probe to study three subphases of the granulation procedure, and were able to determine the endpoints of each subphase using a multivariate technique to analyse the aquired NIR spectra.

One safety aspect is to incorporate an identity check of the final product before packaging. Herkert et al.[62] showed that a NIR spectroscopic system could be used to identify two different products, and a mixture of them, on-line at production speeds with 100 % accuracy.

3.3.4 Measurements of physical parameters

The impact of scattering on NIR spectra is not only a drawback, but have also been used to extract information about the physical state of the samples. Chen et al.[63] used reflectance NIR data to evaluate the hardness of the measured intact tablets, a measure that is important e.g. for the dissociation of the tablets and drug delivery. Otsuka[64] used NIR spectroscopy to extract the mean particle size of powders, in the range from 37 μm to 590 μm. The mean error of the size determination was measured to be 25 %. Gupta et al.[65] used the baseline slope of reflectance spectra, both as a measure of the strength of compacted pharmaceutical powders as well as a measure of the particle size distribution of the powders after milling of the compacts.

NIR spectroscopy using microscopes allow imaging of pharmaceutical preparations. Clarke et al.[66] used a combination of NIR and Raman microscopy to image the surface of intact tablets, to evaluate the specific formulation used.

Chapter 4

Multivariate data analysis

Spectroscopy has been used as a tool for chemical analysis for many years. In the early days the wavelengths used in a specific experiment was limited by the restrictions set by the instrumentation. At that time, a large set of required data might have included five wavelengths. In recent years the rapid development of spectroscopic devices and computers has resulted in much larger data sets. Today, the resulting data from a spectroscopic measurement often consists of thousands of measured wavelengths. To cope with these enormous matrices of data, the use of multivariate data analysis has become increasingly important. The main tools within multivariate data analysis are based on projection[67] techniques, in which the original data set with thousands of wavelengths, is projected onto a set of new, and fewer variables.

The projection techniques have several important and unique features:

- They are not restricted by any shape of the original data matrix.
- They handle collinearities.
- They handle missing data.
- They are robust to noise.
- They do not require any *a priori* knowledge of the recorded data.
- They provide informative diagnostic and graphical tools.

Projection tools are mainly available for the two major analytical problems, classification and calibration. Several methods have been proposed[68;69], but in this work only the basic, and most commonly used methods, principal component analysis (PCA) and partial least squares (PLS) will be described.

4.1 Principal component analysis

Principal component analysis (PCA) extracts the information residing in a multidimensional data matrix by projecting the original measured variables onto a set of new, orthogonal variables[70–72]. These new variables are called principal components (PC), and are constructed to include as much as possible of the relevant information in the original data matrix, without including the noise.

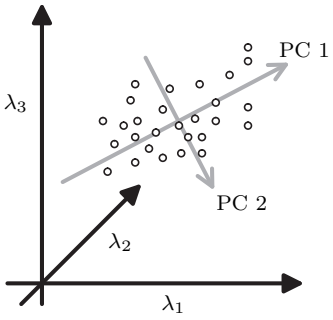


Figure 4.1. *Visualisation of the two first PCs in a PCA model based on a data set containing three wavelengths.*

The basic principles of PCA can be described by constructing a model from a simple data set containing only three measured wavelengths. When the samples are plotted as a function of the wavelengths, they form a cluster of points in a three-dimensional coordinate system. The first PC (PC 1) is then a new coordinate axis oriented in such a way that it covers as much of the variations in the swarm of points as possible. A second PC (PC 2) can be added to the model if PC 1 did not describe the variations in the original data set in a satisfactory way. PC 2 is a second new coordinate axis that covers as much as possible of the remaining variations, and is orthogonal to PC 1, see Fig. 4.1.

In this example it is not needed to continue adding PCs. If the original data set is more complex, more PCs can be calculated using the same criteria as for the first two. New PCs should be added until the remaining variations can be considered as noise. The maximum number of PCs that can be calculated in a model is the minimum of the number of observations and wavelengths in the data set.

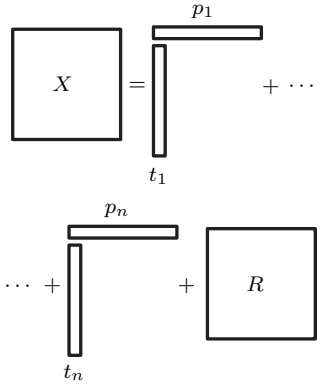


Figure 4.2. *Schematic description of the calculation of the first n number of PCs.*

The calculation of PC 1 is carried out by the construction of two vectors (t_1 and p_1) of which the product should match the original matrix, X, in the best possible way, minimising the eigenvalue of the residual matrix, R. The t vector is called the score vector and contains the new coordinates of the samples after projection onto the PC. The p vector is named the loading vector and can be interpreted as the cosines of the angles between the original variables and the new PC.

The residual matrix is used as input matrix when calculating the next PC. This procedure is continued until the information in the residual matrix is considered as only noise, see Fig. 4.2, giving the final model as:

$$X = T \times P' + R \tag{4.1}$$

Generally, a spectroscopic data set containing thousands of measured wavelengths is adequately described using less than ten PCs.

When evaluating a data set using PCA, both the score and loading vectors are important. As the score vectors establish the coordinates of the samples in the new coordinate system formed by the PCs, plotting the scores from different PCs may reveal a lot of information about the relationships between the samples. Such a plot is known as a score plot, and is one of the most commonly used tools in multivariate data analysis. By analysing the scores, classification of the samples can be achieved and outliers found. To know which wavelengths were influential for the PCA model, and how they are correlated to each other, the loadings have to be investigated by plotting the loading values as a function of wavelength. The plot unravels the importance of the wavelength to the calculation of the PC, where a high positive/negative loading value means high importance and a value close to zero means little or no importance for that PC. If two wavelengths have a similar loading value they are to be thought as highly correlated.

An example of a score plot that extracts information from a data set containing 300 measured wavelengths into just two PCs is shown in Fig. 4.3. The data set used contains evaluated time-resolved data from 27 tablets, where the tablets had different thickness and different concentration of the active substance. From the raw data it was impossible to separate the thickness information from the variations in concentration of the active substance, but using the scores from the PCA makes the information visible and the samples easy to classify.

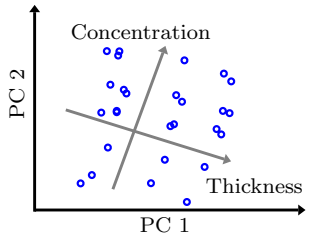


Figure 4.3. *Score plot from a PCA model based on spectral data from 27 tablets with varying thicknesses and concentration of active substance.*

4.2 Partial Least Squares

4.2.1 Algorithm

PLS is a regression extension of PCA that handles two blocks of variables, predictors (X) and responses (Y)[71;73]. The two data sets can be decomposed separately by means of PCA. This gives the outer relations:

$$X = T \times P' + R_X \tag{4.2}$$

$$Y = U \times Q' + R_Y \tag{4.3}$$

These calculations minimise the residuals, R_X and R_Y, without making any effort to correlate the data sets. A correlation between the two data sets can be found by forming a linear inner relation between the scores for each PC:

$$\hat{u}_h = b_h \times t_h \tag{4.4}$$

where h is the number of the specific PC and b is the regression coefficient. This model exhibits only a weak relation between the data sets.

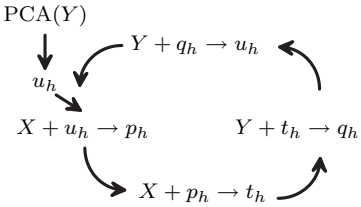


Figure 4.4. *Schematic illustration of the PLS algorithm. The procedure is repeated until a convergence criterium is reached.*

To improve the model, information from the decomposition of one of the two blocks must give information to the other and vice versa, giving slightly rotated PCs, but with a stronger coupling between the predictors and the responses. To accomplish this an iterative procedure is introduced, which is depicted in Fig. 4.4. The algorithm starts with a PCA modeling of the Y matrix. The score vector, u_h, from this model is used together with the X matrix to calculate a new loading vector, p_h. The loading vector is used together with the X matrix to calculate a new score vector, t_h, which is then used together with the Y matrix to calculate a new loading vector, q_h. The loop is closed when the loading vector and the Y matrix give a new score vector, u_h, corresponding to the score vector from the original PCA model. This procedure is repeated until the algorithm converges. When the convergence criterium is accomplished, the vectors are orthogonalised and the next partial least squares component, PLSC, can be calculated.

4.2.2 Validation of PLS models

The best way to evaluate the predictive power of a PLS model is to use an external test set with known responses (y values). The y values of the test samples are predicted by the PLS model and compared with the known y values. This strategy needs a lot of samples to get calibration and test sets that are big enough. If not enough samples are available, an alternative approach is to use a cross-validation (CV) procedure. The CV starts by splitting the data set into several different groups. A PLS model is then made using all but one of the groups, whose response values are predicted and saved. This procedure is redone until every group has been predicted once, and the total error can be calculated. The big drawback of using CV is that it is easy to over fit the PLS model and thereby drawing the wrong conclusions of the models quality.

The standard measure to use when evaluating a PLS model is the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{4.5}$$

where n is the number of samples, y_i is the known value and $\hat{y}_i$ is the value predicted by the PLS model. The RMSE value is often divided by the average y value to be able to get a relative measure of the introduced error.

It is important to notice that the RMSE gives the total errors, including errors in the spectral measurements, in the reference analysis and in the modeling procedure.

4.3 Preprocessing techniques

The data is often pretreated in order to transform it into the most suitable form for multivariate calibration. The use of preprocessing can make the difference between the construction of a successful and an unsuccessful calibration model. The preprocessing techniques range from simple data transformations and variable scaling to more advanced variable selection techniques and data correction algorithms.

4.3.1 Variable selection

Variable selection techniques can be used for two major purposes. One is to be a guide when designing systems, or when optimising a measurement procedure. Variable selection can for instance be used when a spectroscopic system has the capabilities to conduct measurements in a large wavelength range, but the timing of the measurement protocol only allows a few wavelengths to be measured. Another example is when a spectroscopic instrumentation is to be designed, and the design only allows the use of a limited number of wavelengths. The other main application is to remove non-informative wavelengths from already recorded spectroscopic data. The goal with this is to only include the parts of the spectra where information about the samples are present, and remove the other parts, that otherwise would have added noise to the data analysis.

A variety of algorithms for variable selection have been developed, from simple iterative procedures, based on a trial and error approach[74], to complex algorithms, nested into the multivariate calibration algorithm[75]. One of the more versatile variable selection techniques is the genetic algorithm[76] (GA), originally proposed by Holland[77]. In the remainder of this section a GA for variable selection will be described as an example of a variable selection algorithm[78–81].

GAs are inspired by evolution theory, according to which the individuals with the best fitness have both the greatest probabilities of surviving and the highest probabilities of winning the battles engaged for reproduction, thereby propagating their genome.

GAs involve four basic steps, see Fig. 4.5, where step 2-4 are repeated until a termination criterion is reached:

1. To allow an easy mathematical treatment of a chromosome, a coding is necessary. This is solved by representing each variable (gene) with a binary code in a vector (chromosome) with one cell for each variable. A variable is selected if a

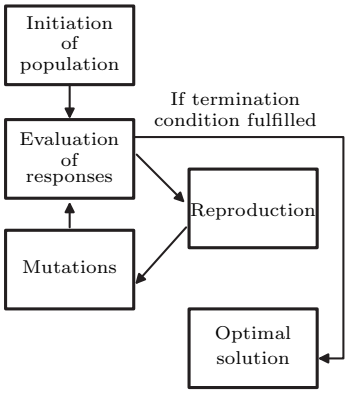


Figure 4.5. *Schematic illustration a genetic algorithm.*

1 is present and not selected if a 0 is present. The original chromosome is then perturbed randomly to make a number of chromosomes, the initial population.

2. For each chromosome the response associated with the corresponding experimental conditions is evaluated. This is usually done by making a PLS model for each chromosome. The models are then evaluated, by external or cross validation, in order to get numeric values describing their quality. These values are known as the fitness values and are the criterion for guiding the GA to the global optimum.

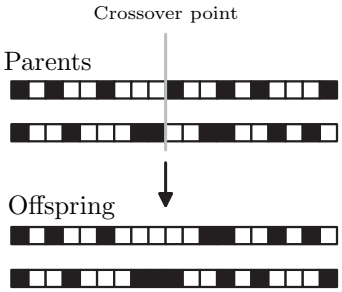


Figure 4.6. *Single-point crossover in the recombination step of a GA.*

3. The reproduction step creates a new population that can be considered as the next generation. The new generation of chromosomes is made up by recombination of the original ones. The recombination can be carried out in several ways, where the most common is single-point crossover, see Fig. 4.6. Single-point crossover is based on two parent chromosomes that are cut in two pieces at a randomly chosen crossover point. They are then crossed and put together again to form two children chromosomes, that will replace the parent chromosomes in the next generation. The fitness value is used to determine which chromosomes that will be used in the reproduction step, all in order to improve the overall fitness of the population.

4. Mutations are necessary to overcome some problems that otherwise would occur. The most essential problem solved is that if a variable is not selected in any of the original chromosomes it will never be selected in the coming generation. A mutation is simply an inversion of a gene in a chromosome.

The algorithm is repeated until a termination condition is fulfilled. The termination condition can be based on a fitness value that has to be reached or on a convergence criterion, a criterion where the algorithm is terminated when a certain percentage of the chromosomes are identical.

4.3.2 Centering and scaling

The most commonly used preprocessing technique is mean-centering. Mean-centering corresponds to subtracting the mean value of each variable from each instance of that variable. Without mean-centering, the first PC of the multivariate model would go from the origin to the center of the data, i.e. the first PC would describe the average of the data (mean spectrum when used with spectroscopic data). Since the average data usually does not carry any important information, mean-centering is almost always used.

Scaling can be conducted in many different ways, but the most commonly used is unit variance scaling. Unit variance scaling is accomplished by dividing each instance of a variable with the standard deviation of that variable. Since the modeling procedures try to capture variances, variables with a large variance will have greater impact on the model. The unit variance scaling changes this fact and makes all variables have the same chance of effecting the model.

Often both mean-centering and unit variance scaling is used simultaneously and is then called auto scaling. A visualisation of auto scaling is depicted in Fig. 4.7

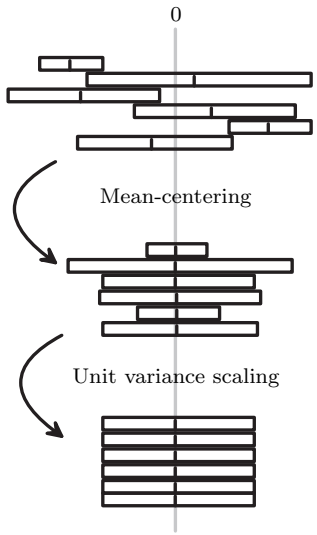


Figure 4.7. *Schematic illustration of auto scaling. The data for each of the six variables are represented by a variance bar and its center.*

4.3.3 Second order derivatives

A commonly used approach to spectral correction is second order derivation[82]. Second order derivation removes not only simple additive offsets, but also first order effects like drift in baseline. Since derivatives are calculated by taking the difference between adjacent spectral points, noise in the data is usually magnified. For this reason, algorithms that smooth the data prior to differentiation are typically employed.

4.3.4 Standard normal variate

Standard normal variate[83] (SNV) has a lot in common with auto scaling, but the calculations are conducted for each sample instead of each variable. In other words the mean value of the sample spectrum is subtracted from the sample spectrum itself. After this the sample spectrum is divided by the standard deviation of the same sample spectrum. The technique aims at reducing the effects of multiplicative effects caused by variations in scattering between samples. Thus SNV pretreated data has been shown to produce better calibration models than calibration models from raw data[83].

4.3.5 Multiplicative scatter correction

Multiplicative scatter correction[84] (MSC) is a relatively simple spectral preprocessing technique that attempts to compensate for the different optical path lengths present when measuring on highly scattering media, such as a pharmaceutical tablet. The technique regresses every spectrum against the average spectrum of the complete data set (or another reference spectrum) and corrects the original spectrum using the slope from the regression. The technique works well for some data sets but has a big drawback, namely that it stipulates that changing scattering properties introduces an error that is linearly dependent on the wavelength. That is contradicted by the results from several research groups

that actually have measured the impact of changing scattering on the measured spectra.

Several variations of MSC have been developed, e.g. the extended inverted signal correction[85;86] that offers a more flexible scatter correction.

4.3.6 Orthogonal signal correction

In situations where the predictors, X, contain a lot of variations that is uncorrelated to the responses, Y, the first couple of PLSCs will be calculated under great influence of the variations in X and only to smaller extent of the correlation between X and Y, resulting in a poor predicting power of the model. In these cases the orthogonal signal correction[87] (OSC) may be a powerful tool.

OSC is a kind of PLS modeling procedure taking place before the calculation of the actual PLS calibration model. The OSC extracts information from the data matrix that is orthogonal to Y and removes that from X before the actual PLS calculation. This makes the variations in X correlated to Y have a larger impact on the first PLSCs.

Chapter 5

Modeling of light transport in turbid media

The most fundamental way to describe light transport is to use Maxwell's equations. These equations describe the interaction of an electromagnetic wave with the dielectric properties of the medium it is traveling through. There are just a few analytical solutions to Maxwell's equations. That means that for most practical applications simplifications and approximations are necessary, leading to the use of different models for light transport in turbid media. The most commonly used models will be described in this chapter.

5.1 Radiative transport theory

When using radiative transport theory[88] the light propagation is treated as a stream of neutral particles (photons), that do not interact with each other, i.e. the phase and the polarisation of the light is not taken into account. The model is therefore simply to be seen as a model of energy transport within the medium. The transport theory is not restricted to light, but has been used in other areas, such as neutron transport and thermodynamics.

The five different contributions to the radiation transport equation are described schematically by looking at the small volume element in Fig. 5.1. When used for calculations regarding light flow the equation is usually expressed as a continuity function of the radiance, $L(\mathbf{r}, \mathbf{s}, t)$ $\left[W/m^2 sr\right]$, i.e. the light intensity per unit area and solid angle. The five different parts of the equation are:

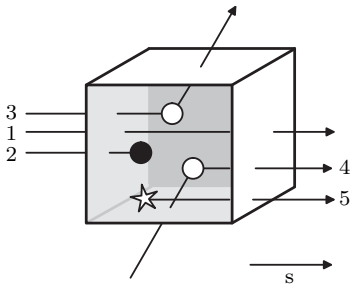


Figure 5.1. *Schematic description of the different parts of the transport equation.*

1. Transmitted light
2. Absorbed light
3. Light scattered from the direction s
4. Light scattered into the direction s
5. Sources within the volume

Together the five parts lead to the equation:

$$\frac{1}{v}\frac{\partial L(\mathbf{r},\mathbf{s},t)}{\partial t} = \underbrace{-\mathbf{s}\cdot\nabla L(\mathbf{r},\mathbf{s},t)}_{1} - \underbrace{(\mu_a+\mu_s)\cdot L(\mathbf{r},\mathbf{s},t)}_{2,3} + \underbrace{+\mu_s \int_{4\pi} L(\mathbf{r},\mathbf{s'},t)\, p(\mathbf{s},\mathbf{s'})\, d\omega'}_{4} + \underbrace{Q(\mathbf{r},\mathbf{s},t)}_{5} \tag{5.1}$$

where μ_a is the absorption coefficient, μ_s is the scattering coefficient, $\mathbf{s}$ is the studied direction and $p(\mathbf{s},\mathbf{s'})$ is the scattering phase function. Although the equation seems relatvely straightforward, complications with coupling between all different directions, $\mathbf{s'}$, boundary conditions and other geometrical aspects make the analytical solutions available for only a few idealised cases[89].

5.2 Diffusion models

When deriving a diffusion model from the radiative transfer equation[90;91], one usually start by expanding the radiance into spherical harmonics and truncate the expansion, the so called P_1 expansion:

$$L(\mathbf{r},\mathbf{s},t) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sqrt{\frac{2l+1}{4\pi}}L_{lm}(\mathbf{r},t)\,Y_{lm}(\mathbf{s}) \approx \approx \frac{1}{4\pi}\left(\phi(\mathbf{r},t) + 3\mathbf{F}(\mathbf{r},t)\cdot\mathbf{s}\right) \tag{5.2}$$

where

$$\phi(\mathbf{r},t) = \int_{4\pi} L(\mathbf{r},\mathbf{s},t)\, d\omega \tag{5.3}$$

and

$$\mathbf{F}(\mathbf{r},t) = \int_{4\pi} L(\mathbf{r},\mathbf{s},t)\,\mathbf{s} d\omega \tag{5.4}$$

The radiance is thereby divided into two parts, the fluence rate, $\phi(\mathbf{r},t)$, which is the light intensity per unit area at position $\mathbf{r}$ at a given time t, and $\mathbf{F}(\mathbf{r},t)$, which is the photon flux describing the

non-isotropic dependence of the radiance.

The same expansion is made for the source term,

$$Q(\mathbf{r},\mathbf{s},t) \approx \frac{1}{4\pi}\left(q_0(\mathbf{r},t) + 3\mathbf{q}_1(\mathbf{r},t)\cdot\mathbf{s}\right) \tag{5.5}$$

giving one isotropic part, q_0, and one non-isotropic part, $\mathbf{q}_1$. In the diffusion approximation only isotropic sources are treated leading to the assumption that $\mathbf{q}_1$ is zero.

Another important equation for the diffusion model is Fick's law giving the photon flux from an isotropic source,

$$\mathbf{F}(\mathbf{r},t) = -D\nabla\phi \tag{5.6}$$

where D is the diffusion coefficient given by,

$$D = \frac{1}{3(\mu_a + \mu_s)} \tag{5.7}$$

Combining equations 5.1 through 5.7 leads to the time-dependent diffusion equation:

$$\frac{1}{v}\frac{\partial\phi(\mathbf{r},t)}{\partial t} - D\nabla^2\phi(\mathbf{r},t) + \mu_a\phi(\mathbf{r},t) = q_0(\mathbf{r},t) \tag{5.8}$$

To get an expression for the light distribution in a medium after the insertion of a short light pulse, an isotropic point source simulated by a Dirac delta function is used:

$$q_0(\mathbf{r},t) = \delta(0,0) \tag{5.9}$$

The solution to Eq. 5.8 using the source term, Eq. 5.9, yields the following expression of the fluence,

$$\phi(\mathbf{r},t) = c(4\pi Dvt)^{-3/2} e^{\left(-\mu_a vt - \frac{r^2}{4Dvt}\right)} \tag{5.10}$$

Eq. 5.10 is also called the Green's function for free diffusion. The equation is only valid for a point source within an infinite homogeneous medium, but it can be used for other simple geometries, such as a semi-infinite geometry or a thin slab.

To find solutions to other geometries, one first has to alter the light source. If a surface is illuminated by light, the source cannot be seen as an isotropic point source at the surface. The most correct way to model the light source in the medium would be to use a line of isotropic light sources, with an exponentially decaying intensity, proportional to $e^{-\mu_{tr}z}$, where z is the depth inside the medium. This will lead to an integration procedure, but the procedure can be simplified by inserting a point source at the depth $z_0 = 1/\mu_s'$ [92]. At this depth the source can be modeled

with the simple Dirac delta function in Eq. 5.9.

The next important factor is to find expressions for the boundary condition at the surface of the medium. Obviously Eq. 5.10 has to be altered in a way so that light exiting through the surface to the non scattering surrounding medium (often air or glass) will not re-enter the medium again. It is, however, not quite correct to assume that the fluence back through the boundary from the surrounding medium is zero, unless the surrounding medium has the same refractive index as the sample[93]. If not, the reflection at the surface due to refractive index mismatch has to be included. The reflection can be included by inserting an extrapolated boundary where the fluence is set to zero[94]. Given a boundary between air and a medium with a refractive index of 1.4, an extrapolated boundary at the distance $z_e \approx 5.5D$ is found to roughly satisfy the surface reflections. This distance will change when the conditions change, but it can be calculated for different refractive indices[95;96]. A second boundary can be added if a slab is to be modeled.

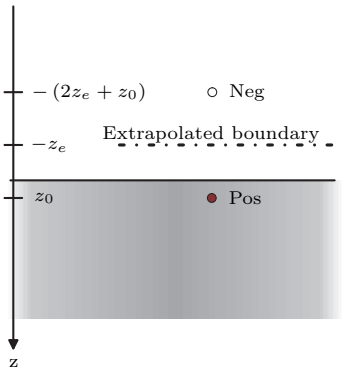


Figure 5.2. *Schematic picture of the first two sources in the extrapolated boundary condition.*

To get the fluence to be zero at the extrapolated surface/surfaces, an infinite number of isotropic sources, both positive and negative have to be added. For practical reasons, the calculations are often conducted using less than 40 sources. A schematic picture of the sources can be seen in Fig. 5.2. The final equation for the time-resolved fluence in cylindrical coordinates for a slab geometry is thereby derived to be[95]:

$$\phi(r,z,t) = \frac{c}{(4\pi Dct)^{3/2}} e^{-\mu_a ct - \frac{r^2}{4Dct}} \times \qquad (5.11)$$
$$\times \left[\sum_{m=-\infty}^{\infty} e^{\left(\frac{-(z-z_{+,m})^2}{4Dct}\right)} - \sum_{m=-\infty}^{\infty} e^{\left(\frac{-(z-z_{-,m})^2}{4Dct}\right)} \right]$$

where

$$z_{+,m} = 2m(d + 2z_e) + z_0$$
$$z_{-,m} = 2m(d + 2z_e) - 2z_e - z_0$$

where d is the thickness of the slab and m is the number of the source.

When predicting the optical properties from a time-resolved measurement, the most common technique is to fit the theoretical diffusion model to the recorded time-resolved data using a Levenberg-Marquardt algorithm with μ_a, μ_s' and an amplitude factor as free parameters[97]. It is also important to incorporate a convolution with the instrumental response function in the

algorithm, see Fig. 5.3. An instrumental response function that is measured incorrectly or a time scale that is slightly shifted, could easily introduce errors of 5%[98], stressing the importance of having a good measurement protocol for the evaluations to work properly. There are other evaluation schemes, where the diffusion equation has been transformed or simplified into forms that can be evaluated without a fitting procedure, e.g. by using two measurements at different distances[99].

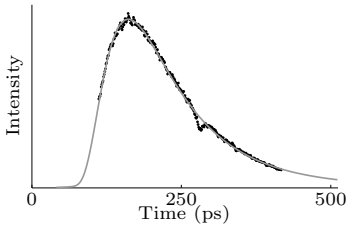


Figure 5.3. *Example of a diffusion evaluation. The required data is represented by the dotted line, while the theoretical diffusion data is shown by the solid gray line.*

For the evaluations of time-resolved measurements to be correct it is important that some requirements are fulfilled. First of all, the reduced scattering coefficient must be much larger than the absorption coefficient, $\mu_s' >> \mu_a$. Furthermore the distance between the light source and detector must be larger than approximately ten transport mean free paths[100], $\mu_{tr}^{-1} > 10$. Both these requirements ascertain that a diffuse light flux will be measured.

5.3 Monte Carlo simulations

Monte Carlo simulations offer a more flexible approach to light propagation than the diffusion models, e.g. it can handle non isotropic light[101] and complex geometries. However, both methods yield similar result in cases where the diffusion solution is valid[102], see Fig. 5.4. Monte Carlo uses a random walk procedure for the photon transport. The principle of the Monte Carlo method is to follow one photon at a time as it is propagating through the medium. The propagation is simulated using probability distributions for the different parameters. The trajectory of the photon is followed until the photon is lost over a boundary or is absorbed. The procedure is then repeated for a large number of photons until the noise in the resulting data is good enough. Depending on the requirements, the number of needed photons may vary from 10,000 for a steady state reflectance simulation, to 10,000,000 for a model of the light fluence inside the medium with a high spatial resolution.

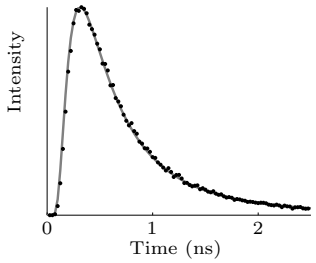


Figure 5.4. *Comparison between a time-resolved reflectance MC simulation (dots) and diffusion model (solid line) using the same optical properties. $\mu_a = 0.01$ cm^{-1}, $\mu_s' = 75$ cm^{-1} and a distance between the source and detector of 0.55 cm.*

One limitation with Monte Carlo simulations is that one simulation has to be done for each set of optical properties, which lead to extensive calculation times to build a library of simulations. Some efforts have been done to try to make the simulation procedure faster. One approach is to scale the results from one Monte Carlo simulation to obtain values for other optical properties[101;103;104].

The basic Monte Carlo model[105] that will be described in this section considers a layered media. It treats the photons as neutral particles, not taking phase and polarisation into account, although

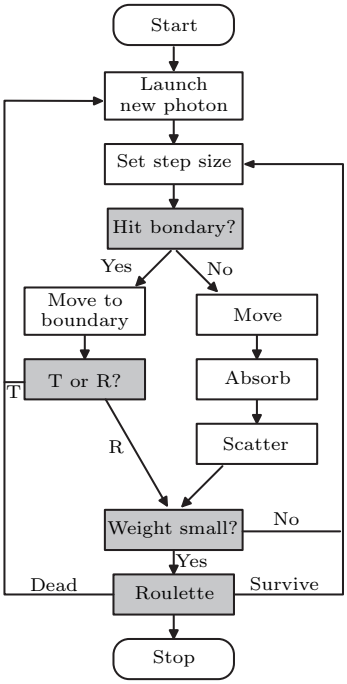


Figure 5.5. *Flowchart describing a Monte Carlo simulation.*

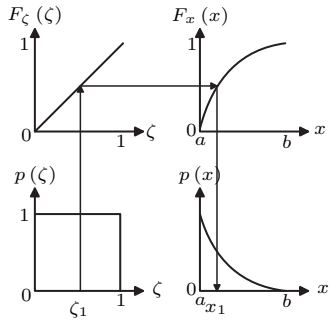


Figure 5.6. *Sampling procedure for non uniform physical properties.*

extended Monte Carlo algorithms have tried to incorporate both those parameters in the models[106;107]. A flowchart showing the main steps in a Monte Carlo simulation is seen in Fig. 5.5. The different steps will be described in the following sections.

5.3.1 Sampling variables

One core task in a Monte Carlo simulation is to properly sample probability distributions. Usually a random number (ζ) generator has a uniformly distribution over the range 0 to 1, that has to be transformed to a non-uniform probability distribution of a physical parameter (x)[108].

The probability function, $p(x)$, of the physical parameter in the entire possible interval $a \leq x \leq b$ is normalised by:

$$\int_a^b p(x)\,dx = 1 \tag{5.12}$$

The cumulative distribution function, $F_x(x_1)$, describes the probability of $a \leq x \leq x_1$

$$F_x(x_1) = \int_a^{x_1} p(x)\,dx \tag{5.13}$$

The cumulative distribution function for the uniformly random number generator becomes:

$$F_\zeta(\zeta_1) = \int_0^{\zeta_1} p(\zeta)\,d\zeta = \zeta_1 \tag{5.14}$$

By setting $F_\zeta(\zeta_1)$ equal to $F_x(x_1)$ the random number generator is able to properly sample the non uniform probability distribution of the physical parameter according to Fig. 5.6. The resulting equation for the sampling procedure is then given by:

$$\zeta_1 = \int_a^{x_1} p(x)\,dx \tag{5.15}$$

5.3.2 Launch of photons

The launch of photons is made by assigning starting values to the photon. The photon is characterised by its position, direction and weight. The weight is a number between one and zero stating how much of the photon is still alive. The use of weights is a way to make the algorithm more time-efficient, compared to

use a photon only as dead or alive. A weight of 1 means that the photon has not yet been partially reflected or absorbed, while a value of 0 means that the photon is dead. In its most basic form the photon is launched at the center of the sample coordinate system with a direction orthogonal to the first layer. The calculation then simulates the incidence of an infinitely narrow beam of light. However, the results from the model can be changed by a simple convolution to simulate results from a top-hat or gaussian light beam with a finite width as light source[109]. Another option to simulate more complex light source geometries is to alter the position and direction of the photons to, for example, simulate light emitted through an optical fibre tip[110].

Another aspect to take into account is the specular reflection at the surface if the light is not launched inside the medium. The specular reflection is calculated using Fresnel's law for reflectance. The reflected fraction is subtracted from the weight of the photon before the photon enters the medium.

5.3.3 Movement of photons

When calculating the movement of a photon inside the medium, first of all the step size has to be calculated. The step size of one photon step is given by the mean free path of propagation, $1/\mu_{tr}$. Since the attenuation is exponential, the step size, s, can be related to a random number, ζ, by:

$$s = \frac{-ln\left(\zeta\right)}{\mu_{tr}} \tag{5.16}$$

The equation satisfies the statement that the average step size will be $1/\mu_{tr}$, since the statistical average of $-ln\left(\zeta\right)$ is one. After the step size is calculated the photon is moved that distance in the medium, in the direction decided by the previous scattering event.

If the photon hits a boundary within the calculated step size the photon is moved to the boundary position and the angle of incidence, α_i, and transmission, α_t, are calculated and the internal reflectance is calculated by[28]:

$$R = \frac{1}{2}\left(\frac{sin^2\left(\alpha_i - \alpha_t\right)}{sin^2\left(\alpha_i + \alpha_t\right)} + \frac{tan^2\left(\alpha_i - \alpha_t\right)}{tan^2\left(\alpha_i + \alpha_t\right)}\right) \tag{5.17}$$

being an average over the two different polarisations. The transmitted fraction is subtracted from the photon weight and the remainder of the photon is moved in the direction of the reflection until the next scattering or reflection event occurs.

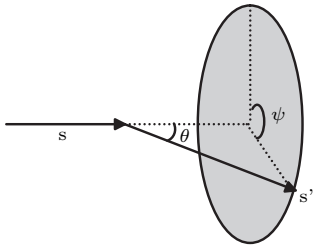


Figure 5.7. *Definition of the scattering angles θ and ψ.*

5.3.4 Absorption and scattering

The absorption is handled by subtracting a fraction of the current photon weight after one step. The fraction subtracted is defined as:

$$\Delta W = W \frac{\mu_a}{\mu_{tr}} \tag{5.18}$$

where W is the photon weight before the step.

The scattering is defined by two angles, the deflection angle, θ, and the azimutal angle, ψ. The definition of the angles are seen in Fig. 5.7.

The probability function for the deflection angle is most often described by the Henyey-Greenstein phase function[6]:

$$p\left(cos\left(\theta\right)\right) = \frac{1-g^2}{2\left(1+g^2-2gcos\left(\theta\right)\right)^{3/2}} \tag{5.19}$$

where the anisotropy, g, is the expectation value for $cos\left(\theta\right)$, ranging from -1 to 1. A value near 1 means an almost totally forward scattering, while a value of 0 indicates isotropic scattering and a value close to -1 indicates total backscattering.

Although the Henyey-Greenstein phase function works properly in most cases other phase functions, e.g. derived from Mie theory[111] or other more empirical functions have been used[112] and compared[113;114].

5.3.5 Photon termination

When using photon weights the photons can, in principal live forever. This would lead to very long computation times, since the photons with low weights do not affect the results in any major way. It is important though to ensure the conservation of energy, i.e. photons cannot just be terminated, since the results then would be incorrect. Instead a roulette procedure is used, where the photons whose weight is below a predefined threshold, e.g. 0.0001, is given one chance in e.g. $m = 10$ to survive. If it does not survive its weight is reduced to 0 and a new photon is launched. If it survives its weight is multiplied by m, and a new step is taken.

Chapter 6

Time-resolved spectroscopy

Time-resolved spectroscopy of turbid media is a measurement technique based on the fact that short light pulses are broadened when propagating through turbid media. The dispersion can be understood as the photons will have different path lengths through the sample. Analysis of the light pulses before and after the passage through the sample allows the calculation of the scattering and absorption properties of the sample. The possibility to measure the absorption and scattering properties of a sample independently is the main advantage of this technique compared to conventional NIR spectroscopy, described in chapter 3, where the recorded data contains a mixture of both effects.

In this chapter instrumentations for time-resolved spectroscopy and their capabilities will be described. A couple of alternative techniques used for the deconvolution of scattering and absorption effects will also be treated.

6.1 Instrumentation

6.1.1 Light sources

Numerous light sources can be used for time-resolved spectroscopy, but this section will focus on the most widely used. An example of a light source used in individual applications, but not described in detail here is Q-switched lasers[115].

Diode lasers

Pulsed diode lasers are, due to their compact size and stable performance, becoming more and more popular as light sources for spectroscopic applications. A laser diode is based on a sandwich

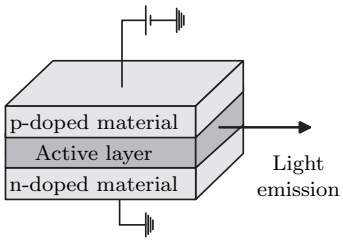


Figure 6.1. *Schematic picture of a diode laser.*

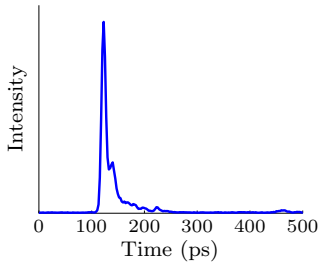


Figure 6.2. *Measured time profile from a pulsed diode laser.*

structure consisting of different semiconductor layers, see Fig. 6.1. The light emission is occurring in the active layer, also called the junction, when holes from the p-doped layer and electrons from the n-doped layer are recombining. The excess energy from the recombination is released as photons, of which the wavelength is determined by the band gap in the semiconductor material. An external applied voltage acts as a pump source, constantly supplying new electrons and holes. To induce a population inversion the external current must exceed a threshold value, above which lasing action is possible. The laser cavity is defined by the end surfaces of the layered material. By modulating the driving current, pulsed emission from the laser diode can be obtained. Commercial pulsed laser diodes are available from 375 nm to 1550 nm having pulse lengths (FWHM) below 100 ps[116;117]. A measurement of the temporal width of a typical pulsed diode laser can be seen in Fig. 6.2

Systems for time-resolved measurements using one single laser diode[118–121] and multiple laser diodes[97;122] have been reported in the literature.

Mode-locked lasers

Within a laser cavity a number of longitudinal modes with different frequencies are allowed. If the laser medium has a very wide gain profile, like the most often used Ti:Sapphire, the number of allowed modes may be as many as 10^5. If all these modes are forced to oscillate in phase, the output light transforms from a continuous beam of light into a train of very short pulses[123].

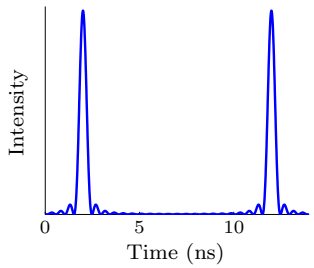


Figure 6.3. *Simulation of time profile of the emission from a mode-locked laser containing 10 longitudinal modes.*

The temporal width of the pulses is defined by the wavelength range spanned by the longitudinal modes of the laser. When using Ti:Sapphire 250,000 modes can be locked, making pulses with a temporal width shorter than 50 fs possible. The time distance between the pulses in the pulse train is defined by the time of a cavity round trip. Different optical set-ups exist, most of them result in a repetition rate of about 80 MHz[124;125]. A simple simulation of the pulses from a mode-locked laser with 10 modes is seen in Fig. 6.3

Due to the short pulses and wavelength tuneability (approximately 700 to 1000 nm) of the mode-locked Ti:Sapphire laser, several groups have reported its use in different time-resolved applications[126–128].

Pulsed continuum sources

If a high power laser is available continuum generation can be conducted by focusing the laser into a cuvette containing a liquid, e.g. water or ethylene glycole[129;130]. Nonlinear effects, such as self-phase modulation of the refractive index due to the high peak power, result in an instantaneous frequency shift. As a result, almost structureless light pulses covering the whole visible and near infrared wavelength region can be obtained and used for time-resolved measurements[131], see Fig. 6.4.

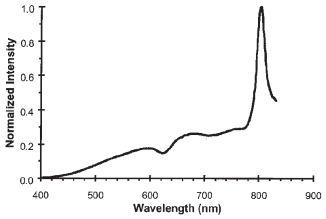


Figure 6.4. *Spectrum of a light pulse created by focusing a high-power laser pulse into a 5 cm-long cuvette containing water*[131].

The conversion efficiency of the continuum generation can be improved by using a tailor made photonic crystal fibre (PCF). There are two fundamental classes of PCFs: PCFs that guide light through the photonic band gap effect and PCFs that guide the light through differences in the refractive index, like a standard optical fibre.

Like in standard optical fibres, index guiding PCFs guide the light by means of total internal reflection between the high refractive index core and the cladding with a lower refractive index. In a PCF the differences in the refractive index is obtained by forming a matrix of different materials (normally air and silica), see Fig. 6.5. By changing the structure of the matrix the properties of the fibre can be tailored with respect to e.g. dispersion and mode structure. One example of the unique features only available in a PCF is their possibility to work as a single-mode fibre for all wavelengths simultaneously[132]. The fact that a PCF can be made with a small core (1-3 μm) in combination with the possibility to have a zero dispersion at the pump laser wavelength, helps the pulses from the pump laser to be confined in a small area and time through the entire length of the fibre. This gives rise to an efficient spectral broadening of the pump pulse into a continuum. The spectral broadening is mainly due to nonlinear effects such as self-phase modulation and stimulated Raman scattering[133]. A continuum generated in a PCF can easily span the entire NIR wavelength range[134]. Continuum generation in PCFs are usually conducted using mode-locked lasers. Some time-resolved systems for characterisation of turbid media using continuum generation in a PCF have been reported in the literature[135;136].

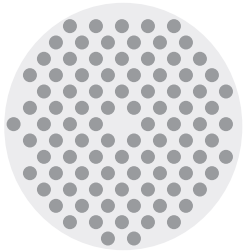


Figure 6.5. *Schematic picture of a PCF. The dark circles represent holes and the light is guided in the core, situated where one hole is left out. The core is usually between 1 and 3 μm.*

6.1.2 Detection schemes

Gated detection

It is possible gate to an intensifier, such as the multi channel plate at very high speeds. The time the intensifier is active can be less than 1 ns, and by combining the amplifier with a camera, a gated view of the light exiting the sample at a certain time-point can

be achieved[137;138]. By synchronising the gating of the intensifier to the light source a time reference is gained. By sequentially delaying the synchronising signal from the light source a series of images can be obtained. These pictures can be combined in order to obtain the time-resolved emission from the sample surface. The time-resolution of a time-gated system cannot compete with the systems decribed in the following sections, but is well suited for imaging applications[139].

Time-correlated single photon counting

The idea of time-correlated single photon counting (TCSPC) is simple and aims at detecting the time delay between two events, e.g. between a light pulse entering a sample and a detected transmitted/reflected photon. By doing this for a large number of transmitted photons a statistical representation of the time-resolved transmission/reflectance is gained.

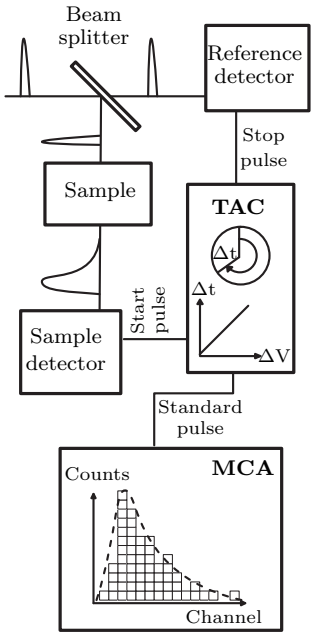


Figure 6.6. *Basic components of a time-correlated single photon counting system.*

A TCSPC system, see Fig. 6.6, consists of two main components, a time-to-amplitude converter (TAC) and a multichannel analyzer (MCA). The light pulses from the light source are fed to the sample and the stretched pulses leaving the sample are collected and sent to the sample detector. When the sample detector detects a photon a response pulse is sent to the TAC. When the pulse arrives to the TAC, an internal clock in the TAC starts and is not stopped until a reference pulse from the light source is detected. The time delay between the pulses is converted into a standard pulse with an amplitude proportional to the time delay. The standard pulse is fed to the MCA that converts the pulse amplitude to a channel and adds one count to that channel. After repeating this procedure for many light pulses, a histogram is formed by the MCA corresponding to the time profile of the measured light pulse. One important constrain in the method is that two photons from the same light pulse must not reach the sample detector. The errors introduced by multiple photons reaching the detector are called pile up errors and tend to skew the time distribution toward earlier times. By reducing the probability of detecting one photon per light pulse to below 0.01, the problem with pile up becomes negligible[140]. With laser sources driven at a repetition rate of 50 to 100 MHz this is not a problem, since most detectors will be overloaded before that limit is reached[141].

The sensitivity of the TCSPC technique is mainly limited by the dark current of the sample detector. Typical values for PMT dark currents are 100 to 500 dark counts per second. There are several sources to the PMT dark current, where the most prominent is thermionic emission from the photocathode and

dynodes[142]. The time-resolution of a TCSPC system is dependent on the timing accuracy of the sample channel only. This accuracy is determined by the transit time spread in the detector for a measured photon and the trigger accuracy in the electronics. The timing accuracy can be up to 10 times better than the detector rise time, resulting in a time-resolution of about 300 ps for regular PMTs while a MCP-PMT based TCSPC system can exhibit time resolutions as low as 30 ps[141].

The compact layout of TCSPC systems has lead to the development of several systems for medical applications[97;122;141].

Streak camera

A schematic picture of a streak camera can be seen in Fig. 6.7. When light impinges on the photocathode it is converted into electrons. The electrons are accelerated by an accelerating mesh and directed toward the sweep electrodes. A high speed voltage sweep is applied to the sweep electrodes, deflecting the electrons. This sweep converts the time difference between the different electrons into a spatial distribution. The voltage sweep is synchronised with the incoming light beam by a trigger signal from the light source. The electrons are multiplied by an MCP before they are converted back to light by a fluorescent screen. The light from the fluorescent screen is imaged onto a CCD, by which the streak image is recorded. By having a slit in front of a large photocathode the time dispersion of a sheet of light can be measured simultaneously by the streak camera. This makes simultaneous time-resolved spatial or spectral measurements possible. A time-resolved spectral measurement is made by simply adding a spectrometer before the streak camera, letting light of different wavelengths impinge on different parts of the entrance slit of the streak camera. The resulting streak image therefore have one time and one wavelength dimension, see Fig. 6.8.

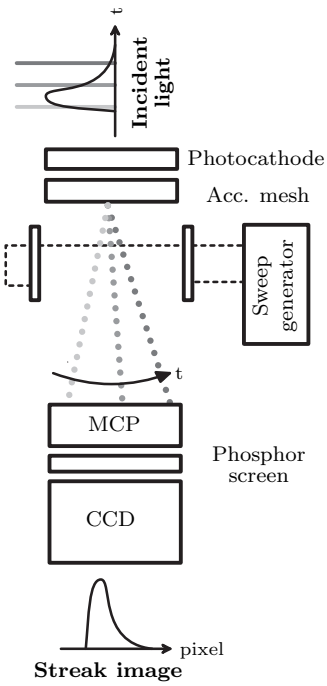


Figure 6.7. *Schematic picture of a streak camera.*

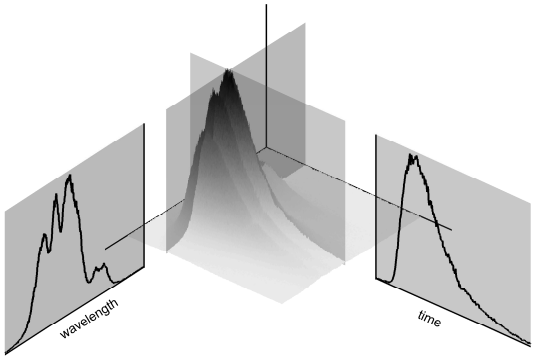


Figure 6.8. *Typical example of a streak image, when the streak camera is combined with a spectrometer to get spectral resolution.*

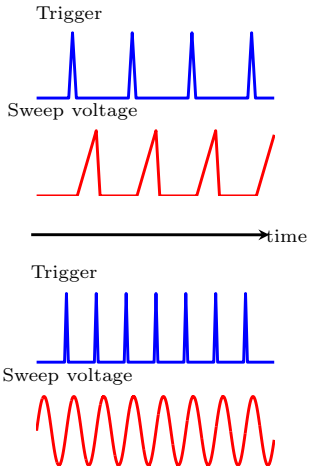


Figure 6.9. *Comparison between single-sweep and synchroscan operation.*

By looking at data from different cross sections of the streak image, the time dispersed light for different wavelengths as well as the spectral shape of light at different times can be studied.

Normally a streak camera can operate the sweep in two different ways, see Fig. 6.9. At low repetition rates, less than 1 kHz, the streak camera sweep is triggered by the triggering signal from the light source. The streak is linear during a specified time, after which the system is blanked before receiving the next trigger signal. The repetition rate of this single-sweep operation is limited by the blanking procedure. At higher repetition rates an operation mode called synchroscan is used. In synchroscan mode the streak is made continuously by applying a sinusoidal modulation to the sweep voltage. The frequency of the sinusoidal signal is controlled by the triggering signal from the light source.

The time-resolution of a streak camera is limited by three factors: transit time spread of photoelectrons, time spread due to the slit width and time spread due to variations in the deflection field. Compared to a TCSPC system, the streak camera has a superior time-resolution with FWHM as low as 1 ps. The sensitivity of the streak camera measurement is limited by the dark current from the photocathode and MCP, and by the read out noise of the CCD. The maximum useful intensity of the incoming light is limited by the fact that the time-resolution gets worse at high intensities. This is thought to be due to the space charge effect. The signal-to-noise ratio is generally in the order of 1000.

To facilitate the high time-resolution of the streak camera, it is often used in combination with light sources with very short pulses, such as the Ti:Sapphire laser.

6.2 Alternative techniques

6.2.1 Spatially resolved spectroscopy

One technique to separate the scattering from the absorption properties of a sample is to use spatially resolved spectroscopy. In a spatially resolved measurement light is usually delivered in one point and the transmitted or reflected light is recorded in several points at different distances from the light source. The detection can be made by a mapping scheme, where the detector is moved to different distances, or by an imaging scheme, where a camera is used.Systems using monochromatic light sources[143] and systems using polychromatic light sources in combination with a spectrometer[144;145] have been reported in the literature.

One benefit of using spatially resolved spectroscopy is that the instrumentation is both cheap and robust, but since the simultaneous measurements are made at different source detector distances the technique is very sensitive to inhomogeneities within the sampling volume. An example of a spatially resolved measurement can be found in Fig. 6.10.

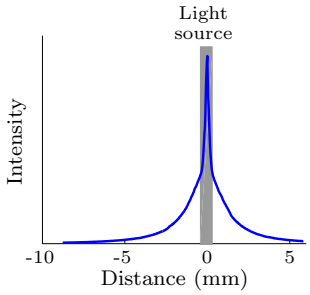


Figure 6.10. *Spatially resolved measurement using a laser at 633 nm and a camera. The measurement was conducted on a pork chop.*

6.2.2 Frequency domain techniques

The purpose of a frequency domain instrument is to detect how the phase and amplitude of the light from an intensity modulated light source is changed when passing through a turbid sample, see Fig. 6.11.

The modulation is usually in the MHz frequency range and is conducted by direct modulation of laser diodes or by Pockels cell modulation of a continuous light source. Laser diodes are the most widely used due to their superior modulation depth of up to 90% compared to 10% for a Pockels cell system. Several laser diodes can be combined in one system using time multiplexing[146].

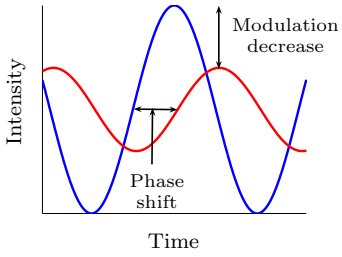


Figure 6.11. *Basic measures in frequency domain techniques.*

As detectors, both PMTs and avalanche photodiodes can be used. The big advantage when using PMTs are their large gain factor, which is 100 times larger than the gain in an avalanche photodiode. The avalanche photodiodes can though detect modulation frequencies up to 1 GHz[147] while the PMTs have problems detecting modulation frequencies above 500 MHz.

There are two main detection schemes used in frequency domain set-ups: homodyne and heterodyne detection. A homodyne system detects the phase shift at the radio frequency, while the heterodyne system shifts the radio frequency down to a lower frequency for phase shift detection. A homodyne detection scheme is less complex, but the more complex heterodyne detection scheme is often used due to its superior phase accuracy and insensitivity to RF noise[148].

Employing one of the models described in chapter 5, the phase shift and amplitude of the detected light can be correlated to the optical properties of the measured sample. By using several different modulation frequencies, an improved prediction of the optical properties can be achieved. Several groups report systems with the ability to predict optical properties with an accuracy better than $\pm 10\%$[147;149–151].

6.3 Pharmaceutical applications

Although NIR spectroscopists working with pharmaceutical samples have been struggling with problems separating scattering from absorption effects in their recorded data, surprisingly few groups have searched for answers in the field of biomedical optics. Time-resolved spectroscopy, and the alternative techniques described above, have been used for biomedical applications for at least 15 years and a solid theoretical foundation for light interaction with turbid media has been developed. A few studies have though been made with a similar main goal as this thesis, namely to use a measurement technique that enables the deconvolution of the scattering and absorption properties of the measured samples.

Shinde et al.[152] used a frequency domain system, based on a diode laser, for quantitative monitoring of a blending process. The application is limited to a binary composition, but by introducing more lasers at other wavelengths the system will have the capability to monitor a blend with several constituents. Pan et al.[153] used a four wavelength frequency domain system to measure the concentration of the active ingredient in a low dose formulation. The measurements were conducted in a blender to follow the mixing procedure. Sun et al.[154] found that the reduced scattering coefficient, μ_s', measured using a frequency domain system, increased linearly with the reciprocal mean particle size of pharmaceutical powders. The same conclusion was made by Torrance et al.[155], that also used their frequency domain system for measurements of the concentration of active ingredients in binary powder mixtures. This shows that the measured concentrations were independent of the samples particle size distribution.

Jiang et al.[156], Richter et al.[157] and Dali et al.[158] all used frequency domain systems to characterise colloidal suspensions, with respect to particle size distribution and volume fraction. In the studies polystyrene spheres were used as scatterers. Pan et al.[156] conducted the same type of study using lactose and resin suspensions.

Acknowledgments

Arbetet med denna avhandling hade inte varit möjligt utan hjälp och uppmuntran från alla i min omgivning.

Jag vill börja med att tacka mina tre handledare. Jonas Johansson för att du gav mig chansen som examensarbetare och för att du trodde på mig och ledde in mig på tanken att doktorera. Utan dig skulle jag inte vara här idag. Stefan Andersson-Engels för all hjälp och allt stöd i det dagliga arbetet. Du har alltid ställt upp, vare sig det handlat om stora vetenskapliga diskussioner eller små reseräkningar som behöver attesteras. Sune Svanberg för din aldrig sinande entusiasm och för den glädje du utstrålar när du får dela med dig av dina enorma kunskaper. Ni har alla tre hjälpt mig att nå fram till mitt mål, och samtidigt hjälp mig att ha skoj längs vägen med paddlingar i Ringsjön, fotbollsmatcher, sommarskolor, resor och utflykter.

Jag vill tacka för det finansiella stöd från AstraZeneca som gjorde detta projekt möjligt. Jag är dessutom tacksam för den hjälp jag fått av alla vid mina besök uppe i Mölndal. Speciellt vill jag tacka Staffan Folestad för din värdefulla vetenskapliga input.

Jag vill tacka alla nuvarande och gamla medlemmar i medicingruppen. Jenny för alla mer, och ibland mindre, intellektuella samtal om morgnarna innan alla andra dykt upp. Ann för ditt, oftast glada, humör, men framför allt för din förmåga att tillsammans med Jenny skratta, ofta och länge. Tomas för trevliga resor och för att jag fått en inblick i fågelskådarnas underbara, men ibland underliga, värld. Thomas för att du får mig att känna mig något mindre förvirrad än jag egentligen är. Marcelo för att du vet vilket hockeylag som är bäst. Sara, Johannes, Johan, Nazila, Florian, Bastiaan och alla andra som hjälp till att göra medicingruppen till det den är.

I would like to take the opportunity to thank Fabien Chauchard for a good collaboration and for teaching me how to select the best

apples when shopping in a supermarket.

Jag vill även tacka alla mina kollegor på avdelningen för atomfysik för alla intresanta diskussioner och fikastunder. Speciellt vill jag tacka Petter för samarbetet i lidarbussen och Laila för all hjälp jag fått med praktiska och administrativa saker.

Jag vill tacka mina föräldrar och min bror för att ni alltid ställt upp för mig. Utan er kärlek och ert stöd hade jag aldrig kommit så här långt.

Jag vill även tacka min nya familj i Trelleborg för att ni fått mig att känna mig som hemma i Skåne.

Sist, och allra mest, vill jag tacka Jenny. Du är det bästa som någonsin hänt mig.

Summary of papers

1. **Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets**
Abrahamsson C et al.

In this work, four different methods for variable selection in partial least square (PLS) regression were studied. The raw data was recorded on a NIR transmission instrument, and the tablets measured were constructed using an experimental design containing five concentration levels of the active substance, two different batches of the active substance as well as one of the excipients and two levels of compressing force. The methods used were genetic algorithm (GA), iterative PLS (IPLS), uninformative variable elimination by PLS (UVE-PLS) and interactive variable selection for PLS (IVS-PLS). All methods improved the predictive abilities of the model compared to a model where the wavelengths were selected manually. For the data set used in this work, IVS-PLS and GA achieved the best results with improvements in prediction error by 20 %.

Contribution:
I conducted major parts of the experimental work, including tablet manufacturing and reference analysis. I made all calculations and wrote the manuscript.

2. **Multi-component chemical analysis of gas mixtures using a continuously tuneable lidar system**
Weibring P et al.

The study introduces the concept of a multi-wavelength lidar system with multivariate analysing techniques for measuring hydrocarbon gas mixtures. The concept was successfully implemented and tested. The procedure uses a genetic algorithm to select what wavelengths to measure and database

spectra for calibration. The calibration procedure is flexible and could be conducted before a measurement campaign without making any reference measurements. The concept was validated using both gas cell measurements as well as measurements on an open ended cell at a distance of 60 m.

Contribution:
I took part in the experimental work. I programmed the genetic algorithm, made all data evaluations and took part in writing the manuscript.

3. **Time and wavelength resolved spectroscopy of turbid media using light continuum generated in a crystal fiber**
Abrahamsson C et al.

The paper is a general description of the novel system for time-resolved measurements, that is the experimental base for this thesis. The system is based on short light continuum pulses generated in an index-guiding crystal fibre, and a spectrometer-equipped streak camera. The system enables spectral recordings of absorption and reduced scattering coefficients of turbid media in the wavelength range 500 - 1200 nm with a spectral resolution of 5 nm and a temporal resolution of 30 ps.

Contribution:
I conducted all experimental work, including setting up the system. I made the data evaluation, but did not construct the evaluation software. I took part in writing the manuscript.

4. **MADSTRESS: A linear approach for evaluating scattering and absorption coefficients of samples measured using time-resolved spectroscopy in reflection**
Chauchard F et al.

In this study present new method for processing time-resolved spectra. The algorithm is based on linear regression and a two dimensional interpolation procedure. The method was tested using time-resolved measurements on apples and fructose powder using the system described in Paper 3. The accuracy of the method was good enough allow the construction of a calibration model predicting the sugar content of the apples.

Contribution:
I took part in all of the experimental work. I took part in writing the manuscript.

5. **Least Squares-Support Vector Machines modelisation for Time Resolved Spectroscopy**
Chauchard F et al.

   In this study, a prediction model for time-resolved data was obtained using a semi-parametric modeling technique : the Least-Squares Support Vector Machines. The main advantage of this technique is that it uses theoretical time dispersion curves during the calibration step. Predictions of the scattering and absorption coefficient was made from measurements on apples. The measurements were made using the system described in Paper 3.

   Contribution:
   I took part in all of the experimental work. I took part in writing the manuscript.

6. **Time-resolved NIR/Vis spectroscopy for analysis of solids: Pharmaceutical tablets**
Johansson J et al.

   Time-resolved spectroscopy in the visible and near infrared regions was used in a feasibility study for analysis of solid pharmaceuticals. The objective of this study was to improve the understanding of the interaction of light with pharmaceutical solids and to investigate the usefulness of the time-resolved spectroscopy as an analytical tool for spectroscopic analysis. In conclusion, the study shows that time-resolved NIR spectroscopy yields more information about solid pharmaceutical samples than conventional NIR spectroscopy.

   Contribution:
   I took a small part the data evaluation and in writing the manuscript.

7. **Time-resolved NIR spectroscopy for quantitative analysis of intact pharmaceutical tablets**
Abrahamsson C et al.

   In this work, time-resolved transmission NIR spectroscopy was utilised to conduct quantitative measurements of intact tablets. The experiments were conducted using the system described in Paper 3. A comparison of the results from the time-resolved technique with the results from transmission NIR spectroscopy was made using tablets containing different concentrations of iron oxide and manufactured with different thicknesses. A PLS model made with data from the time-resolved technique predicted samples 5 times better than a PLS model made data from the conventional NIR

transmission technique. Furthermore, an improvement to predict samples with physical properties outside those included in the calibration set was demonstrated.

Contribution:
I conducted all experimental work, including tablet manufacturing and reference analysis. I made all calculations and wrote the manuscript.

8. **Scatter correction of transmission NIR spectra by photon migration data - quantitative analysis of solids**
Abrahamsson C et al.

The scope of this study was a new methodology to correct transmission NIR data for scattering effects. The method is based on time-resolved measurements, made using the system described in paper 3, and modeling of light transport by diffusion theory. This provides an independent measure of the scattering properties of the samples, and therefore the path length of light. This yields a clear advantage over other pre-processing techniques, where scattering effects are estimated and corrected for by using the shape of the measured spectrum only. PLS calibration models shows that, by using the proposed evaluation scheme, the predictive ability is improved by 50 % as compared to a model based on transmission NIR data only. The method also makes it possible to predict the concentration of active substance in samples with other physical properties than the samples included in the calibration model.

Contribution:
I took part in all of the experimental work, including tablet manufacturing. I took part in the data evaluation and in writing the manuscript.

Bibliography

1. J.M. Hollas. *Modern spectroscopy*. John Wiley & Sons, Chichester, Great Britain, 1997.
2. S. Svanberg. *Atomic and Molecular Spectroscopy - Basic Aspects and Practical Applications Springer series on atoms and plasmas*. Springer Verlag, Heidelberg, Germany, 2004.
3. G. Reich. Near-infrared spectroscopy and imaging:basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews*, **57**:1109–1143, 2005.
4. M.J. Weber. *Handbook of optical materials*. CRC Press, Boca Raton, USA, 2003.
5. C.F. Bohren and D.R. Huffman. *Absorption and scattering of light by small particles*. John Wiley & Sons, New York, USA, 1983.
6. L.G. Henyey and J.L Greenstein. Diffuse radiation in the galaxy. *Astrophysics Journal*, **93**:70–83, 1941.
7. J. Swartling, J.S. Dam, and S. Andersson-Engels. Comparison of spatially and temporally resolved diffuse-reflectance measurement systems for determination of biomedical optical properties. *Applied Optics*, **42**(22):4612–4620, 2003.
8. T. Davies. The history of near infrared spectroscopic analysis: Past, present and future - "from sleeping technique to the morning star of spectroscopy". *Analusis*, **26**(4):M17–M19, 1998.
9. E. Stark and K Luchter. Diversity in nir instrumentation near infrared spectroscopy: 11th international conference: Nir 2003. pages 55–66, Cordoba, Spain, 2003.
10. N.S. Trivedi, A.F. Ghouri, E. Lai, N.K. Shah, and S.J. Barker. Pulse oximeter performance during desaturation and resaturation: A comparison of seven models. *Journal of Clinical Anesthesia*, **9**(3):184–188, 1997.
11. S. Taue, M. Fukuda, H. Yamamoto, Y. Hayasaki, and N. Nishida. Near-infrared spectroscopy probe with position sensor. *Optical Review*, **12**(2):149–154, 2005.
12. S. Oya, H. Inoue, T. Nakade, A. Ogata, M. Tamura, and S. Kato. Near-infrared spectroscopy evaluated as a technique for estimating udder haemodynamics in the lactating cow. *Journal of Veterinary Medicine Series A-Physiology Pathology Clinical Medicine*, **50**(5):230–234, 2003.
13. Y.Q. Lin, G. Lech, S. Nioka, X. Intes, and B. Chance. Noninvasive, low-noise, fast imaging of blood volume and deoxygenation changes in muscles using light-emitting diode continuous-wave imager. *Review of Scientific Instruments*, **73**(8):3065–3074, 2002.
14. A.C. Wilson, T.H. Barnes, P.J. Seakins, T.G. Rolfe, and E.J. Meyer. A low-cost, high-speed, near-infrared hygrometer. *Review of Scientific Instruments*, **66**(12):5618–5624, 1995.
15. M. Sjöholm, G. Somesfalean, J. Alnis, S. Andersson-Engels, and S. Svanberg. Analysis of gas dispersed in scattering media. *Optics Letters*, **26**(1):16–18, 2001.

16. J. Alnis, B. Anderson, M. Sjöholm, G. Somesfalean, and S. Svanberg. Laser spectroscopy of free molecular oxygen dispersed in wood materials. *Applied Physics B*, **77**(6-7):691–695, 2003.
17. S. Morimoto, W.F. McClure, and D.L. Stanfield. Hand-held nir spectrometry: Part 1: An instrument based upon gap-second derivative theory. *Applied Spectroscopy*, **55**(2):182–189, 2001.
18. D.M. Scott. A 2-color near-infrared sensor for sorting recycled plastic waste. *Measurement Science & Technology*, **6**(2):156–159, 1995.
19. K.G. Adamopoulos and A.M. Goula. Application of near-infrared reflectance spectroscopy in the determination of major components in taramosalata. *Journal of Food Engineering*, **63**(2):199–207, 2004.
20. K.G. Adamopoulos, A.M. Goula, and H.J. Petropakis. Quality control during processing of feta cheese - nir application. *Journal of Food Composition and Analysis*, **14**(4):431–440, 2001.
21. H.J. Masterson, G.D. Sharp, and K.M. Johnson. Ferroelectric liquid-crystal tunable filter. *Optics Letters*, **14**(22):1249–1251, 1989.
22. F. Clarke. Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. *Vibrational Spectroscopy*, **34**(1):25–35, 2004.
23. M. Attas, T. Posthumus, B. Schattka, M. Sowa, H. Mantsch, and S.L. Zhang. Long-wavelength near-infrared spectroscopic imaging for in-vivo skin hydration measurements. *Vibrational Spectroscopy*, **28**(1):37–43, 2002.
24. M. Attas, M. Hewko, J. Payette, T. Posthumus, M. Sowa, and H. Mantsch. Visualization of cutaneous hemoglobin oxygenation and skin hydration using near-infrared spectroscopic imaging. *Skin Research and Technology*, **7**(4):238–245, 2001.
25. J. Rantanen, S. Lehtola, P. Ramet, J.P. Mannermaa, and J. Yliruusi. On-line monitoring of moisture content in an instrumented fluidized bed granulator with a multi-channel nir moisture sensor. *Powder Technology*, **99**(2):163–170, 1998.
26. C.D. Tran and M. Bartelt. Performance-characteristics of an acoustooptic tunable filter for optical spectrometry. *Review of Scientific Instruments*, **63**(5):2932–2939, 1992.
27. C. Palmer and E. Loewen. *Diffraction gratings handbook*. Thermo RGL, Rochester, USA, 2002.
28. F.L. Pedrotti and S.J.L.S. Pedrotti. *Introduction to Optics*, volume 2. Prentice Hall Inc., Upper Saddle River, USA, 1996.
29. M.P. Fuller and P.R. Griffiths. Diffuse reflectance measurements by infrared fourier-transform spectrometry. *Analytical Chemistry*, **50**(13):1906–1910, 1978.
30. M. Blanco and M.A. Romero. Near infrared transflectance spectroscopy determination of dexketoprofen in a hydrogel. *Journal of Pharmaceutical and Biomedical Analysis*, **30**(3):467–472, 2002.
31. G.P. Agrawal. *Fiber-optic communication systems*. John Wiley & Sons, New York, USA, 1997.
32. H.G. Hecht. Comparison of continuum models in quantitative diffuse reflectance spectrometry. *Analytical Chemistry*, **48**(12):1775–1779, 1976.
33. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, and C. de la Pezuela. Near-infrared spectroscopy in the pharmaceutical industry. *Analyst*, **123**(8):135R–150R, 1998.
34. C. Pasquini. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, **14**(2):198–219, 2003.
35. O. Svensson, M. Josefson, and F.W. Langkilde. Classification of chemically modified celluloses using a near-infrared spectrometer and soft independent modeling of class analogies. *Applied Spectroscopy*, **51**(12):1826–1835, 1997.
36. K. Kramer and S. Ebel. Application of nir reflectance spectroscopy for the identification of pharmaceutical excipients. *Analytica Chimica Acta*,

420(2):155–161, 2000.

37. A. Candolfi, R. De Maesschalck, D.L. Massart, P.A. Hailey, and A.C.E. Harrington. Identification of pharmaceutical excipients using nir spectroscopy and simca. *Journal of Pharmaceutical and Biomedical Analysis*, **19**(6):923–935, 1999.
38. C.I. Gerhausser and K.A. Kovar. Strategies for constructing near-infrared spectral libraries for the identification of drug substances. *Applied Spectroscopy*, **51**(10):1504–1510, 1997.
39. M. Andre. Multivariate analysis and classification of the chemical quality of 7-aminocephalosporanic acid using near-infrared reflectance spectroscopy. *Analytical Chemistry*, **75**(14):3460–3467, 2003.
40. M.A. Dempster, B.F. MacDonald, P.J. Gemperline, and N.R. Boyer. A near-infrared reflectance analysis method for the noninvasive identification of film-coated and non-film-coated, blister-packed tablets. *Analytica Chimica Acta*, **310**(1):43–51, 1995.
41. M. Laasonen, T. Harmia-Pulkkinen, C. Simard, M. Rasanen, and H. Vuorela. Development and validation of a near-infrared method for the quantitation of caffeine in intact single tablets. *Analytical Chemistry*, **75**(4):754–760, 2003.
42. M.W. Scheiwe, D. Schilling, and P. Aebi. Near infrared spectroscopy analysis of intact pharmaceutical diclofenac coated tablets in transmission. *Pharmazeutische Industrie*, **61**(2):179–183, 1999.
43. A. Eustaquio, P. Graham, R.D. Jee, A.C. Moffatt, and A.D. Trafford. Quantification of paracetamol in intact tablets using near-infrared transmittance spectroscopy. *Analyst*, **123**(11):2303–2306, 1998.
44. N.W. Broad, R.D. Jee, A.C. Moffat, and M.R. Smith. Application of transmission near-infrared spectroscopy to uniformity of content testing of intact steroid tablets. *Analyst*, **126**(12):2207–2211, 2001.
45. P. Chalus, Y. Roggo, S. Walter, and M. Ulmschneider. Near-infrared determination of active substance content in intact low-dosage tablets. *Talanta*, **66**(5):1294–1302, 2005.
46. S.V.S. Tumuluri, S. Prodduturi, M.M. Crowley, S.P. Stodghill, J.W. McGinity, M.A. Repka, and B.A. Avery. The use of near-infrared spectroscopy for the quantitation of a drug in hot-melt extruded films. *Drug Development and Industrial Pharmacy*, **30**(5):505–511, 2004.
47. W. Fountain, K. Dumstorf, A.E. Lowell, R.A. Lodder, and R.J. Mumper. Near-infrared spectroscopy for the determination of testosterone in thin-film composites. *Journal of Pharmaceutical and Biomedical Analysis*, **33**(2):181–189, 2003.
48. J. Gottfries, H. Depui, M. Fransson, M. Jongeneelen, M. Josefson, F.W. Langkilde, and D.T. Witte. Vibrational spectrometry for the assessment of acitve substance in metoprolol tablets: A comparison between transmission and diffuse reflectance near-infrared spectrometry. *Journal of Pharmaceutical and Biomedical Analysis*, **14**:1495–1503, 1996.
49. M. Iyer, H.R. Morris, and J.K. Drennen. Solid dosage form analysis by near infrared spectroscopy: comparison of reflectance and transmittance measurements including the determination of effective sample mass. *Journal of Near Infrared Spectroscopy*, **10**(4):233–245, 2002.
50. S.S. Thosar, R.A. Forbess, N.K. Ebube, Y.X. Chen, R.L. Rubinovitz, M.S. Kemper, G.E. Reier, T.A. Wheatley, and A.J. Shukla. A comparison of reflectance and transmittance near-infrared spectroscopic techniques in determining drug content in intact tablets. *Pharmaceutical Development and Technology*, **6**(1):19–29, 2001.
51. P. Corti, G. Ceramelli, E. Dreassi, and S. Mattii. Near-infrared transmittance analysis for the assay of solid pharmaceutical forms. *Analyst*, **124**:755–758, 1999.
52. J.L. Ramirez, L.K. Bellamy, and R.J. Romanach. A novel method for analyzing thick tablets by near infrared spectroscopy. *AAPS PharmSciTech*, **2**(3):–Article 11, 2005.
53. F.C. Clarke, S.V. Hammond, R.D. Jee, and A.C. Moffat. Determination

of the information depth and sample size for the analysis of pharmaceutical materials using reflectance near-infrared microscopy. *Applied Spectroscopy*, **56**(11):1475–1483, 2002.

54. O. Berntsson, T. Burger, S. Folestad, L.G. Danielsson, J. Kuhn, and J. Fricke. Effective sample size in diffuse reflectance near-ir spectrometry. *Analytical Chemistry*, **71**(3):617–623, 1999.
55. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, and N. Pou. Influence of the procedure used to prepare the calibration sample set on the performance of near infrared spectroscopy in quantitative pharmaceutical analyses. *Analyst*, **126**(7):1129–1134, 2001.
56. M. Andersson, O. Svensson, S. Folestad, M. Josefson, and K.-G. Wahlund. Nir spectroscopy on moving solids using a scanning grating spectrometer–impact on multivariate process analysis. *Chemometrics and Intelligent Laboratory Systems*, **75**(1):1–11, 2005.
57. O. Berntsson, L.G. Danielsson, and S. Folestad. Characterization of diffuse reflectance fiber probe sampling on moving solids using a fourier transform near-infrared spectrometer. *Analytica Chimica Acta*, **431**(1):125–131, 2001.
58. O. Berntsson, L.G. Danielsson, B. Lagerholm, and S. Folestad. Quantitative in-line monitoring of powder blending by near infrared reflection spectroscopy. *Powder Technology*, **123**(2-3):185–193, 2002.
59. A.S. El Hagrasy, H.R. Morris, F. D'Amico, R.A. Lodder, and J.K. Drennen. Near-infrared spectroscopy and imaging for the monitoring of powder blend homogeneity. *Journal Of Pharmaceutical Sciences*, **90**(9):1298–1307, 2001.
60. O. Berntsson, G. Zackrisson, and G. Ostling. Determination of moisture in hard gelatin capsules using near-infrared spectroscopy: Applications to at-line process control of pharmaceutics. *Journal of Pharmaceutical and Biomedical Analysis*, **15**(7):895–900, 1997.
61. J. Rantanen, H. Wikstrom, R. Turner, and L.S. Taylor. Use of in-line near-infrared spectroscopy in combination with chemometrics for improved understanding of pharmaceutical processes. *Analytical Chemistry*, **77**(2):556–563, 2005.
62. T. Herkert, H. Prinz, and K.A. Kovar. One hundred percent online identity check of pharmaceutical products by near-infrared spectroscopy on the packaging line. *European Journal of Pharmaceutics and Biopharmaceutics*, **51**(1):9–16, 2001.
63. Y.X. Chen, S.S. Thosar, R.A. Forbess, M.S. Kemper, R.L. Rubinovitz, and A.J. Shukla. Prediction of drug content and hardness of intact tablets using artificial neural network and near-infrared spectroscopy. *Drug Development and Industrial Pharmacy*, **27**(7):623–631, 2001.
64. M. Otsuka. Comparative particle size determination of phenacetin bulk powder by using kubelka-munk theory and principal component regression analysis based on near-infrared spectroscopy. *Powder Technology*, **141**(3):244–250, 2004.
65. A. Gupta, G.E. Peck, R.W. Miller, and K.R. Morris. Nondestructive measurements of the compact strength and the particle-size distribution after milling of roller compacted powders by near-infrared spectroscopy. *Journal Of Pharmaceutical Sciences*, **93**(4):1047–1053, 2004.
66. F.C. Clarke, M.J. Jamieson, D.A. Clark, S.V. Hammond, R.D. Jee, and A.C. Moffat. Chemical image fusion. the synergy of ft-nir and raman mapping microscopy to enable a move complete visualization of pharmaceutical formulations. *Analytical Chemistry*, **73**(10):2213–2220, 2001.
67. K. Pearson. On lines and planes of closest fit to systems of points in space. *Phil.Mag.*, **6**(2):559–572, 1901.
68. V. Centner, J. Verdu-Andres, B. Walczak, D. Jouan-Rimbaud, F. Despagne, L. Pasti, R. Poppi, D-L. Massart, and O.E. de Noord. Comparison of multivariate calibration techniques applied to experimental nir data sets. *Applied Spectroscopy*, **54**(4):608–629, 2000.
69. B.D. Marx and P.H.C. Eilers. Multivariate calibration stability: a com-

parison of methods. *Journal of Chemometrics*, **16**(3):129–140, 2002.
70. S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**(1-3):37–52, 1987.
71. L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold. *Introduction to multi- and megavariate data analysis using projection methods.* Umetrics AB, Umeå, Sweden, 2005.
72. R.G. Brereton. Introduction to multivariate calibration in analytical chemistry. *Analyst*, **125**(11):2125–2154, 2000.
73. P. Geladi and B.R. Kowalski. Partial least-squares regression - a tutorial. *Analytica Chimica Acta*, **185**:1–17, 1986.
74. S.D. Osborne, R.B. Jordan, and R. Kunnemeyer. Method of wavelength selection for partial least squares. *Analyst*, **122**(12):1531–1537, 1997.
75. F. Lindgren, P. Geladi, S Rännar, and S. Wold. Interactive variable selection (ivs) for pls. part 1: Theory and algorithms. *Journal of Chemometrics*, **8**:349–363, 1994.
76. K.F. Man, K.S. Tang, and S. Kwong. *Genetic Algorithms, concepts and designs.* Springer, London, UK, 1999.
77. J.H. Holland. *Adaption in natural and artificial systems.* MIT Press, Cambridge, USA, 1975.
78. A.S. Bangalore, R.E. Shaffer, G.W. Small, and M.A. Arnold. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least-squares regression: Application to near-infrared spectroscopy. *Analytical Chemistry*, **68**(23):4200–4212, 1996.
79. R. Leardi, R. Boggia, and M. Terrile. Genetic algorithms as a strategy for feature selection. *Journal of Chemometrics*, **6**:267–281, 1992.
80. R. Leardi and A.L. Gonzalez. Genetic algorithms applied to feature selection in pls regression: how and when to use them. *Chemometrics and Intelligent Laboratory Systems*, **41**(2):195–207, 1998.
81. J.M. Roger and V. Bellon-Maurel. Using genetic algorithms to select wavelengths in near-infrared spectra: Application to sugar content prediction in cherries. *Applied Spectroscopy*, **54**(9):1313–1320, 2000.
82. Y.D. Wang, M.J. Lysaght, and B.R. Kowalski. Improvement of multivariate calibration through instrument standardization. *Analytical Chemistry*, **64**(5):562–564, 1992.
83. R.J. Barnes, M.S. Dhanoa, and S.J. Lister. Standard normal variate transformation and de-trending of near- infrared difuse reflectance spectra. *Applied Spectroscopy*, **43**(5):772–777, 1989.
84. P. Geladi, D. MacDougall, and H. Martens. Linearization and scatter correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy*, **39**:491–500, 1985.
85. D.K. Pedersen, H. Martens, J.P. Nielsen, and S.B. Engelsen. Near-infrared absorption and scattering separated by extended inverted signal correction (eisc): Analysis of near-infrared transmittance spectra of single wheat seeds. *Applied Spectroscopy*, **56**(9):1206–1214, 2002.
86. H. Martens, J.P. Nielsen, and S.B. Engelsen. Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry*, **75**(3):394–404, 2003.
87. S. Wold, H. Antii, F. Lindgren, and J. Ohman. Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, **44**:175–185, 1998.
88. J. Mobley, T. Vo-Dinh, and T. Vo-Dinh. *Optical properties of tissue Biomedical Photonics Handbook*, pages 2–1–2–74. CRC Press, Boca Raton, USA, 2005.
89. A.J. Welch and M.J.C. van Gemert. *Optical-Thermal Response of Laser-Irradiated Tissue.* Plenum Press, New York, USA, 1995.
90. W.M. Star, A.J. Welch, and M.J.C. van Gemert. *Diffusion theory of light transport Optical-Thermal Response of Laser-Irradiated Tissue*, pages 131–206. Plenum Press, New York, USA, 1995.
91. S.L. Jacques, A.M. Verga Scheggi, S. Martellucci, A.N. Chester, and

R. Pratesi. *Modelling light transport in tissues Biomedical optical instrumentation and laser-assisted biotechnology NATO Advanced Studies Institute*, pages 21–32. Kluwer Academic Publishers, Dordrecht, 1996.

92. M.S. Patterson, B. Chance, and B.C. Wilson. Time resolved reflectance and transmittance for the non-invasive measurement of optical properties. *Applied Optics*, **28**:2331–2336, 1989.
93. A. Ishimaru. *Wave propagation and scattering in random media*. Academic Press, New York, USA, 1978.
94. M. Keijzer, W.M. Star, and P.R.M. Storchi. Optical diffusion in layered media. *Applied Optics*, **27**(9):1820–1824, 1988.
95. D. Contini, F. Martelli, and G. Zaccanti. Photon migration through a turbid slab described by a model based on diffusion approximation . 1. theory. *Applied Optics*, **36**(19):4587–4599, 1997.
96. R.C. Haskell, L.O. Svaasand, T.-T. Tsay, T.-C. Feng, M.S. McAdams, and B.J. Tromberg. Boundary conditions for the diffusion equation in radiative transfer. *Journal of the Optical Society of America A*, **11**(10):2727–2741, 1994.
97. T. Svensson, J. Swartling, P. Taroni, A. Torricelli, P. Lindblom, C. Ingvar, and S. Andersson-Engels. Characterization of normal breast tissue heterogeneity using time-resolved near-infrared spectroscopy. *Physics in Medicine and Biology*, **50**:2559–2571, 2005.
98. V. Ntziachristos and B. Chance. Accuracy limits in the determination of absolute optical properties using time-resolved nir spectroscopy. *Medical Physics*, **28**(6):1115–1124, 2001.
99. R.K. Wang and Y.A. Wickramasinghe. Fast algorithm to determine optical properties of a turbid medium from time-resolved measurements. *Applied Optics*, **37**(31):7342–7351, 1998.
100. K.M. Yoo, F. Liu, and R.R. Alfano. When does the diffusion approximation fail to describe photon transport in random media? *Physical Review Letters*, **64**(22):2647–2650, 1990.
101. A. Pifferi, P. Taroni, G. Valentini, and S. Andersson-Engels. Real-time method for fitting time-resolved reflectance and transmittance measurements with a monte carlo model. *Applied Optics*, **37**(13):2774–2780, 1998.
102. S.T. Flock, B.C. Wilson, M.S. Patterson, and D.R. Wyman. Monte carlo modeling of light propagation in highly scattering tissues - i: Model predictions and comparison with diffusion theory. *IEEE Transactions on Biomedical Engineering*, **36**(12):1162–1168, 1989.
103. A. Sassaroli, C. Blumetti, F. Martelli, L. Alianelli, D. Contini, A. Ismaelli, and G. Zaccanti. Monte carlo procedure for investigating light propagation and imaging of highly scattering media. *Applied Optics*, **37**(31):7392–7400, 1998.
104. A. Kienle and M.S. Patterson. Determination of the optical properties of turbid media from a single monte carlo simulation. *Physics in Medicine and Biology*, **41**:2221–2227, 1996.
105. L. Wang, S.L. Jacques, and L. Zheng. Mcml - monte carlo modeling of light transport in multi-layered tissues. *Computer Methods and Programs in Biomedicine*, **47**:131–146, 1995.
106. J.C. Ramella-Roman, S.A. Prahl, and S.L. Jacques. Three monte carlo programs of polarized light transport into scattering media: part i. *Optics Express*, **13**(12):4420–4438, 2005.
107. M. Xu. Electric field monte carlo simulation of polarized light propagation in turbid media. *Optics Express*, **12**(26):6530–6539, 2004.
108. I Lux and L. Koblinger. *Monte Carlo particle transport methods: neutron and photon calculations*. CRC Press, Boca Raton, USA, 1990.
109. L.H. Wang, S.L. Jacques, and L.Q. Zheng. Conv - convolution for responses to a finite diameter photon beam incident on multi-layered tissues. *Computer Methods and Programs in Biomedicine*, **54**(3):141–150, 1997.
110. J.R. Mourant, J. Boyer, A.H. Hielscher, and I.J. Bigio. Influence of

scattering phase function on light transport measurements in turbid media performed with small source-detector separations. *Optics Letters*, **21**(7):546–548, 1996.

111. A. Kienle, F.K. Forster, R. Diebolder, and R. Hibst. Light propagation in dentin: influence of microstructure on anisotropy. *Physics in Medicine and Biology*, **48**(2):N7–14, 2003.
112. L.O. Reynolds and N.J. McCormick. Approximate two-parameter phase function for light scattering. *Journal of the Optical Society of America*, **70**(10):1206–1212, 1980.
113. A.N. Yaroslavsky, I.V. Yaroslavsky, T. Goldbach, and H.-J. Schwarzmaier. Influence of the scattering phase function approximation on the optical properties of blood determined from the integrating sphere measurements. *Journal of Biomedical Optics*, **4**(1):47–53, 1999.
114. S.K. Sharma and S. Banerjee. Role of approximate phase functions in monte carlo simulation of light propagation in tissues. *Journal Of Optics A-Pure And Applied Optics*, **5**(3):294–302, 2003.
115. S. Tsuchikawa and S. Tsutsumi. Application of time-of-flight near-infrared spectroscopy to wood with anisotropic cellular structure. *Applied Spectroscopy*, **56**(7):869–876, 2002.
116. Picoquant product catalogue, *http://www.picoquant.com*, 2005.
117. Advanced laser diode systems product catalogue, *http://www.alsgmbh.com/*, 2005.
118. R. Berg, O. Jarlman, and S. Svanberg. Medical transillumination imaging using short-pulse diode lasers. *Applied Optics*, **32**(4):574–579, 1993.
119. M.Q. Brewster and Y. Yamada. Optical-properties of thick, turbid media from picosecond time-resolved light-scattering measurements. *International Journal of Heat and Mass Transfer*, **38**(14):2569–2581, 1995.
120. R. Cubeddu, C. D'Andrea, A. Pifferi, P. Taroni, A. Torricelli, G. Valentini, M. Ruiz-Altisent, C. Valero, C. Ortiz, C. Dover, and D. Johnson. Time-resolved reflectance spectroscopy applied to the nondestructive monitoring of the internal optical properties in apples. *Applied Spectroscopy*, **55**(10):1368–1374, 2001.
121. K. Suzuki, Y. Yamashita, K. Ohta, and B. Chance. Quantitative measurement of optical parameters in the breast using time-resolved spectroscopy. phantom and preliminary in vivo results. *Investigative Radiology*, **29**(4):410–414, 1994.
122. L. Spinelli, A. Torricelli, A. Pifferi, P. Taroni, G.M. Danesini, and R. Cubeddu. Bulk optical properties and tissue components in the female breast from multiwavelength time-resolved optical mammography. *Journal of Biomedical Optics*, **9**(6):1137–1142, 2004.
123. O. Svelto. *Principles of Lasers*, volume fourth edition. Plenum Publishing Corporation, New York, USA, 1998.
124. S. Link, H.A. Durr, and W. Eberhardt. Femtosecond spectroscopy. *Journal of Physics-Condensed Matter*, **13**(34):7873–7884, 2001.
125. P.M.W. French. The generation of ultrashort laser-pulses. *Reports on Progress in Physics*, **58**(2):169–262, 1995.
126. E.B. de Haller, C. Depeursinge, and C.Y. Genton. Resolution of time-resolved breast transillumination: in vitro measurements compared with theoretical predictions. *Optical Engineering*, **34**(7):2084–2091, 1995.
127. A. Torricelli, A. Pifferi, P. Taroni, E. Giambattistelli, and R. Cubeddu. In vivo optical characterization of human tissues from 610 to 1010 nm by time-resolved reflectance spectroscopy. *Physics in Medicine and Biology*, **46**(8):2227–2237, 2001.
128. J.C. Hebden, M. Tziraki, and D. Delpy. Evaluation of the temporally extrapolated absorbance method for dual-wavelength imaging through tissuelike scattering media. *Applied Optics*, **36**(16):3802–3810, 1997.
129. R.R. Alfano. *The supercontinuum laser source*. Springer-Verlag, New York, 1989.
130. R.L. Fork, C.V. Shank, C. Hirlimann, R. Yen, and W.J. Tomlinson. Femotsecond white-light continuum pulses. *Optics Letters*, **8**(1):1–3,

1983.
131. C. af Klinteberg, A. Pifferi, S. Andersson-Engels, R. Cubeddu, and S. Svanberg. In vivo absorption spectrum of tumor sensitizers with femtosecond white light. *Applied Optics*, **44**(11):2213–2220, 2005.
132. T.A. Birks, J.C. Knight, and P.S.J. Russell. Endlessly single-mode photonic crystal fiber. *Optics Letters*, **22**(13):961–963, 1997.
133. G. Genty, M. Lehtonen, H. Ludvigsen, J. Broeng, and M. Kaivola. Spectral broadening of femtosecond pulses into continuum radiation in microstructured fibers. *Optics Express*, **10**(20):1083–1098, 2002.
134. W.J. Wadsworth, A. Ortigosa-Blanch, J.C. Knight, T.A. Birks, T.P.M. Man, and P.S. Russell. Supercontinuum generation in photonic crystal fibers and optical fiber tapers: a novel light source. *Journal of the Optical Society of America B-Optical Physics*, **19**(9):2148–2155, 2002.
135. J. Swartling, A. Bassi, C. D'Andrea, A. Pifferi, A. Torricelli, and R. Cubeddu. Dynamic time-resolved diffuse spectroscopy based on supercontinuum light pulses. *Applied Optics*, **44**(22):4684–4692, 2005.
136. A. Bassi, J. Swartling, C. D'Andrea, A. Pifferi, A. Torricelli, and R. Cubeddu. Time-resolved spectrophotometer for turbid media based on supercontinuum generation in a photonic crystal fiber. *Optics Letters*, **29**(20):2405–2407, 2004.
137. Lavision product catalogue, *http://www.lavision.de*, 2005.
138. Andor technology product catalogue, *http://www.andor.com/*, 2005.
139. C. D'Andrea, D. Comelli, A. Pifferi, A. Torricelli, G. Valentini, and R. Cubeddu. Time-resolved optical imaging through turbid media using a fast data acquisition system based on a gated ccd camera. *Journal of Physics D*, (14):1675–1681, 2003.
140. D.V. O'Connor and D. Phillips. *Time-correlated single photon counting*. Academic Press, London, UK, 1984.
141. Time-correlated single photon counting modules - operating manual, *http://www.becker-hickl.de/*, 2005.
142. Photomultiplier tube handbook, *http://www.hamamatsu.com/*, 2005.
143. J.S. Dam, C.B. Pedersen, T. Dalgaard, P.E. Fabricius, P. Aruna, and S. Andersson-Engels. Fiber optic probe for non-invasive real-time determination of tissue optical properties at multiple wavelengths. *Applied Optics*, **40**(7):1155–1164, 2001.
144. R.M.P. Doornbos, R. Lang, M.C. Aalders, F.W. Cross, and H.J.C.M. Sterenborg. The determination of in vivo human tissue optical properties and absolute chromophore concentrations using spatially resolved steady-state diffuse reflectance spectroscopy. *Physics in Medicine and Biology*, **44**:967–981, 1999.
145. T.H. Pham, F. Bevilacqua, T. Spott, J.S. Dam, B.J. Tromberg, and S. Andersson-Engels. Quantifying the absorption and reduced scattering coefficients of tissue-like turbid media over a broad spectral range using a non-contact fourier interferometric, hyperspectral imaging system. *Applied Optics*, **39**(34):6487–6497, 2000.
146. B. Chance, M. Cope, E. Gratton, N. Ramanujam, and B. Tromberg. Phase measurements of light absorption and scatter in human tissue. *Review of Scientific Instruments*, **69**(10):3457–3481, 1998.
147. T.H. Pham, O. Coquoz, J.B. Fishkin, E. Anderson, and B.J. Tromberg. Broad bandwidth frequency domain instrument for quantitative tissue optical spectroscopy. *Review of Scientific Instruments*, **71**(6):2500–2513, 2000.
148. N. Ramanujam, C. Du, H.Y. Ma, and B. Chance. Sources of phase noise in homodyne and heterodyne phase modulation devices used for tissue oximetry studies. *Review of Scientific Instruments*, **69**(8):3042–3054, 1998.
149. T. Tu, Y. Chen, J. Zhang, X. Intes, and B. Chance. Analysis on performance and optimization of frequency-domain near-infrared instruments. *Journal of Biomedical Optics*, **7**(4):643–649, 2002.
150. S.J. Madsen, E.R. Anderson, R.C. Haskell, and B.J. Tromberg. Portable,

high-bandwidth frequency-domain photon migration instrument for tissue spectroscopy. *Optics Letters*, **19**(23):1934–1936, 1994.

151. M. Gerken and G.W. Faris. High-precision frequency-domain measurements of the optical properties of turbid media. *Optics Letters*, **24**(14):930–932, 1999.
152. R.R. Shinde, G.V. Balgi, S.L. Nail, and E.M. Sevick-Muraca. Frequency-domain photon migration measurements for quantitative assessment of powder absorbance: A novel sensor of blend homogeneity. *Journal Of Pharmaceutical Sciences*, **88**(10):959–966, 1999.
153. T.S. Pan, D. Barber, D. Coffin-Beach, Z.G. Sun, and E.M. Sevick-Muraca. Measurement of low-dose active pharmaceutical ingredient in a pharmaceutical blend using frequency-domain photon migration. *Journal Of Pharmaceutical Sciences*, **93**(3):635–645, 2004.
154. Z. Sun, S. Torrance, F.K. McNeil-Watson, and E.M. Sevick-Muraca. Application of frequency domain photon migration to particle size analysis and monitoring of pharmaceutical powders. *Analytical Chemistry*, **75**(7):1720–1725, 2003.
155. S.E. Torrance, Z. Sun, and E.M. Sevick-Muraca. Impact of excipient particle size on measurement of active pharmaceutical ingredient absorbance in mixtures using frequency domain photon migration. *Journal Of Pharmaceutical Sciences*, **93**(7):1879–1889, 2004.
156. H.B. Jiang, J. Pierce, J. Kao, and E. SevickMuraca. Measurement of particle-size distribution and volume fraction in concentrated suspensions with photon migration techniques. *Applied Optics*, **36**(15):3310–3318, 1997.
157. S.M. Richter and E.M. Sevick-Muraca. Characterization of concentrated colloidal suspensions using time-dependent photon migration measurements. *Colloids and Surfaces A-Physicochemical and Engineering Aspects*, **172**(1-3):163–173, 2000.
158. S.S. Dali, J.C. Rasmussen, Y.Q. Huang, R. Roy, and E.M. Sevick-Muraca. Particle sizing in dense suspensions with multiwavelength photon migration measurements. *Aiche Journal*, **51**(4):1116–1124, 2005.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemometrics and Intelligent Laboratory Systems 69 (2003) 3–12
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets

Christoffer Abrahamsson[a,*], Jonas Johansson[a], Anders Sparén[a,1], Fredrik Lindgren[b]

[a] AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden
[b] Umetrics AB, Storgatan 2, SE-211 42 Malmö, Sweden

Received 4 October 2002; received in revised form 8 April 2003; accepted 12 April 2003

Abstract

Near infrared (NIR) transmission spectroscopy is a promising method for fast quantitative measurements on pharmaceutical tablets, but there are still some problems to overcome in order to incorporate the technique as a control tool in tablet production. The main problem is the limited precision for multivariate calibrations based on NIR transmission data. The precision is affected by several factors, where one of the most important is which variable to include in the multivariate calibration model.

In this work, four different methods for variable selection in partial least square (PLS) regression were studied and compared to a calibration made with manually selected wavelengths. The methods used were genetic algorithm (GA), iterative PLS (IPLS), uninformative variable elimination by PLS (UVE-PLS) and interactive variable selection for PLS (IVS-PLS).

All methods improved the predictive abilities of the model compared to the model where the wavelengths were selected manually. For the data set used in this work, IVS-PLS and GA achieved the best results with improvements in prediction error by 20%, but further measurements and investigations have to be made before any general conclusion can be drawn.

Keywords: Near-infrared spectroscopy; Transmission; Precision; Pharmaceuticals; Intact tablets; Variable selection

1. Introduction

When using analytical calibrations, it is important to reduce all possible sources of error in order to get as good results as possible. When working with near infrared (NIR) spectroscopy on intact tablets, these sources can be divided into three major categories: instrumentation, reference analysis and data analysis.

Since the first commercial NIR instruments were introduced in the pharmaceutical industry, great effort has been made to further develop the instrumentation to encounter the high demands set by the controlling authorities. Using the transmission mode instead of the more often used reflectance mode has shown to give improvements not only in precision but also in robustness [1,2]. Not only the measuring mode is thought to be important but also the sample presentation has a large effect on the results [3], as well as other optical components in the instrument [4,5].

* Corresponding author. Present address: Department of Physics, Lund Institute of Technology, PO Box 118, SE-221 00 Lund, Sweden. Tel.: +46-46-2223120; fax: +46-46-2223177.
E-mail address: christoffer.abrahamsson@fysik.lth.se (C. Abrahamsson).
[1] Present address: AstraZeneca Tablet Production Sweden, SE-151 85 Södertälje, Sweden.

0169-7439/$ - see front matter © 2003 Published by Elsevier B.V.
doi:10.1016/S0169-7439(03)00064-9

As in all spectroscopic calibrations, it takes a lot of samples in order to compensate for limited precision in the reference analysis [6,7]. The most common technique to determine the concentration of the active substance in a pharmaceutical tablet is high performance liquid chromatography (HPLC). If new reference analysis techniques with better precision than HPLC could be found, it would be beneficial when building calibration models.

NIR calibrations are often made with multivariate data analysis, e.g. partial least squares (PLS) regression. Many spectral pre-treatment methods have been developed to reduce the effects of variations in the spectral data that are not related to the chemical variations in the samples [8–10]. These methods often improve the calibrations, but they do not take into account that there might be spectral regions that do not contain any information about the chemical variations in the samples. In fact, one of the major problems in multivariate data analysis is to decide which variables to include in a calibration model in order to achieve the best performance. Classically, this selection is made from the basic knowledge about the spectroscopic properties of the sample, but it has been shown that there are mathematical strategies for variable selection that are more efficient. These algorithms work in different ways and they have been developed for different applications. It is therefore hard to get an overview of which algorithm is suited for what kind of data.

The goal of this paper was to compare four of these algorithms in order to improve the precision of calibration models built on data from NIR transmission measurements on intact tablets.

2. Theory

All algorithms used in this work have been developed from their original forms. The genetic algorithm (GA) [11–13], iterative PLS (IPLS) [14] and uninformative variable elimination by PLS (UVE-PLS) [15] have just undergone small adjustments in order to work with windows of several neighbouring wavelengths instead of single wavelengths. The reason for this is that neighbouring wavelengths in spectroscopic data are often highly correlated, especially in the near infrared region. The algorithm for interactive variable selection for PLS (IVS-PLS) used in this work is however quite different from the one presented by Lindgren et al. [16]. The new algorithm works with windows instead of single wavelengths and also has the ability to use two different weighing modes.

2.1. The genetic algorithm

Genetic algorithms (GAs) are variable selection methods based on the principles of genetics and natural selection. Several articles about the use of GAs have been published [11,12]. Only a description of the GA implementation used in this research will be provided here.

The GA consists of four basic steps, where steps 2–4 are repeated until the termination criterion is reached:

(1) To allow easy mathematical treatment of a chromosome, a coding is necessary. This is solved by representing each variable/window (gene) with a binary code in a vector (chromosome) with one cell for each variable/window. The original chromosome is then perturbed randomly to make a number of chromosomes, the initial population.
(2) For each chromosome, the response associated with the corresponding experimental conditions is evaluated. This is done by making a PLS model for each chromosome. The model is then evaluated by cross validation in order to get a numeric value describing the quality of the model. This value is known as the fitness value and is the criterion for guiding the GA to the global optimum.
(3) The reproduction step creates a new population that can be considered as the next generation. The new generation of chromosomes is made up by recombination of the original chromosomes. This recombination is made by single-point crossover, which is based on two parent chromosomes that are cut in two pieces, each at a randomly chosen point. They are then crossed and put together again to form two children chromosomes that will replace the parent chromosomes in the new generation. The chromosomes with a high fitness value have a higher probability to reproduce than a chromosome with a low fitness value, all in order to improve the overall fitness of the population.
(4) Mutations are necessary to overcome some problems that may occur. The most essential

problem to be solved is that if a variable should not be selected in any of the original chromosomes, it would never be selected in the coming generation if mutations were not present. A mutation is simply an inversion of a gene in a chromosome. The mutation rate is user defined and often set to 0.001–0.01.

The algorithm is repeated until the termination condition is fulfilled. The termination condition is based on a convergence criterion, where the algorithm is terminated when a certain percentage of the chromosomes are identical.

2.2. Iterative PLS

Iterative PLS (IPLS) is a variable selection method that is designed to start with a small number of variables/windows and subsequently add new variables/windows to or remove original ones from the data set if that improves the model [14].

The method used consists of four steps:

(1) The original variables/windows are selected randomly.
(2) An ordinary PLS calculation, using the selected wavelengths, is made and the model is evaluated using cross validation.
(3) The variable/window to be added or withdrawn from the model is chosen randomly and a new PLS model is built and evaluated by means of cross validation.
(4) If the new cross validation value (root mean square error of cross validation, RMSECV) is lower than the original, the new set of variables replaces the original. If the new cross validation value is higher than the original, the original set of variables is kept.

The algorithm is terminated when every variable/window has been tested once without giving any improvements.

2.3. Uninformative variable selection by PLS

Uninformative variable elimination by PLS (UVE-PLS) is a method of variable selection based on an analysis of the *b* regression coefficients. One advantage of UVE-PLS compared to other variable selection methods is that it is user independent and therefore does not present any configuration problems [15].

The algorithm used consists of five basic steps:

(1) The first step involves building up a matrix of random numbers (0.0–1.0) of the same size as the *X* matrix. The numbers are multiplied with a small constant (10^{-10}) to make them at least one order of magnitude smaller than the imprecision of the instrument. The multiplication retains the variation of the variables but makes their influence on the model negligible.
(2) The new matrix is added to the original one to form an extended matrix with twice as many variables as the original one.
(3) As many models as there are samples are made, where every model contains all but one sample. This leads to a matrix of *b* values with as many rows as samples and one column for each variable, both original and random.
(4) The *c* values are calculated as the average of the *b* values of each column divided by the standard deviation of that column. When windows are used, the *c* value of one window is calculated as the mean of the *c* values of the variables in that window.
(5) The cut-off limit is set by the largest *c* value of the randomly calculated variables/windows. Every variable/window with equal or lower *c* value is left out of the final model. This means that all random variables and all original variables that are assumed to contain nothing but noise are eliminated.

2.4. Interactive variable selection for PLS

The selection of variables in interactive variable selection for PLS (IVS-PLS) is made dimension-wise instead of model-wise as in the other methods. This means that variable selection is carried out in every PLS dimension in order to maximise the predictive ability of the final model. The algorithm uses cross validation as guidance to the best variable selections.

The algorithm contains four basic steps that are repeated for each PLS component (PLSC) (see Fig. 1).

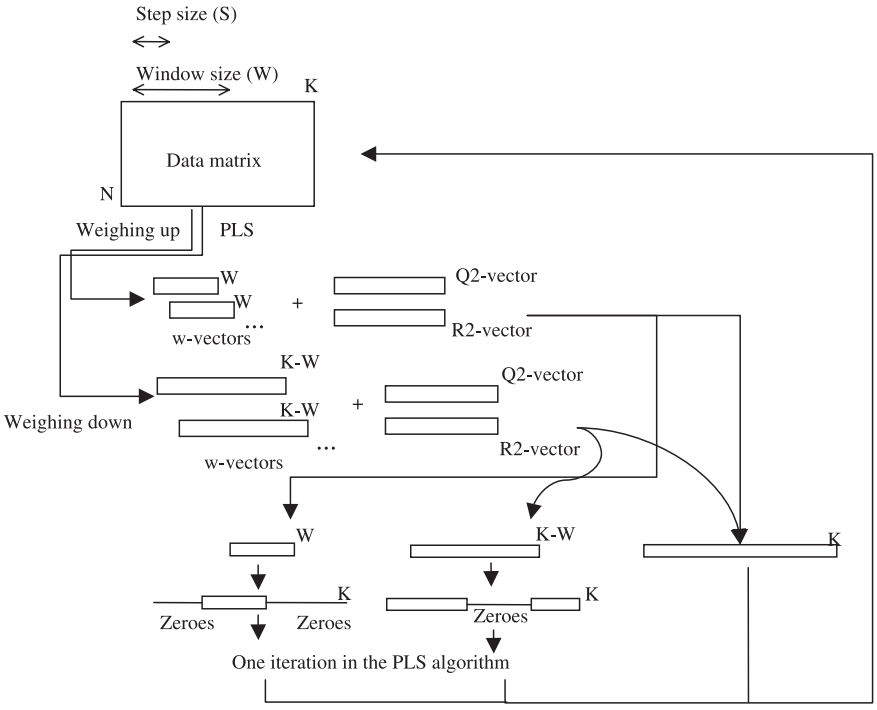


Fig. 1. Schematic picture of the modified IVS-PLS algorithm. The algorithm works with two weighing modes and adjustable window sizes.

(1) The key to a successful IVS-PLS calculation is to find the best set-up for the in-parameters in each PLSC. There are four parameters that guide the algorithm:

(i) Window size. The window size is very much dependent on the type of data to be analysed. It also varies between PLSCs, e.g. as PLSC1 finds a maximum in Q^2 in a window of 100 wavelengths, PLSC2 might find a maximum in Q^2 in a window of 300 wavelengths. The window size can therefore be varied between PLSCs.

(ii) Step size. The step size tells how many wavelengths the window should move forward between each calculation. If a step size of 1 is selected, every possible location of the window in the data matrix will be tested, while the step size selected is the same as the window size one would calculate PLSCs for adjacent windows. The only benefit when using a step size larger than 1 is faster calculations.

(iii) Weighing up or down. See Section 2.

(iv) Cross validation segments. The algorithm is evaluated by means of cross validation, and the user defines the number of cross validation segments.

(2) The calculation of the PLSC of a window can be made in two different ways, weighing up or down. The first one means to calculate a PLSC with the variables in the window included and the rest excluded. In the second method, the variables in the window are excluded and the rest included. The calculation results in one separate PLS component for each possible window. For each component, an R^2 value and a Q^2 value are calculated, where Q^2 is calculated as the cross-validated R^2.

(3) The evaluation of the calculation is made by looking at a plot of R^2 and Q^2 versus the number of the window. By investigating the plot, one should decide if any of the window-based components presents an improvement compared to a PLSC including all variables, if it is best to include all variables in the PLSC or if it is best to restart the iteration by

selecting another set of in-parameters. There are a few rules to regard while looking at the evaluation plots:

(i) Only pay attention to windows with improvements in the Q^2 value by 10% or more. Smaller improvements may be of local nature and only appear in the training set.
(ii) Leave at least the square root of the number of variables in each dimension.
(iii) Dimensions with $Q^2>1$ should be left unchanged.

These general rules often do not give enough guidance to find a simple solution, if one exists, but need to be complemented by common sense and sometimes a trial and error approach.

(4) The calculation of the final PLSC can be done in three different ways, depending on the in-parameters and the result of the evaluation. If the evaluation shows that the model with all variables gives the best PLSC, the PLSC will be calculated as an ordinary PLSC. If the evaluation shows that a window PLSC gives the best results, the calculation can be made in two different ways depending if weighing up or down is used. If weighing up is used, the weights from the window-based model will be lengthened with zeros at both ends, until the vector is of the length of the original matrix. In the same way, the w-vector from the weighing down PLSC will be lengthened with a vector of zeros, with the length of the window size, inserted at the location of the selected window. The final PLSC is calculated by performing one iteration of the PLS algorithm with the new w-vector as the starting point.

3. Experimental

3.1. Data set

The data set consisted of 360 spectra. Tablets (120, three from each batch) were measured three times each. Between the measurements, each tablet was moved 10 steps in the sample holder, e.g. measured in wells 3, 13 and 23.

3.1.1. Tablets

Homogenous tablets were manufactured according to an experimental design. The design contained five concentration levels between 90% and 110% of the nominal content of the active substance, two different batches of the active substance as well as one of the excipients and two levels of compressing force, giving a total of 40 different batches.

3.1.2. NIR apparatus

All NIR spectra were measured with a NIRSystems 6500 monochromator, with a NIRSystems InTact MultiTab Analyzer presenting the samples. The instrument is a holographic grating instrument with a tungsten-halogen lamp as the light source and a single indium–gallium–arsenide (InGaAs) detector. The spectra were collected in the wavelength range of 600–1900 nm with a spectral resolution of 10 nm and a data point spacing of 2 nm. The sample holder consisted of a carousel with 30 sample wells.

3.1.3. Reference analysis

The UV spectrophotometer used was an HP 8453 (Hewlett Packard Sverige, Spånga, Sweden). The HP

Table 1
Summary of the input parameters and the results of the variable selection made by the genetic algorithm

Model	Input parameters			Responses			
	Window size	Population size	Percent wavelengths included initially	Number of wavelengths	Components	RMSEP	RMSECV
GA1	25	64	25	150	4	1.072	0.671
GA2	25	64	50	180	4	1.018	0.689
GA3	25	256	25	140	4	1.073	0.672
GA4	25	256	50	160	4	1.058	0.673
GA5	100	64	25	165	4	1.148	0.83
GA6	100	64	50	165	4	1.120	0.808
GA7	100	256	25	260	3	1.090	0.829
GA8	100	256	50	325	4	1.159	0.873
Original	–	–	–	–	3	1.194	0.886

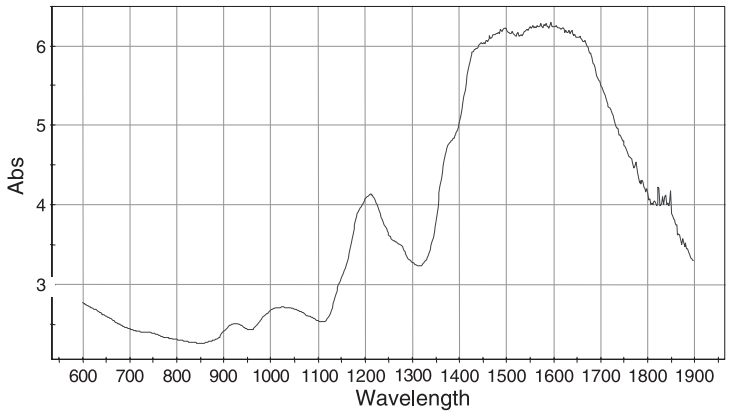


Fig. 2. NIR transmission spectrum of the tablet. In the region above 1350 nm, the signal becomes noisy due to increased absorption in the tablet and the decreased sensitivity of the detector. The region below 800 nm suffers from instrumental properties that make the spectra less dependable.

8453 is a diode array instrument that collects spectra in the wavelength range of 190–1100 nm with a resolution of 1 nm. In this study, the absorption was measured at 274 nm and the background at 550 nm.

Each tablet was weighed and transferred into a 25 ml volumetric flask. Twenty milliliters of phosphate buffer pH 3.0 was added and the flask was shaken vigorously in a mechanical shaker for 30 min until the tablet was completely disintegrated. The samples were diluted to volume with phosphate buffer pH 3.0 and left to sedimentate for 3 h. Five milliliters of the clear solution was transferred to a 50 ml volumetric flask that was filled to volume with phosphate buffer pH 3.0.

3.2. Modelling and variable selection

All multivariate models were made in Simca-P 8.0 (Umetrics, Umeå, Sweden) and programming and variable selection in Matlab 5.3 (The Math Works, Natic, MA).

All spectra were mean centered before calculations and the number of PLSCs selected in the models were as many as Simca-P 8.0 found suitable. The program uses the cross-validated predicted fraction for both X and Y to find the optimal number of PLSCs.

The GA used in this work was from PLS Toolbox 2.0 (Eigenvector Research, Manson, WA) and it was run eight times according to an experimental design where the population size, the window size and the number of variables included in each chromosome at initiation were varied (see Table 1). The algorithm was terminated after 100 generations or when 80% of the chromosomes was identical. The mutation rate was set to 0.005, single point crossover was used and the fitness value for every chromosome was calculated as cross validation value of the PLS model made with four PLSCs. Locking the number of PLSCs may not give the optimal wavelength selection since different wavelength sets normally have different number of significant PLSCs, but this approach gives a faster evaluation scheme and also helps to minimise the risk of overfitting the models.

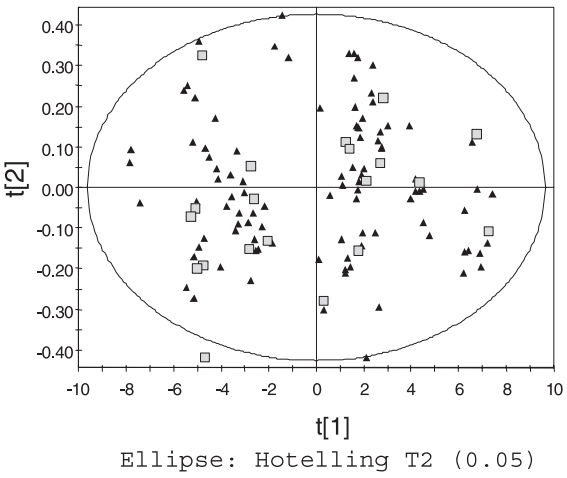


Fig. 3. Score plot for component 1 ($R2 = 0.997$) and component 2 ($R2 = 0.002$) in a PCA model with all samples included. The samples marked as squares are the test set that is used for external validation.

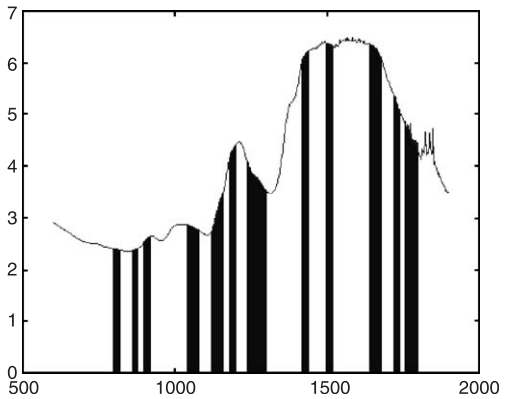


Fig. 4. Overview of the best variable selection obtained with GA, GA2. Wavelengths included are marked black. The model based on these wavelengths lowered the prediction error with 15% compared to the original model.

The variable selection with IPLS was made using randomly selected start variables and cross validation was used for evaluation. The window size and the number of wavelengths included initially were varied at three levels each.

The UVE-PLS algorithm was run three times with different window sizes.

4. Results and discussion

4.1. Original model

The spectral features around 1200 nm are characteristic for the active substance (see Fig. 2). This region is the second overtone region of C–H stretch vibrations. It has earlier been found that the best wavelength range for PLS calibrations is 800–1350 nm for a similar formulation [1].

Three spectra were scanned for each sample. A PLS model with all 360 spectra was calculated and the values predicted by the model using cross validation were analysed. The relative standard deviation, at a single wavelength of the three spectra from the same tablet, varied between 0.13% and 1.72% with a mean value of 0.58%.

To overcome problems with cross validation that may occur when more than one spectrum of each sample is present, mean spectra were calculated. This calculation reduced the number of spectra from 360 to 120, but when the two models, one with all 360 spectra and one with the calculated 120 spectra, were compared, no deterioration in the predictive ability of the model of the mean spectra was seen.

A PCA model was then made with all 120 spectra where no outliers were detected. From this PCA model, a test set of 20 samples was selected. The test set was selected to cover most of the variation in the first two principal components and was used as source for external validation in models where variable selection was accomplished (Fig. 3).

In order to have a measure of the quality of the variable selection algorithms, a model was built with the previously found suitable wavelength region. This original model was a PLS model with three PLSC, which gave a root mean square error of prediction (RMSEP) of 1.194 mg/tablet when predicting the test set and a corresponding root mean

Table 2
Results and input parameters from the variable selection made by iterative PLS

Model	Input parameters		Responses			
	Window size	Wavelengths included initially	Number of wavelengths	Components	RMSEP	RMSECV
IPLS1	10	50	50	2	1.157	0.79
IPLS2	10	100	110	3	1.124	0.8
IPLS3	10	650	70	2	1.118	0.787
IPLS4	25	50	100	2	1.132	0.79
IPLS5	25	100	100	2	1.132	0.79
IPLS6	25	650	100	2	1.132	0.79
IPLS7	50	50	150	3	1.168	0.834
IPLS8	50	100	150	2	1.305	0.886
IPLS9	50	650	150	2	1.305	0.886
Original	–	–	–	3	1.194	0.865

square error of cross validation (RMSECV) of 0.865 mg/tablet.

4.2. Variable selection

One can see (in Table 1) that all models after the variable selection by GA gave slightly better RMSEP values than the original model, which implies better predictive power. The improvements in RMSEP were between 2% and 15%, where the largest improvements were seen in the models where the GAs had been working with small windows. GA2 (Fig. 4) had the lowest RMSEP value of the models made with this variable selection technique.

Seven out of nine models built by data after variable selection by IPLS exhibited improvements in the RMSEP values, compared to the original model (Table 2). The largest improvements (6%) were obtained when small windows were applied (Fig. 5).

One interesting result is that the algorithm selected the same variables in all three models with a window size of 25 wavelengths (IPLS4–IPLS6) although they started with different and randomly selected wavelengths.

After using the UVE algorithm, it can be seen that the model with the smallest window size (UVE1) was the most successful (Fig. 6), but all models gave lower RMSEP values than the original model. The improvements were in the range of 4–7% (Table 3).

The IVS-PLS algorithm was tested using many different combinations and the most successful is summarised in Table 4 and Fig. 7. Three out of five models gave lower RMSEP values than the original model, where IVS1 showed the best improvement, almost 20%.

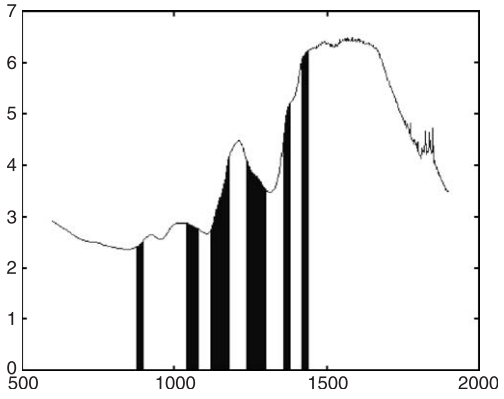


Fig. 5. Wavelengths included in the model IPLS2 are marked black. The improvements in RMSEP achieved with IPLS were 6%.

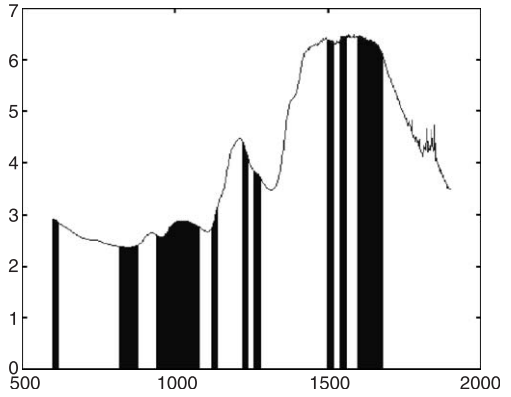


Fig. 6. Overview of the best UVE-PLS model, UVE1. The selected wavelengths are marked black. The improvement in RMSEP, compared to the original model, was 7%.

All variable selection techniques improved the predictive ability of the model, but two of the techniques showed larger improvements than the others, IVS-PLS and GA. These are also the two algorithms that needed most input from the user. The GA is known for its enormous configuration challenge and IVS-PLS requires the user to choose the region and weighing mode for every PLSC, while UVE-PLS and IPLS are more automated.

When comparing the time needed for the computer to make one run of the algorithms, there are two algorithms that need far more time than the others, GA and IPLS.

All techniques, except IVS-PLS, worked best with the smallest windows. This may seem strange since

Table 3
Summary of the input parameters and the results of the variable selection made by uninformative variable selection by PLS

Model	Input parameter	Responses			
	Window size	Number of wavelengths	Components	RMSEP	RMSECV
UVE1	10	200	3	1.113	0.847
UVE2	25	300	4	1.117	0.855
UVE3	50	250	4	1.151	0.942
Original	–	–	3	1.194	0.865

Table 4
Results from the variable selection with IVS-PLS

Model	PLS 800–1350	IVS1		IVS2		IVS3		IVS4		IVS5	
	RMSEP	RMSEP	Selection	RMSEP	Selection	RMSEP	Selection	RMSEP	Selection	RMSEP	Selection
PLSC1	3.447	3.4599	200; 50; 02; up	3.4869	All	3.4869	All	3.4698	200; 10; 39; down	3.4698	200; 10; 39; down
PLSC2	1.605	3.1795	250; 50; 02; up	1.0423	100; 10; 36; up	2.748	100; 10; 45; down	2.8427	200; 10; 44; down	1.5213	150; 10; 23; up
PLSC3	1.194	0.9576	All	1.1714	All	1.754	100; 10; 46; down	1.3282	50; 10; 36; down	1.0999	50; 10; 46; down

The codes under selection stand for: window size; step size; window selected; weighing method. The two weighing methods are weighing up (up) and weighing down (down).

the features in the NIR region are much broader than the smallest window size, but this can be explained by the fact that the windows are selected consecutively, e.g. when a window size of 100 wavelengths is used, the windows starts with wavelength 1, 101, 201, etc. The outcome of the different window sizes may have been different if the windows could be placed anywhere in the wavelength range.

All variable selection methods selected wavelengths above the region selected manually (800–1350 nm), despite the high absorbance values seen in Fig. 2. The fact that some information is still found is

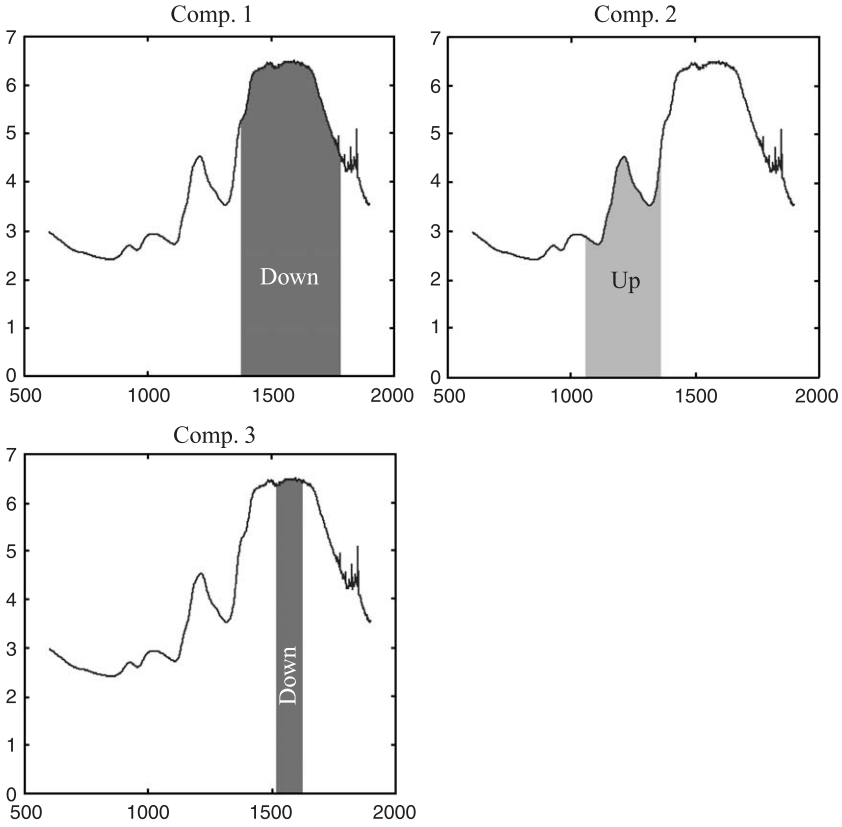


Fig. 7. Overview of one of the best IVS-PLS models, IVS5.

thought to be due to stray light leaking to the detector between the tablet and the wall of the tablet holder in the instrument. Although the low signals in this region are outside the detector's linear regime, the response in this region may contain quantitative information that improves the model.

When comparing the different algorithms, it is important to keep in mind that IVS-PLS is not a variable selection algorithm in the classical sense since it selects different variables in different PLSC.

5. Conclusions

When different variable selection techniques were conducted on NIR transmission data, small improvements of the predictive ability, compared to a model where the wavelength range was manually selected, were seen in all models made after variable selection, no matter what method used. A comparison shows that the largest reductions of RMSEP values were found when using the genetic algorithm or interactive variable selection for PLS. However, it is important to notice that these results are valid only for this data set and that further measurements and investigations have to be made before any general conclusion can be drawn.

The genetic algorithm improved the RMSEP value by 2–15%. This test was though far from complete and it is probably possible to find even better configurations of the algorithm. Although only small changes were made in the input factors, all runs gave different results, which proves that there is an enormous configuration challenge to overcome when using genetic algorithms.

Iterative PLS improved the RMSEP in all runs, but in all cases, it only decreased with less than 6%. Unlike the genetic algorithm, all runs selected similar wavelengths and several of them gave exactly the same results.

Uninformative variable elimination by PLS led to reductions of the RMSEP value by 4–7%.

The method giving the best results was interactive variable selection for PLS. The calculated reduction of RMSEP was almost 20% in the best run. The searches for the optimal selections are however very slow when using IVS-PLS because the algorithm needs to be run a lot of times before finding a good solution.

References

[1] J. Gottfries, H. Depui, M. Fransson, M. Jongeneelen, M. Josefson, F.W. Langkilde, D.T. Witte, J. Pharm. Biomed. Anal. 14 (1996) 1495–1503.
[2] P. Corti, G. Ceramelli, E. Dreassi, S. Mattii, Analyst 124 (1999) 755–758.
[3] A. Sparén, M. Malm, M. Josefson, S. Folestad, J. Johansson, Appl. Spectrosc. 56 (2002) 586–592.
[4] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, C. de la Pezuela, Analyst 123 (1998) 135R–150R.
[5] W.F. McClure, Anal. Chem. 66 (1994) 43A–53A.
[6] T. Fearn, NIR News 9 (1996) 3.
[7] T. Fearn, NIR News 10 (1999) 13–14.
[8] S. Wold, H. Antii, F. Lindgren, J. Ohman, Chemom. Intell. Lab. Syst. 44 (1998) 175–185.
[9] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Appl. Spectrosc. 43 (1989) 772–777.
[10] P. Geladi, D. MacDougall, H. Martens, Appl. Spectrosc. 39 (1985) 491–500.
[11] A.S. Bangalore, R.E. Shaffer, G.W. Small, M.A. Arnold, Anal. Chem. 68 (1996) 4200–4212.
[12] R. Leardi, R. Boggia, M. Terrile, J. Chemom. 6 (1992) 267–281.
[13] R. Leardi, A.L. Gonzalez, Chemom. Intell. Lab. Syst. 41 (1998) 195–207.
[14] S.D. Osborne, R.B. Jordan, R. Kunnemeyer, Analyst 122 (1997) 1531–1537.
[15] V. Centner, D.L. Massart, O.E. deNoord, S. deJong, B.M. Vandeginste, C. Sterna, Anal. Chem. 68 (1996) 3851–3858.
[16] F. Lindgren, P. Geladi, S. Rännar, S. Wold, J. Chemom. 8 (1994) 349–363.

Appl. Phys. B 79, 525–530 (2004)
DOI: 10.1007/s00340-004-1565-8

Applied Physics B
Lasers and Optics

P. WEIBRING
C. ABRAHAMSSON✉
M. SJÖHOLM
J.N. SMITH*
H. EDNER
S. SVANBERG

Multi-component chemical analysis of gas mixtures using a continuously tuneable lidar system

Department of Physics, Lund Institute of Technology, 22100 Lund, Sweden

Received: 26 September 2003/Revised version: 14 May 2004
Published online: 7 July 2004 •

ABSTRACT Differential absorption lidar (DIAL) measurements are usually made on single compounds by alternately switching the wavelength between on and off a resonance line. The selection of more than two wavelengths is a mathematical necessity for simultaneous measurement of multiple species or for resolving interference effects between a compound of interest and a background gas such as water vapour or carbon dioxide. This is especially true in the mid-IR region, where many hydrocarbon compounds have important spectral features. We present a method for remote measurement of gas mixtures in the mid-IR region based on a newly developed fast-switching, frequency-agile optical parametric oscillator lidar transmitter. A multivariate statistical procedure has also been applied for this system, which combines a genetic algorithm for wavelength selection with a partial least squares method for identifying individual compounds from their combined absorption spectrum. A calibration transfer is performed for compounds of interest using reference spectra from an absorption spectra database. Both indoor absorption cell measurements and outdoor remote range resolved measurements of hydrocarbon mixtures were performed to explore the performance of the method.

PACS 42.62 Fi; 42.79 Qx; 02.50 Sk

1 Introduction

The increased awareness of the environmental impact of certain industrial activities, and more stringent regulations on emissions, call for more powerful measurement techniques for air pollution monitoring. Especially interesting are the hydrocarbons due to their absorption of terrestrial radiation in the infrared wavelength region, which contributes to the greenhouse effect [1]. The amount and composition of the anthropogenic hydrocarbon emissions from sources such as petrochemical industries, pipelines, road transportation, use of solvents, rice paddies, and biomass burning are also crucial for understanding their impact on the chemistry of the atmosphere on regional and global scales.

In this context, optical remote sensing techniques [2] are particularly advantageous, allowing large-area monitoring and avoiding sample extraction and preparation difficulties. The mid-IR wavelength region is an important spectral range where many hydrocarbon compounds have fundamental rotational–vibrational transitions while major constituents such as N_2 and O_2 do not. In Fig. 1 the absorption spectra of some simple n-alkanes and water vapour are presented. It can be concluded that there are significant overlaps in the absorption spectra of the hydrocarbons, as well as interference effects from water vapour. Spectrometers have frequently been used for multiple-species detection over long atmospheric paths, utilising different multivariate techniques to overcome these interference effects. In these cases, the whole spectral signature, containing hundreds of measurement wavelengths, is utilised for the detection. In the case of lidar systems, it would be possible to scan the entire wavelength range using a tuneable, pulsed transmitter, but this is not feasible due to the limited repetition rate of most lasers. Still, a substantial number of different wavelengths need to be employed in the multi-species lidar measurement, rather than just a pair of on- and off- resonance wavelengths, as is customary in differential absorption lidar (DIAL). The multi-wavelength lidar concept raises new considerations, such as how many wavelengths and which wavelengths should be used during a particular measurement. In addition, the signal obtained from a single laser pulse in a lidar measurement is not sufficient for data extraction. On the contrary, a large number of shots have to be averaged for each measurement wavelength in order to get an acceptable signal-to-noise ratio. Thus, for a concentration measurement the total number of transmission wavelengths must be weighed by the time required to obtain a sufficient signal for each wavelength.

A few laser-based techniques have been proposed for remote measurements of spectrally overlapping species. Kasparian et al. applied a femtosecond white-light source and a time resolved spectrometer to record whole spectral signatures, ranging over hundreds of nanometers [3]. Presently, available intensity in the mid-IR wavelength region makes hydrocarbon monitoring impractical using this approach. Robinson et al. used a hybrid approach, consisting of DIAL measurements and gas sampling, to determine range resolved

✉ Fax: +46-46/222-4250,
E-mail: Christoffer.Abrahamsson@fysik.lth.se
*_On leave from_: Department of Environmental Engineering Science, California Institute of Technology, Pasadena, CA 91125, USA;
Present address: Atmospheric Chemistry Division, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA

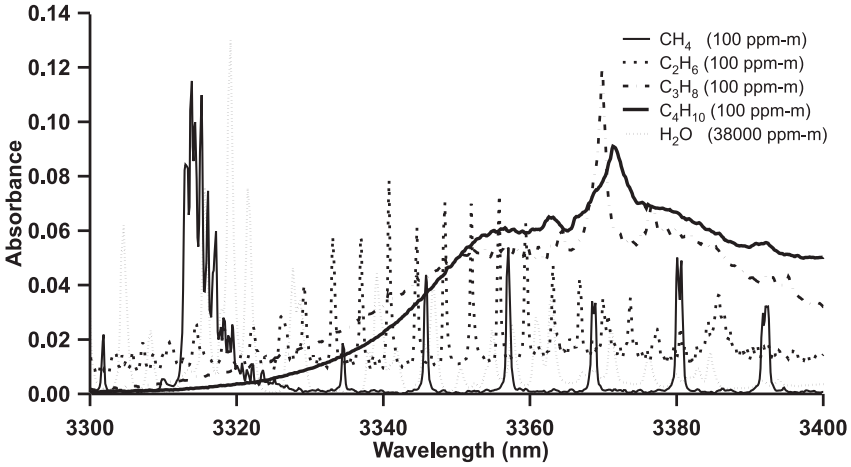


FIGURE 1 Example of overlapping hydrocarbon and water absorption spectra in the middle-IR region

concentrations of hydrocarbon mixtures [4]. Quagliano et al. employed a CO_2 laser in the wavelength range from 9 to 11 μm to make long-path absorption measurements [5]. In that case the discrete nature of the tuning of this laser necessitates the limitation of a set of compounds that have interesting spectral features coinciding with the emitted laser wavelengths. The same group chose to use partial least squares techniques, which we also have explored, for analysing the composition of hydrocarbon mixtures from their combined absorption spectra. The most important distinction between previous papers on the subject and the present one is that our system can be tuned to any wavelength, in any order, in a range of 160 cm^{-1}, and on a shot-to-shot basis providing on-line range resolved measurements of compound mixtures [6]. The systems overall tuning range, 220–4300 nm, enables measurements of a major part of the important trace gases in the atmosphere.

The quantitative identification of multiple, coexisting compounds from their combined spectrum requires not only a measurement system that has the ability to transmit and detect light at multiple wavelengths, but also the implementation of the appropriate evaluation tools and measurement procedures. To manage these requirements, we will present a method, based on a frequency agile transmitter, genetic variable selection and multivariate statistical techniques. The method is tested by absorption measurements in the laboratory and range resolved outdoor measurements on a remote artificial smokestack.

2 Measurement procedure

The task of the multivariate regression technique is to create a mathematical model that relates the absorption of the measurement wavelengths to the concentration of individual components in a gas mixture. This model can then be used to predict the composition of new gas samples.

The multivariate analysis starts with the acquisition of spectra from gas samples with known compositions. A regression model is then built, with the important constraint that the model should be built up using the individual and combined spectra of all compounds and concentrations that one would likely encounter in future measurements. Various methods exist for building a regression model, including the well known multiple linear regression technique that is based on ordinary least squares regression. The problem with applying this technique to spectroscopic measurements is the requirement that the variables (spectra) must be linearly independent. In addition, ordinary least squares regression tends to over-fit noisy data. The partial least squares (PLS) technique [7] overcomes most of these problems by the virtue of being a projection technique, i.e., the PLS technique projects the information onto a lower dimensional subspace. The model is built in such way that the first component conveys as much of the variations in the data as possible, at the same time as it finds the relation between the two data blocks (spectra and gas concentrations). If a one-component model is not good enough, more components are added until the variations modelled by the last component is regarded as noise. The quality of the model is evaluated by predicting the gas concentrations of the samples in a test set and thereby calculate the mean prediction error (RMSEP) value:

$$\mathrm{RMSEP} = \sqrt{\sum_{i=1}^{n} \frac{(y_i - c_i)^2}{n}} \tag{1}$$

where y_i is the predicted, c_i is the real concentration value of the sample and n is the total number of samples. In most studies that have been performed up to date, the multivariate regression technique of PLS has been shown to provide the most accurate predictive models [8] that are the least sensitive to background noise, and therefore this technique will be used for analysing the data in this work.

A recording of the whole spectrum would be desirable, but due to the limited repetition rate of the laser, such measurements could not be carried out in a reasonable time scale, during which the atmospheric circumstances and the compounds present could be considered as constant in the absorption path. Therefore there is a need to reduce the number of measured wavelengths, which is easily done when measuring mixtures containing only a few compounds. When the number of compounds increases, the selection of suitable measurement wavelengths gets more and more difficult, and in the case of severely overlapping spectral structures, it is not possible by the eye to determine which wavelengths to choose. Therefore a genetic algorithm (GA) can advantageously be used as

guidance when selecting which wavelengths to measure. The GA has proven to be a valuable tool for automated variable selection [9–11]. The drawback with GA is normally the substantial set-up preparations that have to be considered. For the present application only some basic set-up properties of the GA were explored, since the goal is not to optimise the GA but just to get some guidance when selecting wavelengths. The wavelengths selected by the GA could then be used as a base when measuring with the system.

For a fixed number of wavelengths the GA consists of four basic steps, where steps 2–4 are performed until a stop criterion is fulfilled:

1. A number of chromosomes (different wavelength sets) are constructed randomly. This is called the initial population. The number of wavelengths in each chromosome is locked to a certain number during the entire wavelengths selection procedure.
2. For each chromosome the response associated with the corresponding experimental conditions is evaluated. This is done by making a PLS model for each chromosome. The model is then evaluated by means of predicting a test set and thereby calculating the RMSEP-value. This value is used as the fitness value and is the criterion for guiding the GA to the global optimum.
3. The reproduction step creates a new population that can be considered as the next generation. The new generation of chromosomes is made up by recombination of the original chromosomes. The chromosomes with a good fitness value have a higher probability to reproduce than a chromosome with a bad fitness value, all in order to improve the overall fitness of the population. The chromosome with the best fitness value is always transferred unchanged to the next generation.
4. Mutations are necessary to overcome some problems that may occur. The most essential problem to be solved is that if a wavelength should not be selected in any of the original chromosomes, it would never be selected in the coming generation if mutations were not present. Another important aspect is that the mutations help the GA not to find local minima, but to find the global minimum. A mutation is simply an exchange of one of the wavelengths in the chromosome to a randomly chosen one.

Although it is optimal to experimentally record these absorption spectra using gases in a test cell, this is not practical for such a large set of compounds that normally can be found in a polluted atmosphere. A better alternative is to use spectroscopic databases, which include most species in the atmosphere. The database approach also has the advantage that it will allow the multivariate model to be modified in the field if a previously unidentified compound would be found to exist in the gas mixture. It is though important to adjust the database data so it also takes the instrumental parameters into account. For example, since field experiments are associated with noisy signals, white noise has to be added to the spectra included in the model. In addition, differences in laser linewidth, wavelength stability, wavelength offset, and wavelength slope lead to further differences in the recorded spectra. All above stated parameters can be taken into account through a calibration transfer procedure, which establishes a relationship between the instrument and database responses.

Due to the above-mentioned considerations, a measurement must sometimes be performed in an iterative way, especially if there is no detailed knowledge about the composition of the pollution.

A systematic way of dealing with this is proposed in the following procedure.

- Building a model based on experience and basic knowledge of the source's emission inventory and atmospheric conditions. For instance, a petrochemical industry emits a wide range of hydrocarbons. Based on this knowledge the model is built on the expected existing compounds in the emission.
- A GA is used to select the appropriate wavelengths for later PLS determination of the concentrations of the different pollutants. The wavelength selection is made based on database spectra, with added noise etc.
- A PLS model is built from database spectra, with the GA selected wavelengths.
- The lidar measurement is performed with the required number of shots to reach the desired signal-to-noise ratio.
- Calibration transfer is performed between the instrument recording at the selected wavelengths and the database spectra.
- Prediction of the individual concentrations of the compounds in measured mixtures by importing the calibration transferred data to the model.
- Checking the model prediction error statistics. If the measurement gives large residuals, probably the compounds included in the model do not match the composition of the emission. The model is then rebuilt, based on experience, by either increasing or decreasing the number of included compounds until the residual information not described by the model is minimised.

3 Experimental set-up

Experiments for this investigation were performed with the Lund University mobile lidar system [12], which is shown schematically in Fig. 2. The key element of the system is a fast tuning lidar transmitter, using all-solid-state technology, providing range- and temporally resolved atmospheric measurements of gas concentrations. The instrument is based on a commercial OPO laser system (Spectra Physics MOPO-730) which has been redesigned introducing piezoelectric transducers mounted on the wavelength tuning mirror and on the crystal angle tuning element in the OPO. A piezoelectric transducer similarly controls the frequency-mixing and doubling stages, which have been implemented to extend system capabilities to the mid-IR and UV regions. The system is able to produce radiation, with an average linewidth better than 0.2 cm^{-1} and a shot-to-shot tunability of 160 cm^{-1}, with an accuracy better than the linewidth. For instance, this means that the fast tuning ability is within a range of 170 nm in the mid-IR wavelength region. The output power in the wavelength regions 220–1800 nm and 2600–4300 nm reaches 100 mJ and 20 mJ, respectively [12]. The system performance in terms of wavelength, linewidth and power stability is monitored on a shot-to-shot basis in real time by

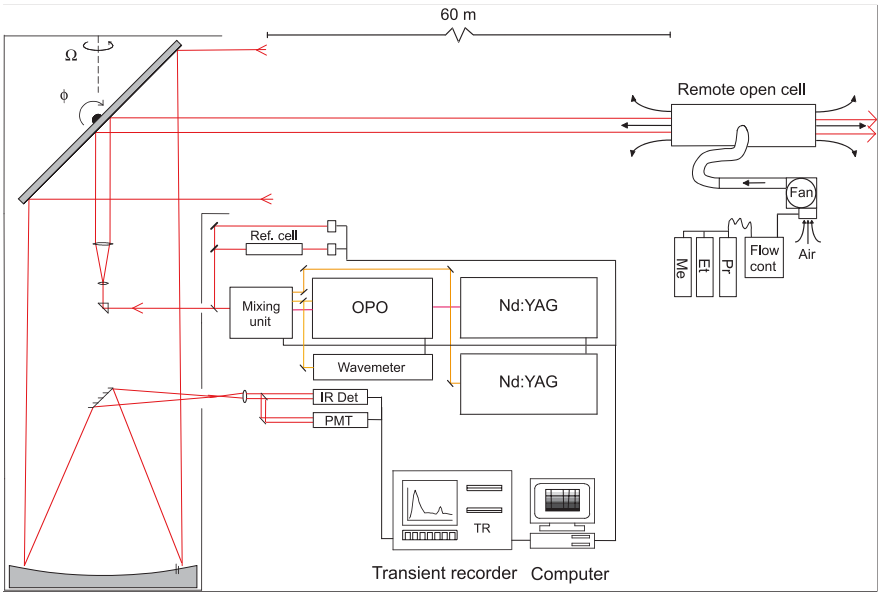


FIGURE 2 Optical schematics of the Lund University lidar system and outdoor open-ended flow cell

a surveillance system based on a Fabry–Pérot interferometer. If the desired requirements are not met for a specific light pulse the system discards the corresponding collected data and repeats the measurement until the requirements are fulfilled. For concentration and wavelength reference purposes, an absorption cell system is incorporated in the lidar system, enabling different reference mixtures to be measured at the same time as the lidar measurement is performed. The outgoing beam from the lidar system is expanded by a variable beam expander and transmitted out into the atmosphere by a computer controlled transmitting/receiving unit; see Fig. 2. A 40 cm diameter Newtonian telescope is used as a receiver. Two parallel detection channels enable full spectral coverage in the UV, VIS and mid-IR regions by utilizing both a PMT and a liquid-nitrogen-cooled InSb detector. A filter system in front of the detectors enables high signal-to-background ratios for compounds in the whole wavelength region. For an optimally adjusted IR lidar system, a measurement range of about 400 m can be estimated.

To test the potential of the measurement procedure in determining the concentration of different compounds in a mixture, two test stations were built. One station consisted of an ordinary absorption cell arrangement simulating an ideal measurement situation. The other test station consisted of an open-ended remote gas cell simulating a remote plume.

The absorption cell arrangement consisted of a 2 cm long cell. All optics including the cell windows were made of CaF_2. To eliminate interference fringes and achieve a similar power at both detectors, the beams for the cell and reference paths were split off by reflection from the beam line with separate wedged beam splitters. The detectors (PbSe) were tested to assure that the experiment was conducted in the linear regime of their response. To create a gas mixture the following steps were carried out: The absorption cell was flushed with nitrogen and then evacuated, after which it was filled with different partial pressures of methane, ethane, propane and butane ($>$ 99% purity) using a pressure gauge. Following this, the cell was filled with nitrogen to atmospheric pressure resulting in concentrations of 0–140 ppm m of the above stated hydrocarbons. For validation purposes a gas mixture was measured using a FTIR instrument (Mattsson Instruments), which showed a good agreement with the lidar system measurement.

In order to simulate a real-life measurement, an outdoor open-ended flow cell was built for testing the lidar system under conditions that closely resemble emission sources such as factory smokestacks. The facility was located 60 m from the lidar system, and consisted of a 3 m long, 40 cm diameter pipe, through which ambient air was forced from a side pipe using a fan. At the inlet of the fan, controlled flows of methane, ethane and propane were added to a controlled flow of carrier gas consisting of ambient air. This mixed gas was then introduced into the center of the main flow cell and allowed to pass out each open end of the pipe, as shown in Fig. 2. The total flow of the gas mixture was measured by a pressure gauge inserted in the side pipe. The concentration of the different gases in the cell was calculated by dividing the flow of the individual gases by the total flow. All flows could be set and monitored from a computer at the lidar system. The average flow data of the gases and air were stored on the computer disc every 10 s. The mass flow controllers were tested to have a maximum error of 15%, and the variation of the air flow was constantly measured and found to be within 2% which results in a total concentration error of approximately 15%. By firing the lidar beam through the open-ended flow cell, range resolved absorption measurements were performed. The maximal range-resolution allowed by the system is 7.5 m. The integrated gas contents (ppm m) due to the gas released was evaluated from the size of the resulting steps in the lidar curves, evaluated with some smoothing of the raw data, using an effective range interval of 15 m.

4 Results

The measuring and analysis procedure for both the absorption cell and the open ended flow cell was performed according to the multivariate concept described above. To make a PLS model, first database spectra of methane, ethane, propane, butane and water were utilised to make 150 resulting

spectra, spanning the expected concentrations of the compounds (0–200 ppm m). Also 40% of white noise and a wavelength jitter of 0.2 nm were added to the spectra to enable a modelling resistant to noisy data and laser wavelength jitter. A GA was utilised to select the most significant wavelengths to use in the PLS model according to the procedure described above. The GA input properties were kept constant at 100 separate chromosomes (wavelength sets), 100 generations and a 5% mutation rate, with the exception of the number of selected wavelengths in order to examine the effect of that factor. Four different sets of wavelengths were compared: 10 and 20 wavelengths selected by the GA, 20 wavelengths selected manually based upon experience of a spectroscopist, and 20 wavelengths selected by a combination of the GA and a spectroscopist. These four sets of wavelengths and a fifth case, utilising all scanned wavelengths, were then used to build separate PLS models.

To validate the wavelength selection procedure and the potential of the different PLS models in determining the concentration of different compounds, an experiment with the indoor absorption cell was performed. The gas cell was filled with 0–140 ppm m of methane, ethane, propane and butane spanning all encountered concentration situations. After the cell was filled, a complete wavelength scan from 3305 to 3385 nm, with a resolution of 0.1 nm, was performed. A simple calibration transfer was then performed, consisting of re-scaling and offset adjustment in order to match the measured spectra to the resolution and spectral response of the database spectra. These data were then used to validate the above described PLS models created from the database. The result from this experiment is summarised in Table 1. Also the models based on wavelength selections for 20 wavelengths selected by a spectroscopist (manual) and 20 wavelengths selected by a combination of the GA and a spectroscopist (GA & manual) are shown in Fig. 3. It can be concluded that a PLS model using 10 wavelengths based on a GA selection gives quite large residuals, but a PLS model based on a 20 wavelength GA selection improves the predictive abilities. Further increase of the number of wavelengths, to e.g. 30–40, does not significantly increase the prediction capabilities. The automated GA selection of 20 wavelengths performs actually better for ethane than a manual selection of 20 wavelengths based on the selection of a spectroscopist. The GA & manual model with 10 wavelengths selected by a GA and 10 wavelengths added afterwards by a spectroscopist gives the best results, thus achieving a robust model with predictive powers only succeeded by the model based upon the complete spectrum. The prediction error in the combined model is about 7 ppm m for methane, and between 20 and 25 ppm m for the other species. The main improvement in accuracy of the combination model compared to the manual model is achieved for ethane and propane. For ethane, the selection of measurement wavelengths is completely different among the two models while for propane the number of measurement wavelengths around the main peak is increased in the GA & manual model. Here the aid from GA proves to be superior due to the fact that it is difficult to determine by eye only the wavelengths that enable a good de-convolution of the interference among the different species.

The sharp-featured substances can be modelled in a more accurate and robust way, as in the case of ethane, using a few wavelengths rather than using complete spectra, due to the fact that the errors introduced by varying interfering substances could be minimised. It can also be seen that substances with sharp features, such as methane and ethane, are easily modelled using just a few wavelengths whereas heavier compounds with broader spectral features such as propane and butane need a larger set of wavelengths to be modelled in a proper way.

The cell experiment showed that the combination of GA and manual selection gave the best wavelength selection; therefore, these 20 wavelengths were used when measuring on the outdoor cell. The outdoor cell system was set to deliver 0–100 ppm m of methane, ethane and propane for the differ-

	RMSEP methane (ppm m)	RMSEP ethane (ppm m)	RMSEP propane (ppm m)	RMSEP butane (ppm m)
10 wavelengths GA	15	37	43	32
20 wavelengths GA	9	21	37	25
20 wavelengths manual	9	35	33	25
20 wavelengths GA & manual	7	23	19	26
Complete spectra	6	27	14	15

TABLE 1 Average errors (RMSEP) from the predictions of the samples measured in the indoor cell with variations of the different compounds 0–140 ppm m, showing the differences between the various wavelength selection cases

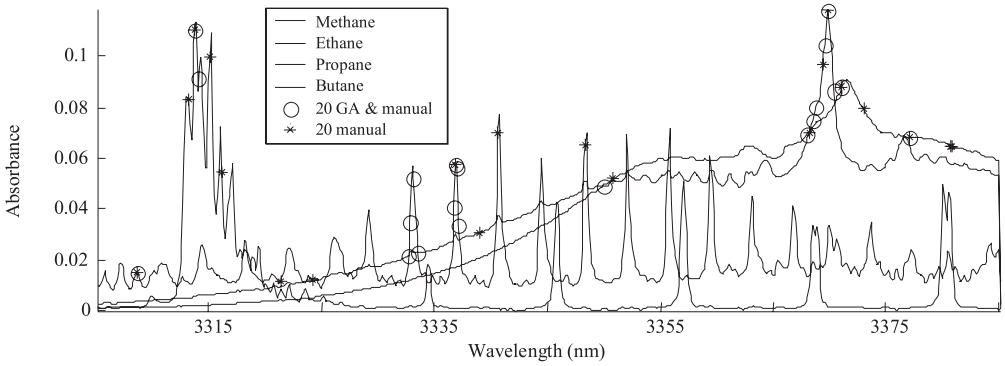


FIGURE 3 Spectra of the four measured substances, with the selected wavelengths for the different wavelength selection cases marked

TABLE 2 Set values for the mass flow controllers and lidar system response for methane, ethane and propane

Mixture	Set methane concentration (ppm m)	Predicted methane concentration (ppm m)	Set ethane concentration (ppm m)	Predicted ethane concentration (ppm m)	Set propane concentration (ppm m)	Predicted propane concentration (ppm m)
1	0	18	50	46	50	64
2	50	53	100	104	0	0
3	0	10	100	118	0	−2
4	0	5	100	127	100	120
5	100	99	100	99	0	−7
6	100	120	100	120	0	14
7	100	101	100	107	100	109
8	50	56	50	63	50	53
9	0	24	0	−7	50	38
10	100	145	100	105	100	129
Average error		13		10		11

ent measurements. On this occasion the typical pulse energy of the emitted IR radiation was 8–10 mJ. The range resolved data were then used to calculate the absorption at the different wavelengths through the cell. The natural background concentration of methane was measured and taken into account during the calculations. The results from the outdoor cell measurements are summarised in Table 2. The prediction errors are comparable between the lidar and the indoor cell measurements, with some exceptions. The major factor for the disagreement is probably the wind conditions. The fact that the range resolution of the lidar system is only 7.5 m and that the wind speed and direction affects the effective length of the absorption path introduces large errors. The length of the absorption path could typically vary from 3 to 6 m between the extremes of the wind conditions. This was confirmed by introducing smoke to the inlet of the cell system and studying the resulting plumes out of the cell openings for different wind conditions.

5 Discussion

The concept of a multi-wavelength lidar system with multivariate analysing techniques for measuring hydrocarbon gas mixtures has been successfully implemented and tested. The tests were carried out with the light hydrocarbons methane, ethane, propane and butane. When larger hydrocarbons such as heptane, hexane, and octane are measured, the spectral overlap will be more severe and it can be difficult to determine the individual concentrations of the different hydrocarbons without increasing the number of wavelengths substantially. Recording all spectral structures over a range of hundreds of nanometers is not possible due to prolonged measurement time compared to the time constants of source emission and atmospheric variability. Thus establishing the concentration of different groups of hydrocarbons, using fewer wavelengths, could be the target for such a measurement. The extension of the multivariate techniques employed in the current study to other classes of compounds is thus one area for future development.

Multi-wavelength measurements can be advantageously applied in other spectral regions as well. In the ultraviolet wavelength region, overlapping aromatic hydrocarbons are more easily measured due to less interfering species. High accuracy measurements of sulphur dioxide with interference of ozone in the ultraviolet wavelength region have been performed with specially designed three wavelength systems [13]. By utilising even more wavelengths, an increase in the accuracy of both sulphur dioxide and ozone could be expected. In the same way one could achieve higher measurement accuracy for isolated species that have been traditionally measured by standard two wavelength measurements (DIAL).

ACKNOWLEDGEMENTS This work was supported by the Swedish Space Board, the Knut and Alice Wallenberg Foundation and the Lund/Caltech Academic Exchange Programme. A fruitful collaboration with Spectra Physics, Inc., is gratefully acknowledged.

REFERENCES

1 J.H. Seinfeld, S.N. Pandi: *Atmospheric Chemistry and Physics* (John Wiley & Sons, New York 1988)
2 M.W. Sigrist: Air monitoring by spectroscopic techniques, In: *Chemical Analysis*, vol. 127 (John Wiley & Sons, New York 1994)
3 J. Kasparian, M. Rodriguez, G. Méjean, J. Yu, E. Salmon, H. Wille, R. Bourayou, S. Frey, Y.-B. André, A. Mysyrowicz, R. Sauerbrey, J.-P. Wolf, L. Wöste: Science **301**, 61 (2003)
4 R.A. Robinson, P.T. Woods, M.J.T. Milton: SPIE **2506**, 140 (1995)
5 J.R. Quagliano, P.O. Stoutland, R.R. Petrin, R.K. Sander, R.J. Romero, M.C. Whitehead, C.R. Quick, J.J. Tiee, L.J. Jolin: SPIE **2702**, 16 (1996)
6 P. Weibring, J. Smith, H. Edner, S. Svanberg: Rev. Sci. Instrum. **74**, 4478 (2003)
7 P. Geladi, B.R. Kowalski: Anal. Chim. Acta **185**, 1 (1986)
8 S.J. Haswell, A.D. Walmsley: Anal. Chim. Acta **400**, 399 (1999)
9 R. Leardi, R. Boggia, M. Terrile: J. Chemom. **6**, 267 (1992)
10 A.S. Bangalore, R.E. Shaffer, G.W. Small, M.A. Arnold: Anal. Chem. **68**, 4200 (1996)
11 L. Davies: *Handbook of Genetic Algorithms* (Van Nostrand Reinhold, New York 1991)
12 P. Weibring, H. Edner, S. Svanberg: Appl. Opt. **42**, 1 (2003)
13 T. Fujii, T. Fukuchi, N. Goto, K. Nemoto, N. Takeuchi: App. Opt. **40**, 949 (2001)

MADSTRESS: A Linear Approach for Evaluating Scattering and Absorption Coefficients of Samples Measured Using Time-Resolved Spectroscopy in Reflection

F. CHAUCHARD,* J. M. ROGER, V. BELLON-MAUREL, C. ABRAHAMSSON, S. ANDERSSON-ENGELS, and S. SVANBERG

Information and Technologies for Agro-processes, Cemagref BP 5095, 34033 Montpellier Cedex 1, France (F.C., J.M.R., V.B.-M.); and Department of Physics, Lund Institute of Technology, P.O. Box 118, SE-221 00, Sweden (C.A., S.A.-E., S.S.)

Time-resolved spectroscopy is a powerful technique permitting the separation of the scattering properties from the chemical absorption properties of a sample. The reduced scattering coefficient and the absorption coefficient are usually obtained by fitting diffusion or Monte Carlo models to the measured data using numerical optimization techniques. However, these methods do not take the spectral dimension of the data into account during the evaluation procedure, but evaluate each wavelength separately. A procedure involving multivariate methods may seem more appealing for people used to handling conventional near-infrared data. In this study we present a new method for processing TRS spectra in order to compute the absorption and reduced scattering coefficients. This approach, MADSTRESS, is based on linear regression and a two-dimensional (2D) interpolation procedure. The method has allowed us to calculate absorption and scattering coefficients of apples and fructose powder. The accuracy of the method was good enough to provide the identification of fructose absorption peaks in apple absorption spectra and the construction of a calibration model predicting the sugar content of apples.

Index Headings: **Time-resolved spectroscopy; Equation of diffusion; Multi-linear regression; Light continuum; Apple absorption coefficient; Fructose absorption coefficient.**

INTRODUCTION

Near infrared (NIR) reflectance spectroscopy has the advantage that it can be used to nondestructively measure chemical compounds residing inside a scattering medium.[1] These media may be agricultural products,[2] pharmaceuticals products,[3] or others. An example of application is the prediction of sugar content in apples.[4] Prominent scattering does, however, present a drawback as it modifies the measured absorption spectra. The measured reflectance spectrum is hence a combination of both absorption and scattering effects. Consequently, calibration models based on NIR reflectance spectra implicitly compensate for scattering effects, resulting in complex and non-robust models.[5] For this reason the calibration procedure is often combined with data preprocessing techniques such as standard normal variate[6] or multiple scatter correction.[7] The benefits of the present preprocessing techniques are, however, limited because they only produce data more correlated to the absorption coefficient, but they are not able to extract the real absorption coefficient. Furthermore, the scattering coefficient cannot be evaluated from NIR reflectance data solely, even though its evaluation would be of deep interest since it carries information about the physical characteristics of the sample. An attractive method for evaluating the scattering properties of highly scattering samples is time-resolved spectroscopy[8] (TRS). TRS was mainly developed for medical applications[9,10] but has found its way into other fields of research, such as pharmaceutical[11,12] and agricultural applications.[13–16] The parameters used to describe the optical properties of a turbid medium are the absorption coefficient $\mu_a(\lambda)$, the scattering coefficient $\mu_s(\lambda)$, and the scattering anisotropy g. Often, $\mu_s(\lambda)$ and g are combined to form the reduced scattering coefficient $\mu_s'(\lambda) = \mu_s(\lambda)(1 - g)$. TRS uses short light pulses to irradiate the sample. The light diffusively re-emitted by the sample at a given distance from the irradiation point is then detected as a function of time.[17] TRS measurements can be conducted in either reflectance or transmission mode. In order to obtain a simultaneous measure of the temporal signal at different wavelengths, light pulses with a broad wavelength profile are used in combination with a streak camera detection system. The broad light pulses can be generated using different techniques; one is using continuum generation by focusing of a high power laser pulse into a photonic crystal fiber.[18]

Once the two-dimensional signal (one temporal and one spectral dimension) is recorded, the reduced scattering coefficient ($\mu_s'(\lambda)$) and the absorption coefficient ($\mu_a(\lambda)$) are obtained by linking the experimental data with theoretical or modeled data. This step is crucial to obtaining correct results, and many evaluation schemes have been proposed. Three approaches are usually used: Monte Carlo simulations,[17,19] numerical optimizations,[20,21] and analytical descriptors of temporal dispersion.[22] Those methods do not, however, take the spectral dimension of the data into account. $\mu_a(\lambda)$ and $\mu_s'(\lambda)$ are calculated at a given wavelength without considering neighboring wavelengths. A method based on a linear solution of the equation of diffusion would allow the implementation of various chemometric tools, such as multi-linear regression (MLR), partial least squares, and N-ways methods.

This study aims at proposing a new method for finding scattering and absorption coefficients using TRS data. The first part is devoted to the diffusion equation and a presentation of the suggested linear approach. In the second part, methods and TRS instrumentation are presented. The last part describes the performance of the linear approach and its results.

Received 21 April 2005; accepted 26 July 2005.

* Author to whom correspondence should be sent. E-mail: fabien.chauchard@montpellier.cemagref.fr.

0003-7028/05/5910-1229$2.00/0

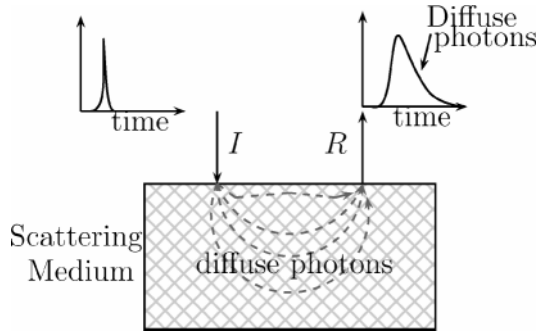


FIG. 1. Propagation of a light pulse in a turbid media.

THEORY

When the photons of a short light pulse penetrate a turbid medium, they scatter around due to the matrix effect (Fig. 1). If the light is detected at a specific distance from the light source, ρ, the detected light pulse will be temporally broader than the pulse sent into the medium. The dispersion of the pulse is governed by the scattering coefficient $\mu_s'(\lambda)$ and the absorption coefficient $\mu_a(\lambda)$ of the medium. The photon transport in turbid media is described by the radiative transport equation,[23] which makes a balance between gained photons and lost photons. In the case of a semi-infinite homogeneous medium measured in reflectance mode, the solution is given by the time-resolved diffusion equation:[17]

$$R(t, \lambda) = [4\pi D(\lambda)v]^{-3/2} z_0(\lambda) t^{-5/2} \exp[-\mu_a(\lambda)vt] \times \exp\left[-\frac{\rho^2 + z_0(\lambda)^2}{4D(\lambda)vt}\right] \quad (1)$$

Here $R(t, \lambda)$ is the signal measured at a given distance ρ at time t, $D(\lambda)$ is the diffusion coefficient with $D(\lambda) = [3(\mu_a(\lambda) + \mu_s'(\lambda))]^{-1}$, $z_0(\lambda) = (\mu_s'(\lambda))^{-1}$ is the mean free path of the photons, and v is the speed of light in the medium, assumed to be constant within the measured wavelength range.

Solving the Equation. Let $t_0(\lambda)$ be the time where $R(t, \lambda)$ has its maximum value. By dividing the signal $R(t, \lambda)$ by its maximum value at $t_0(\lambda)$, $R(t_0(\lambda), \lambda)$, the following equation is obtained:

$$\frac{R(t, \lambda)}{R[t_0(\lambda), \lambda]} = \left[\frac{t}{t_0(\lambda)}\right]^{-5/2} \exp\{\mu_a(\lambda)v[t_0(\lambda) - t]\} \times \exp\left[-\frac{\rho^2 + z_0^2(\lambda)}{4D(\lambda)vt} + \frac{\rho^2 + z_0^2(\lambda)}{4D(\lambda)vt_0(\lambda)}\right]$$

This can also be written as:

$$\frac{R(t, \lambda)}{R[t_0(\lambda), \lambda]}\left[\frac{t}{t_0(\lambda)}\right]^{5/2} = \exp\{\mu_a(\lambda)v[t_0(\lambda) - t]\} \times \exp\left\{\left[\frac{\rho^2 + z_0^2(\lambda)}{4D(\lambda)v}\right]\left[\frac{1}{t_0(\lambda)} - \frac{1}{t}\right]\right\}$$

Taking the logarithm of this equation:

$$\log\left\{\frac{R(t, \lambda)}{R[t_0(\lambda), \lambda]}\right\} + \frac{5}{2}\log\left[\frac{t}{t_0(\lambda)}\right] = \mu_a(\lambda)v[t_0(\lambda) - t] + \left[\frac{\rho^2 + z_0^2(\lambda)}{4D(\lambda)v}\right]\left[\frac{1}{t_0(\lambda)} - \frac{1}{t}\right] \quad (2)$$

When $R(t, \lambda)$ has its maximum, $(\partial R/\partial t)(t_0(\lambda), \lambda) = 0$. $\partial R(t, \lambda)/\partial t$ is given by:

$$\frac{\partial R(t, \lambda)}{\partial t} = [4\pi D(\lambda)v]^{-3/2} z_0(\lambda) \exp[-\mu_a(\lambda)vt] \times \exp\left[-\frac{\rho^2 + z_0(\lambda)^2}{4D(\lambda)vt}\right]\frac{5}{2}t^{-7/2} + [4\pi D(\lambda)v]^{-3/2} z_0(\lambda) \exp[-\mu_a(\lambda)vt] \times \exp\left[-\frac{\rho^2 + z_0(\lambda)^2}{4D(\lambda)vt}\right] t^{-5/2} \times \left[-\mu_a(\lambda)v + \frac{\rho^2 + z_0(\lambda)^2}{4D(\lambda)vt^2}\right]$$

Factoring using $R(t, \lambda)$, the expression at $t = t_0(\lambda)$ is:

$$\frac{\partial R}{\partial t}[t_0(\lambda), \lambda] = R[t_0(\lambda), \lambda]\left[-\frac{5}{2}t_0(\lambda)^{-1}\right] + R[t_0(\lambda), \lambda]\left[-\mu_a(\lambda)v + \frac{\rho^2 + z_0(\lambda)^2}{4D(\lambda)vt_0(\lambda)^2}\right] = 0$$

Taking into account that $R(t_0(\lambda), \lambda)$ is a maximum and thus different from zero, it follows that:

$$-\frac{5}{2}t_0(\lambda)^{-1} - \mu_a(\lambda)v + \frac{\rho^2 + z_0(\lambda)^2}{4D(\lambda)vt_0(\lambda)^2} = 0$$

which leads to:

$$\frac{\rho^2 + z_0(\lambda)^2}{4D(\lambda)v} = \frac{5}{2}t_0(\lambda) + \mu_a(\lambda)vt_0(\lambda)^2 \quad (3)$$

Substituting into Eq. 2 yields:

$$\log\left\{\frac{R(t, \lambda)}{R[t_0(\lambda), \lambda]}\right\} + \frac{5}{2}\log\left[\frac{t}{t_0(\lambda)}\right] = \mu_a(\lambda)v[t_0(\lambda) - t] + \left[\frac{5}{2}t_0(\lambda) + \mu_a vt_0(\lambda)^2\right]\left[\frac{1}{t_0(\lambda)} - \frac{1}{t}\right]$$

This equation can be put under the following form, and $\mu_a(\lambda)$ can be found using experimental data by solving:

$$\mu_a(\lambda)G(t, \lambda) = F(t, \lambda) \quad (4)$$

with:

$$F(t, \lambda) = -\log[R(t, \lambda) + \log[R(t_0, \lambda)] - \frac{5}{2}\log\left(\frac{t}{t_0}\right) + \frac{5}{2}\left(1 - \frac{t_0}{t}\right) \quad \text{and:}$$

$$G(t, \lambda) = v\frac{1}{t}[t - t_0(\lambda)]^2$$

$\mu_a(\lambda)$ can be found by applying a multivariate method to Eq. 4 where $F(t, \lambda)$ contains the values to predict and

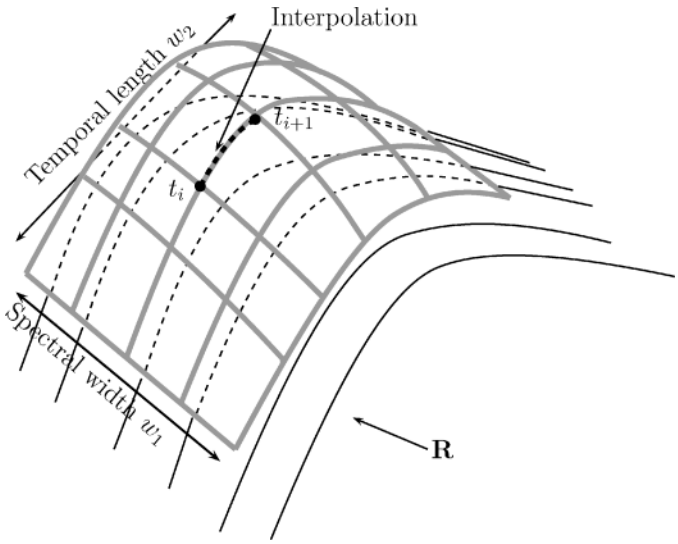


FIG. 2. Adaptive pattern for time interpolation.

$G(t, \lambda)$ contains the variable ($\mu_a(\lambda)$ is then the regression coefficient found).

Once $\mu_a(\lambda)$ has been evaluated, Eq. 3 enables us to obtain $\mu_s'(\lambda)$:

$$\frac{3}{4v}[\rho^2 + z_0(\lambda)^2][\mu_a(\lambda) + \mu_s'(\lambda)] = \frac{5}{2}t_0(\lambda) + \mu_a(\lambda)vt_0(\lambda)^2$$

which gives:

$$\frac{3\rho^2}{4v}[\mu_a(\lambda) + \mu_s'(\lambda)] + \frac{3}{4v}\left[\frac{\mu_a(\lambda)}{\mu_s'(\lambda)^2} + \frac{1}{\mu_s'(\lambda)}\right] = \frac{5}{2}t_0(\lambda) + \mu_a(\lambda)vt_0(\lambda)^2$$

As $\mu_s'(\lambda) \gg \mu_a(\lambda)$, $(\mu_a(\lambda)/\mu_s(\lambda)^2)$ can be neglected compared to $1/\mu_s(\lambda)'$. $\mu_s'(\lambda)$ is then linked to $\mu_a(\lambda)$ by a second-order equation:

$$\frac{3\rho^2}{4v}\mu_s'(\lambda)^2 + \left[-\frac{5}{2}t_0(\lambda) - \mu_a(\lambda)vt_0(\lambda)^2 + \frac{3\rho^2}{4v}\mu_a(\lambda)\right]\mu_s'(\lambda) + \frac{3}{4v} = 0 \quad (5)$$

Implementation on Experimental Data. An accurate determination of $t_0(\lambda)$ may be judged as the keystone for successful results with the suggested approach. Experimental data do, however, contain noise and the time resolution is limited by the measuring apparatus. These constraints prevent the use of the measured signal maximum to find $t_0(\lambda)$, since measurement noise may hide the real maximum value of the time-resolved signal; likewise, the apparatus time resolution restricts $t_0(\lambda)$ prediction accuracy. These problems may be limited by smoothing and artificially increasing the signal time resolution.

Let **R** be a matrix defined by $\{R(t_i, \lambda_j)\}$, where $t_i \in \{1, \ldots, p\}$ and $\lambda_j \in \{1, \ldots, q\}$, which defines a TRS measurement (Fig. 2). A mesh of width w_1 (odd) and length w_2 (even) is wrapped on a given part of **R**. The wrapping is performed by fitting of a two-variable parabolic polynomial function of the form $at^2 + bt + c\lambda^2 + d\lambda + et\lambda + f$, with $R(t, \lambda)$. The segment (t_i, t_{i+1}) situated at the mesh center (bold dotted line in Fig. 2) is then interpolated, where k new points are added between t_i and t_{i+1} using the parabolic equation of the wrapped mesh. By applying this procedure on each part of **R**, a new matrix $\mathbf{R}^+$ of size $(p \times k) \times q$ is created.

For a given wavelength λ_j, each t_i is taken as candidate for $t_0(\lambda_j)$ and the correlation coefficient, $r(t_i(\lambda_j), \lambda_j)$, between $G(t, \lambda_j)$ and $F(t, \lambda_j)$ is calculated. Assuming Eq. 4 must exist at $t_i = t_0(\lambda_j)$, the value for $t_0(\lambda_j)$ is chosen where the highest value of the correlation coefficient is reached, $r(t_0(\lambda_j), \lambda_j)$ (i.e., where Eq. 4 is the most verified and where $\mu_a(\lambda)$ has a positive value). The approach of combining Eq. 4 with the above-described implementation is named MAximum Determination for Solving Time-REsolved Spectroscopy Signal (MADSTRESS).

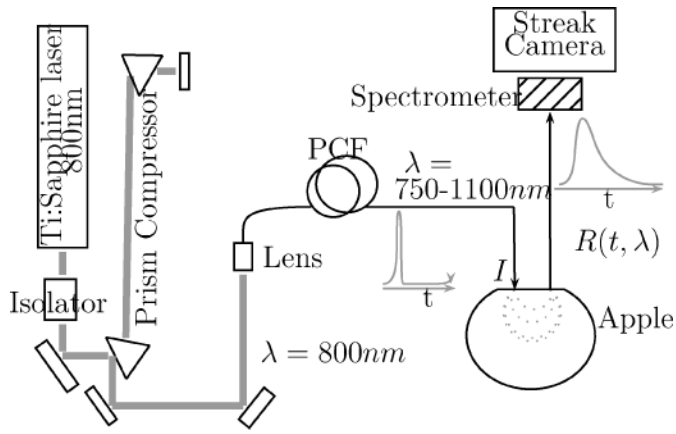


FIG. 3. Setup for TRS measurements.

MATERIALS AND METHODS

Time-Resolved Spectroscopy Instrumentation. Figure 3 depicts the experimental setup. The instrument has been described in detail by Abrahamsson et al.[18] Briefly, a mode-locked Ti:Sapphire laser, pumped by an argon-ion laser, was used to generated 100 fs pulses centered around 800 nm with an 80 MHz repetition rate. The laser pulses were focused into a 100 cm long index guiding crystal fiber (ICF) (Crystal Fiber A/S, Copenhagen, Denmark). The broadband light pulses generated by nonlinear effects in the ICF ranged from 750 nm to 1100 nm. The light was then transferred by a set of lenses into a gradient index fiber guiding the light to the sample. Another gradient fiber, with the distal tip $\rho = 6$ mm from the irradiating one, was used to collect the light re-emitted from the sample. The fibers were put in contact with the sample. A Streak Camera (Hamamatsu, ModelC5680) coupled to an imaging spectrometer (Chromex, Model 250IS) captured the reflected light as a function of time and wavelength $R(t, \lambda)$. The spectral resolution was 0.93 nm distributed over 512 pixels, while the temporal resolution was 2.93 ps in the span from 0 to 1900 ps, spread over 640 pixels. Integration time was 5 min.

Samples. Fifteen Golden Delicious apples were measured using the TRS setup at an ambient temperature of 25 °C. A small part of each apple was carefully removed in order to create a flat surface for applying the fibers. The measurements were performed immediately after the preparation of the apples in order to avoid flesh drying.

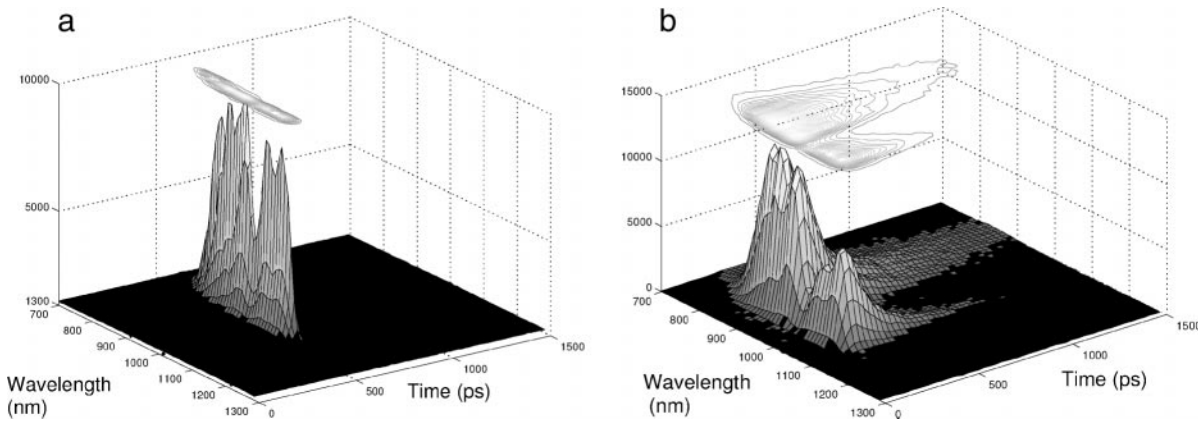


FIG. 4. Recording of a two-dimensional time-resolved measurement on an apple.

After the TRS measurement, the reference sugar content was measured: a drop of apple juice was taken from the scanned surface below the fibers and spread on a Euromex RD645 refractometer with 0.2 °Brix accuracy (i.e., about 0.2 g of sugar per 100 mL).

Two other kinds of samples were also evaluated using TRS. The first one was a cup filled with 0.5 mm diameter fructose powder. The second was a solid tissue phantom, prepared according to Swartling et al.[24] The 6.5 cm diameter and 5.5 cm high epoxy phantom contained TiO_2 particles as scattering centers and toner powder as an absorber.

Prior to each sample measurement, an instrumental response function was recorded by connecting the transmitting and receiving fibers to each end of a thin metal tube. This instrumental response function was used to determine time zero of the streak camera response and to measure the dispersion of the measured pulse due to the system characteristics.

Linear Approach Implementation. The measured signal was interpolated using the interpolation procedure described in the Theory section. In order to reduce the total number of parameters to tune (w_1, w_2, k), the following relations were established: $w_1 = w + 1$ and $w_2 = w$. For each pair (w, k), the correlation coefficient $r(t_0(\lambda_j, \lambda_j)$ was evaluated for 271 wavelengths of the apple measurements. Then, the performance criterion of the chosen pair (w, k) was taken as the mean of all $r(t_0(\lambda_j), \lambda_j)^2$. To improve the MADSTRESS prediction efficiency, the temporal window was chosen where the signal was significantly above zero. Using the interpolated signal and the estimated $t_0(\lambda)$, $\mu_a(\lambda)$ was evaluated by means of classical MLR. $\mu_s'(\lambda)$ was then evaluated using Eq. 5. Two solutions were obtained. The one satisfying $\mu_s'(\lambda) \gg \mu_a$ was chosen.

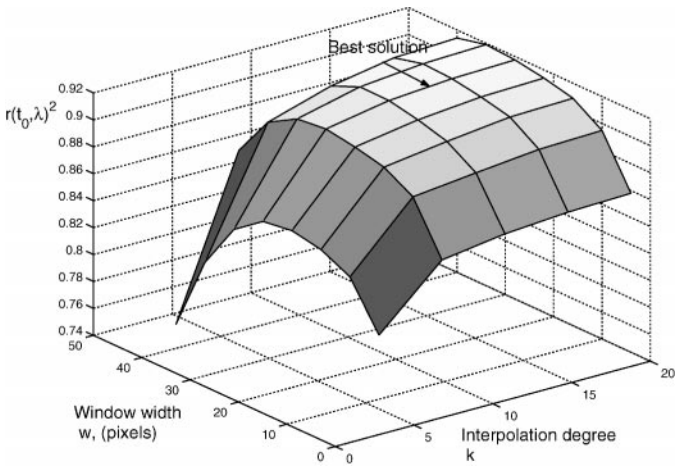


FIG. 5. (w, k) tuning regarding correlation coefficient $r(t_0, \lambda)$.

RESULTS AND DISCUSSION

Time-Resolved Spectroscopy Measurements. Figure 4a shows the light continuum that irradiates the sample. The temporal width was about 23 ps full-width at half-maximum (FWHM) and the spectral width was about 300 nm FWHM. The spectral profile was very sensitive to changes in the laser intensity and variations in the in-coupling efficiency into the ICF. For this reason the signal appeared quite disrupted, but these fluctuations did not critically influence the evaluations as MADSTRESS uses the ratio $R(t, \lambda)/R(t_0(\lambda), \lambda)$. The recorded signal from one apple is shown in Fig. 4b. The temporal dispersion is very high due to the scattering inside the apple. The measured pulse length reach 1000 ps, which implies a 20 cm (!) light path length inside the fruit, while the input/output fibers were only separated by 6 mm. The mean transit time of photons inside the fruit is 234 ps, corresponding to a mean path of 5 cm. Since photon path distribution in turbid media have a typical banana shape,[25] the mean depth probed may be estimated to 2 cm.

Parameters Setting. Figure 5 shows the evolution of $r(t_0, \lambda)^2$ with regard to (w, k) values of the interpolated mesh. The response increased rapidly when interpolation degree, k, exceeds 10 pixels. A good result was found for $w = 30$ pixels (mesh width). The following values were retained for the interpolation procedure: $w = 30$, and $k = 16$.

Figure 6 shows the peak of a TRS measurement. Despite its high intensity level, the signal peak still contains non-negligible noise. The maximum value of the TRS curve is hidden resulting in uncertainty for $t_0(\lambda)$ determination. This illustrates the $t_0(\lambda)$ determination problem and the interest of using $r(t, \lambda)$ as a means for $t_0(\lambda)$ seeking. The $r(t, \lambda)$ curve is smooth, allowing its maximum to be found easily and without any doubt. The maximum value of $r(t, \lambda)$ seems to provide a good estimation of

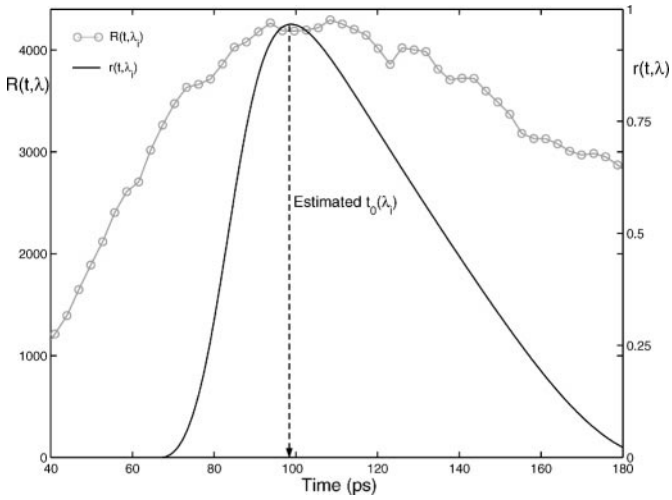


FIG. 6. Determination of $t_0(\lambda)$ for a TRS measurement at $\lambda = 910$ nm.

$R(t, \lambda)$ maximum. Without the $r(t, \lambda)$ function, t_0 determination would have been harder.

MADSTRESS Performance. ***Comparison of Experimental and Analytical Data.*** Figure 7a illustrates four temporal dispersion curves measured on an apple. The MADSTRESS analysis has provided $\mu_a(\lambda)$ and $\mu_s'(\lambda)$ values for each of the four curves. Using the diffusion equation with the estimated coefficient values, analytical TRS curves were constructed. In order to get closer to reality, the width of the irradiating peak was also taken into account. The fitted curves are plotted in black in Fig. 7a. The raw signal and the fitted signal are very similar and difficult to separate. However, a small delay may be observed in the beginning of the rising edge of the curves. This could be due to the temporal width of the irradiating peak, which is neglected in the linear approach.

In order to get a more precise idea of the accuracy of our method, the determination coefficient between the measured TRS curves, $R(t, \lambda)$, and the fitted ones, $R^*(t, \lambda)$, have been calculated for the 271 wavelengths of an apple measurement. The histogram of the calculated determination coefficient is presented in Fig. 7b. The mean value of the determination coefficients is 0.997, which is clearly a high performance. Even the lower values (0.992) demonstrate the good performance of the MADSTRESS method.

Reduced Scattering Coefficient $\mu_s'(\lambda)$**.** In Fig. 8, the reduced scattering coefficients obtained with MADSTRESS are shown. There is a large difference between $\mu_s'(\lambda)$ depending on the sample. Fructose powder logically appears as the most scattering sample. One way of analyzing the results in detail is to use Mie theory, which states that the scattering coefficient may be approximated by the relation

$$\log(\mu_s'(\lambda)) = -b \log(\lambda) + \log(a)$$

where a is proportional to the density of scattering particles and b is linked to the mean size of the scattering particles (Mie diameter). Whereas b values for the apple and the phantom were near 0.5 (respectively, 0.53 and 0.58), the fructose powder b value attains 1.8. Large sizes of the scattering particles of fructose may explain the slope differences. There was a difference in slope between the apple scattering coefficient and the mean of the 15 apple coefficients. We also found that $\mu_s'(\lambda)$ was changing significantly from one apple to another: 18 cm^{-1} for the lowest $\mu_s'(\lambda)$ value, 26 cm^{-1} for the highest scattering coefficient (not shown in the figure). These results imply that the density and size of the scattering centers vary between the apples. Yet it is well known that apple cell size and porosity are different from one apple to another. Another important comment regarding fructose $\mu_s'(\lambda)$: it does not strictly follow a linear evolution. There may be several explanations for this nonlinearity. The first one is the sample morphology, which includes crystallinity that might be important. Another important property of fructose is its light polarization ability. Hence the electric field of the scattering centers will increase the wavelength variability of the scattering coefficient by changing the scattering efficiency coefficient.

Absorption Coefficient $\mu_a(\lambda)$**.** Figure 9a presents $\mu_a(\lambda)$ calculated values using MADSTRESS, from which different absorption peaks may be identified. The phantom sample has a relatively plane and linear absorption coefficient, which was an objective during its making. However, since the exact composition of the toner powder was not known it was not possible to go further in its spectrum analysis.

The fructose presents an absorption peak at 910 nm. This band is a C–H stretch third overtone, which has already been attributed to sugar by Golic et al.[27] Another

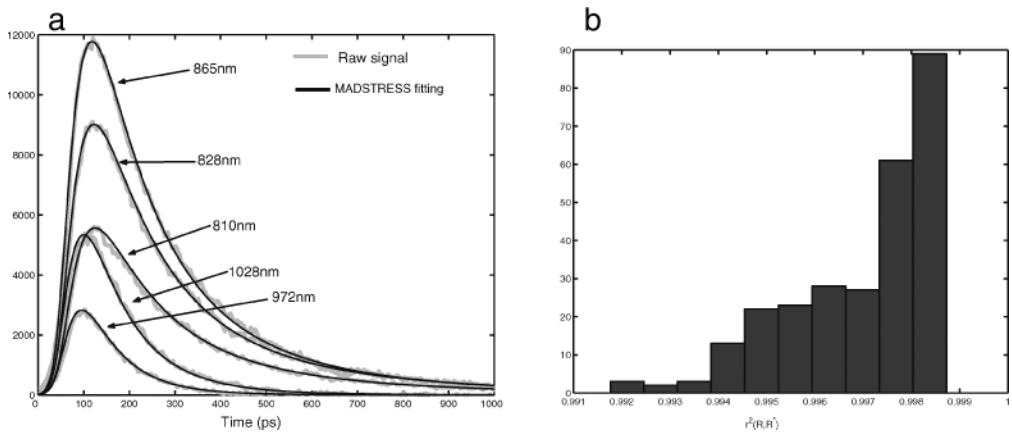


FIG. 7. Similarities between measured and fitted data. (**a**) Five examples. (**b**) Histogram of determination coefficients.

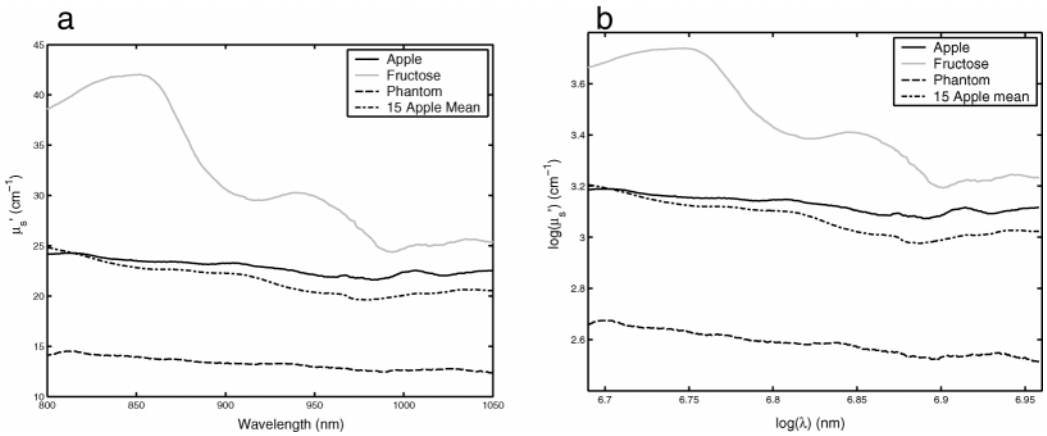


FIG. 8. $\mu_s'(\lambda)$ for the studied samples. (**a**) Normal scale. (**b**) Logarithm scale.

wide peak, also due to fructose, is found near 1010 nm. This peak is due to 2 × C–H + 3 × C–H vibration bands. In-depth analysis also reveals a small peak at 880 nm corresponding to the $C–H_3$ third overtone.

Apple $\mu_a(\lambda)$ contains a dominant peak at 970 nm. This well-known absorption band is due to water, making up 80% of an apple's content. The width of the water peak is large due to different species of water giving rise to different absorption bands (960 nm and 984 nm). The peak also overlaps the previously identified fructose peak, but that peak is still detectable. The bands at 880 nm and 905 nm have already been identified as fructose wavelengths.[28,29]

Another conclusion can be drawn regarding the second-derivative mean spectra of the apples and fructose (Fig. 9b). The derivatives were calculated using a Savitsky–Golay procedure with a window 41 pixels wide. Fructose peaks are found at 880 nm, 910 nm, and 1005 nm. In all three cases the peaks are shifted in the apple spectrum. This phenomenon can be explained by water interacting with the sugar. This effect is also enhanced by changes in sample temperature.[27]

Using the 15 apple absorption coefficients with regard to measured sugar content, an MLR prediction model has been calibrated using four wavelengths. The wavelengths (814 nm, 828 nm, 912 nm, and 1005 nm) were selected by a stepwise algorithm using leave-one-out cross-validation. The model performance (Fig. 10) was very good, with a determination coefficient of 0.92 and standard error of calibration (SEC) of 0.51 °Brix. It is interesting to note that two of the fructose wavelengths were selected by the stepwise procedure. The good performance of the model strengthens the observations made about the fructose peaks.

CONCLUSION

Light continuum generation using photonic crystal fibers has made it possible to efficiently conduct multispectral time-resolved measurements. The power of the newly developed instruments leads to the desire to use the spectral dimension during the evaluation of the scattering and absorption coefficients of the samples in order to improve accuracy. In this study a method for evaluation of the absorption and scattering coefficients using time-resolved reflectance measurements was investigated. The MADSTRESS method is based on a linear regression and a two-dimensional interpolation procedure. The method allowed us to calculate absorption coefficients and scattering coefficients of apples and fructose powder. The accuracy of the method was good enough to provide the identification of fructose absorption peaks and the construction of a calibration model predicting the sugar content of apples. MADSTRESS, as a powerful method

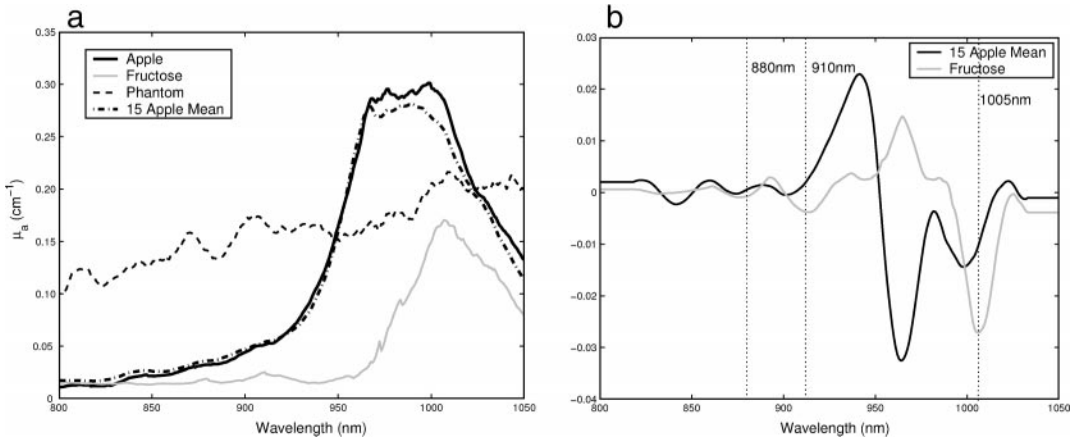


FIG. 9. Results of absorption coefficient evaluation. (**a**) $\mu_s'(\lambda)$ for the studied samples. (**b**) Second derivative of apple and fructose.

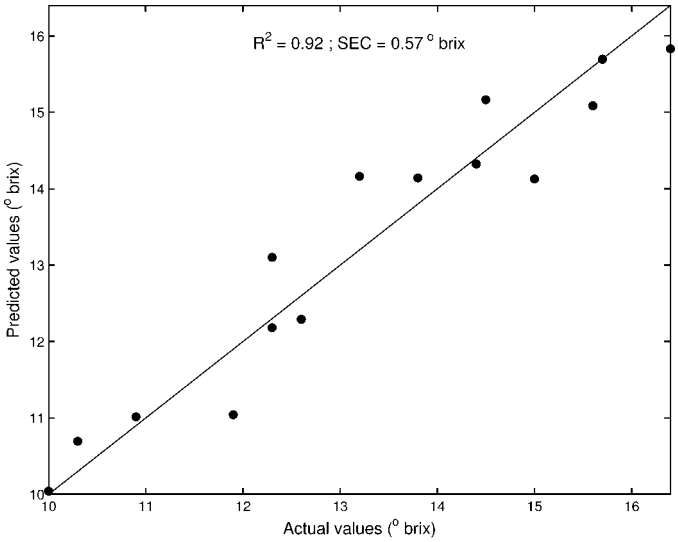


FIG. 10. Prediction of sugar concentration in apples using four wavelengths.

for processing data from spectrally continued TRS data paves the way to accurate determination of scattering and absorption coefficients in domains as varied as agricultural and food products, pharmaceutical, chemistry, and medecine. It will serve as a reference basis for chemometricians who develop specific methods to remove the scattering signal from UV-Vis and NIR spectra.

ACKNOWLEDGMENT

This work was supported by the Integrated Initiative of Infrastructure project LASERLAB-EUROPE, Contract No. RII3-CT-2003-506350.

1. B. Osborne and T. Fearn, *Near Infrared Spectroscopy in Food Analysis* (John Wiley and Sons, New York, 1986).
2. V. Bellon and G. Boisde, ''Remote Near Infrared Spectrometry in the Food Industry with the use of Silica and Uoride Glass Fiber,'' in *OE'Lase*, Proc. SPIE-Int. Soc. Opt. Eng. **1055,** 350 1989.
3. M. A. Dempster, B. F. MacDonald, P. J. Gemperline, and N. R. Boyer, Anal. Chim. Acta **310,** 43 (1995).
4. V. A. McGlone, R. B. Jordan, and P. J. Martinsen, Postharvest Biology and Technology **26,** 135 (2002).
5. V. Centner, J. VerduÁndrs, B. Walczak, D. Jouan-Rimbaud, F. Despagne, L. Pasti, D.-L. Massart, and O. de Noord, Appl. Spectrosc. **54,** 1620 (2000).
6. R. Barnes, M. Dhanoa, and S. Lister, Appl. Spectrosc. **43,** 772 (1989).
7. P. Geladi, D. MacDougall, and H. Martens, Appl. Spectrosc. **39,** 491 (1985).
8. B. Chance, J. Leigh, H. Miyake, D. Smith, S. Nioka, R. Greenfeld, M. Finander, K. Kaufmann, W. Levy, and M. Young, Proc. Natl. Acad. Sci. USA **85,** 4971 (1988).
9. S. Jacques, Appl. Opt. **28,** 2331 (1989).
10. S. Andersson-Engels, R. Berg, A. Persson, S. Svanberg, and A. Jarlman, Opt. Lett. **15,** 1179 (1990).
11. J. Johansson, S. Folestad, M. Josefson, A. Sparen, C. Abrahamsson, S. Andersson-Engels, and S. Svanberg, Appl. Spectrosc. **56,** 725 (2002).
12. C. Abrahamsson, J. Johansson, S. Andersson-Engels, S. Svanberg, and S. Folestad, Anal. Chem. **77,** 1055 (2005).
13. J. Johansson, R. Berg, A. Pifferi, S. Svanberg, and L. Bjron, Photochem. Photobiol. **62,** 242 (1999).
14. P. Zerbini, M. Grassi, R. Cubeddu, A. Pifferi, and A. Torricelli, Postharvest Biology and Technology **25,** 87 (2002).
15. S. Tsuchikawa and T. Hamada, J. Agric. Food Chem. **52,** 2434 (2004).
16. S. Tsuchikawa, E. Sakai, K. Inoue, and K. Miyamoto, J. Am. Soc. Horticultural Sci. **128,** 391 (2003).
17. M. Patterson, B. Chance, and B. Wilson, Appl. Opt. **28,** 2331 (1989).
18. C. Abrahamsson, T. Svensson, S. Svanberg, and S. Andersson-Engels, Opt. Express **12,** 4103 (2004).
19. T. Farrell, M. S. Patterson, and B. Wilson, Med. Phys. **19,** 879 (1992).
20. R. Cubeddu, A. Pifferi, P. Taroni, A. Torricelli, and G. Valentini, Med. Phys. **23,** 1625 (1996).
21. S. J. Madsen, B. C. Wilson, M. S. Patterson, Y. D. Park, S. L. Jacques, and Y. Hefetz, Appl. Opt. **31,** 3509 (1992).
22. L. Leonardi and D. H. Burns, Anal. Chim. Acta **348,** 543 (1997).
23. A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic Press, New York, 1978).
24. J. Swartling, J. Dam, and S. Andersson-Engels, Appl. Opt. **42,** 4612 (2003).
25. S. F. F. Zeng and B. Chance, Appl. Opt. **34,** 3826 (1995).
26. J. Mourant, T. Fuselier, J. Boyer, T. Johnson, and I. Bigio, Appl. Opt. **36,** 949 (1997).
27. M. Golic, K. Walsh, and P. Lawson, Appl. Spectrosc. **57,** 139 (2003).
28. K. Miyamoto and Y. Kitano, J. Near Infrared Spectrosc. **3,** 227 (1995).
29. Y. Roggo, L. Duponchel, B. Noe, and J. Huvenne, J. Near Infrared Spectrosc. **10,** 137 (2002).

Least Squares-Support Vector Machines modelisation for Time Resolved Spectroscopy

Fabien Chauchard

Information and Technologies for Agro-processes
Cemagref BP 5095, 34033 Montpellier Cedex 1, France

Sylvie Roussel

Agrometrix, www.agrometrix.com

Jean-Michel Roger, Véronique. Bellon-Maurel

Information and Technologies for Agro-processes
Cemagref BP 5095, 34033 Montpellier Cedex 1, France

Christoffer Abrahamsson, Tomas Svensson, Stefan Andersson-Engels and Sune Svanberg

Department of Physics
Lund Institute of Technology, P.O Box 118, SE-221 00 Sweden

By using time resolved spectroscopy it is possible to separate light scattering effects from chemical absorption effects in samples. In the study of propagation of short light pulses in turbid samples the reduced scattering coefficient and the absorption coefficient are usually obtained by fitting diffusion or Monte Carlo models to the measured data using numerical optimization techniques. In this study, we propose a prediction model obtained using a semi-parametric modeling technique : the Least-Squares Support Vector Machines. The main advantage of this technique is that it uses theoretical time dispersion curves during the calibration step. Predictions can then be performed using data measured on different kinds of samples, such as apples.

OCIS codes: 300.6500 spectroscopy, time-resolved , 000.3860 mathematical methods in physics, 170.3660 light propagation in tissue, 290.7050 turbid media

1. Introduction

Striking advances have been made in time resolved spectroscopy (TRS)[1] . While conventional near infrared spectroscopic measurements are influenced by light scattering in the sample, TRS deconvolutes absorption from scattering effects. The scattering properties of a sample is dependent on the physical properties of the samples, while the absorption is mostly dependent on the chemical composition of the samples. TRS was first developed for medical applications[2,3] , but is now extended to other fields, such as pharmaceutical applications[4] and agricultural applications[5,6] . TRS uses short laser pulses, of a few picoseconds, to irradiate a sample. The light signal diffusively remitted by the sample at a given distance from the irradiation point is then temporally recorded[7] . The recordings can be made in either reflection or transmission mode. In order to simultaneously measure the temporal signal at different wavelengths, new techniques using a streak camera for detection have been proposed. There are different ways to obtain light pulses with a broad wavelength profile; one is using continuum generation by focusing a high power laser pulse in a cuvette of water.[8] The development of photonic crystal fibers has further simplified the instrumental setups for continuum generation[9] .

Once the 2-dimension signal is recorded, the reduced scattering coefficient (μ_s') and the absorption coefficient (μ_a) are obtained by linking the experimental data with theoretical or modeling data. This step is crucial to get correct results, and many methods have been proposed. Three approaches are usually found: Monte Carlo simulations[7,10] , numerical optimizations[11,12] and analytical descriptors of temporal dispersion[13] .

Since the signal cannot be described by a linear equation, a non-linear multivariate model is required. Semi-parametric methods, such as kernel ones, provide more understandable models than artificial neural networks. Recently Least-Square Support-Vector Machines (LS-SVM)[14] methods have been developed and applied to near infrared spectroscopy issues such as a non-linear discrimination[15,16] and quantitative predictions.[17]

This paper aims at studying LS-SVM models calibrated only using theoretical data calculated from the diffusion equation in reflectance mode. These models are then applied to predict the reduced scattering coefficient and the absorption coefficient of experimental data.

2. Theory

2.A. Diffusion Equation

Photon transport in turbid media is described by the radiative transport equation[18] :

$$\frac{1}{c}\frac{\partial L(r,\mathbf{s},t)}{\partial t} + \mathbf{s}.\nabla L(r,\mathbf{s},t) + (\mu_s+\mu_a)L(r,\mathbf{s},t) = \mu_s \int_{4\pi} L(r,\mathbf{s},t)p(\mathbf{s},\mathbf{s}')d\omega' + Q(r,\mathbf{s},t) \quad (1)$$

Here L is the radiance at a given distance r from the irradiating source at time t and in direction $\mathbf{s}$. $p(\mathbf{s},\mathbf{s}')$ is the Henyey Greenstein phase function, $d\omega'$ is the angle between the

initial photon direction $\mathbf{s}$ and the new one $\mathbf{s}'$. c is the speed of light in vacuum. In order to solve this equation the sample geometry must be taken into account. In the case of a semi-infinite homogeneous medium measured in reflection, the solution is given by the time-resolved diffusion equation[7] :

$$R(\rho,t) = (4\pi Dv)^{-3/2} z_0 t^{-5/2} \exp(-\mu_a v t) \exp\left(-\frac{\rho^2 + {z_0}^2}{4Dvt}\right) \tag{2}$$

Here R is the signal measured at a given distance ρ at time t, D is the diffusion coefficient with $D(\lambda) = [3(\mu_a(\lambda) + \mu'_s(\lambda))]^{-1}$, and $z_0(\lambda) = \frac{1}{\mu'_s(\lambda)}$ is the mean path. v is the speed of light in the medium, assumed to be constant in the measured wavelength range. A theoretical database containing time-resolved curves may be easily obtained using Eq. (2). A model can be derived based on this database, which can then be used to predict μ_a and μ'_s.

2.B. LS-SVM theory

LS-SVM models constitute an alternate formulation of SVM regression[19] proposed by Suykens[14] . Whereas classical multivariate regression is built on variables (e.g. time data for TRS or wavelengths for spectroscopic data) LS-SVM methods are based on a kernel matrix $\mathbf{K}$. The raw data matrix $\mathbf{X}_{n,p}$ containing n samples with p variables (e.g. n time-resolved curves), is then replaced by the $\mathbf{K}_{n,n}$ kernel defined as:

$$\mathbf{K} = \begin{pmatrix} k_{1,1} & \dots & k_{1,n} \\ \vdots & \ddots & \vdots \\ k_{n,1} & \dots & k_{n,n} \end{pmatrix} \tag{3}$$

Here $k_{i,j}$ is given by the RBF function:

$$k_{i,j} = e^{\frac{-\left\|\mathbf{x_i^T} - \mathbf{x_j^T}\right\|^2}{\sigma^2}} \tag{4}$$

and $\mathbf{x_i^T}$ is the time response for a TRS measurement. The variable space is hence replaced by a sample space of a very high dimension where a sample is defined by its distance to the other samples contained in the database. The proper subspace for modeling is tuned with the σ^2 parameter. The higher σ^2, the wider the gaussian kernel is. Put simply, $k_{i,j}$ represents the similarities between the $\mathbf{x_i^T}$ and $\mathbf{x_j^T}$ time responses. The model equation is then:

$$\hat{\mathbf{y}} = \mathbf{K}\boldsymbol{\beta} + \beta_0 \tag{5}$$

where $\hat{\mathbf{y}}$ is the predicted value, $\mathbf{K}$ is the kernel as defined by Eq.3, $\boldsymbol{\beta}$ is the regression vector and β_0 is the offset term. Furthermore, the LS-SVM objective function takes into account

the norm of the regression vector in order to increase the model robustness. The classical squared loss function is thus replaced by the following objective function:

$$min(\mathbf{e}) = min\left[\frac{\sum_{i=1}^{n}(\mathbf{y_i} - \hat{\mathbf{y}}_\mathbf{i})^2}{2} + \frac{1}{\gamma}\frac{(\boldsymbol{\beta}^T\boldsymbol{\beta})}{2}\right] \tag{6}$$

where γ is a regularization parameter analogous to the regularization parameter of regularized artificial neural networks, is used to weigh $\boldsymbol{\beta}$ norm. Once σ^2 and γ are chosen, the model is trained after constructing the Lagrangian by solving the linear Karush-Kuhn-Tucker (KKT) system:

$$\begin{bmatrix} 0 & \mathbf{l_n^T} \\ \mathbf{l_n} & \mathbf{K} + \frac{\mathbf{I}}{\gamma} \end{bmatrix} \begin{bmatrix} \hat{b}_0 \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{7}$$

where $\mathbf{I}$ refers to an $[n \times n]$ identity matrix and $\mathbf{l}_n$is a $[n \times 1]$ unity vector. The solution of Eq. (7) can be found using most standard methods of solving sets of linear equations, such as conjugate gradient descent.

3. Material and methods

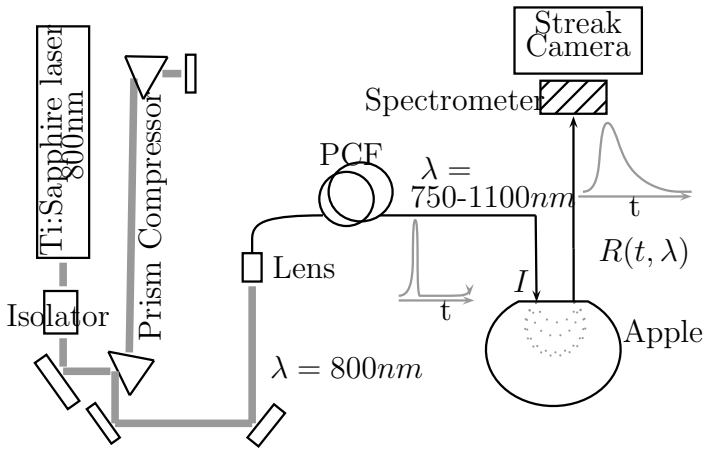


Fig. 1. Setup for TRS spectrum acquisition

3.A. Instrumentation

Fig. 1 depicts the experimental setup. The instrument has been described in detail elsewhere[9] . Briefly, a Ti:Sapphire mode-locked Laser, pumped by an Ar-ion Laser, was used to generated

100 fs pulses centered around 800 nm with 80 MHz repetition rate. The laser pulses were focused into a 100 cm long index guiding crystal fiber (ICF)(Crystal fiber A/S, Copenhagen, Denmark). The broad band light pulses generated by non-linear effects in the ICF ranged from 750 nm till 1100 nm. The light was then transferred by a set of lenses into a gradient index fiber guiding the light to the sample. Another gradient fiber, with the distal tip 6 mm from the irradiating one, was used to collect the light diffusively reflected by the sample. The fibers were put in contact with the sample. A Streak Camera (Hamamatsu, ModelC5680) coupled to an imaging spectrometer (Chromex, Model 250IS) captured the reflected light as a function of time and wavelength $R(t, \lambda)$. The spectral resolution was 0.93 nm distributed over 512 pixels while the temporal resolution was 2.93 ps in the span from 0 to 1900 ps, spread over 640 pixels.

3.B. Measured samples

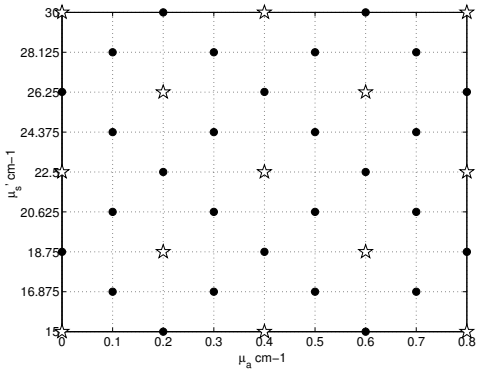
15 Golden Delicious apples were measured using the TRS setup. A small part of the apple was carefully removed in order to create a flat surface for applying the fibers. The measurements were performed immediately after the preparation of the apples in order to avoid flesh drying. Prior to each sample measurement, an instrumental response function was recorded by connecting the transmitting and receiving fibers to each end of a thin metal tube. This instrumental response function was used to determine time zero of the streak camera response and to measure the dispersion of the measured pulse due to the system characteristics.

3.C. LS-SVM model

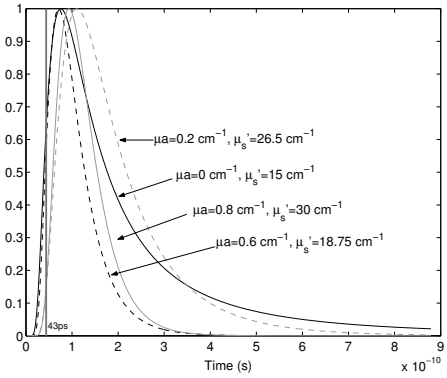
The LS-SVM model was derived using a theoretical calibration set. The data set was obtained using the diffusion equation, applied for an inter fiber distance of $\rho = 6$ mm and a time resolution of 2.93 ps. Each signal was normalized by division with its maximum in order to become independent of variations in the irradiating signal intensity level. To improve the model efficiency, the temporal window between $t = 43$ ps and 900 ps was selected, where the time dispersion curves with different optical properties were significantly different. To span the absorption and scattering variations of apples[20] , a mixture design was used as described in Fig. 2. In order to tune γ and σ^2 the training set was split into two subsets, one for calibration (subset A) and one for validation (subset B). Once the 2 parameters were chosen, the final model was constructed using the whole theoretical data set.

The LS-SVM toolbox (LS-SVM v1.4 [1], Suykens, Leuven, Belgium) was used with MATLAB 6.0 (The MathWorks, Inc., Natick, USA) to derive the LS-SVM models. In

[1] www.esat.kuleuven.ac.be/sista/lssvmlab/



a



b

Fig. 2. (a)Values of μ_a and μ'_s for the training set ; $\star$ calibration(A), $\bullet$ validation(B) for γ and σ^2 parameters tuning. (b) Four theoritical dispersion profiles used are shown

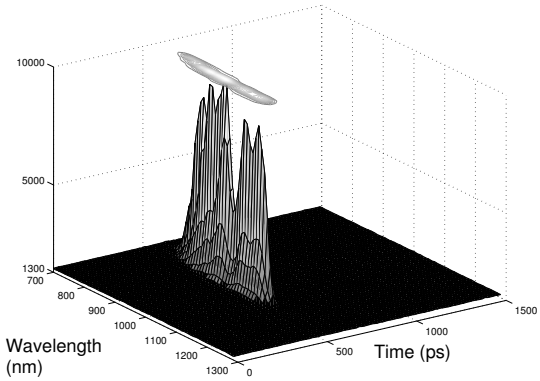
order to evaluate the accuracy of this new method, the predicted values of μ_a and μ'_s were compared to the ones fitted to the diffusion equation using a Levenberg Marquart (LMA) optimization procedure, already previously used on apple TRS measurements[20] .

4. Results and Discussion

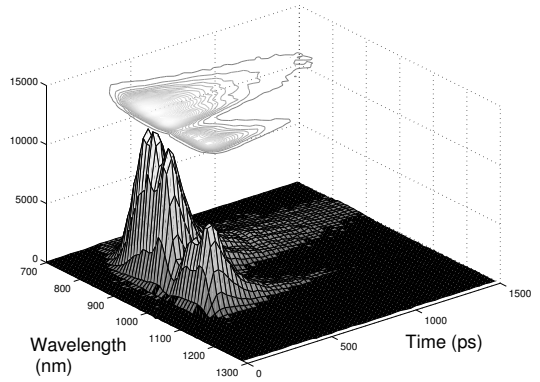
4.A. TRS measurements

Fig. 3-a shows the instrumental response function. The continuum light pulses obtained were 300 nm wide (800 nm - 1100 nm). The temporal width was about 23 ps FWHM. The

h!]



a



b

Fig. 3. Recording of a 2D time resolved measurement: (a) Multispectral light pulse used for sample irradiation (b) Recorded signal for an apple

spectral profile was very sensitive to changes in the laser intensity and variations in incoupling efficiency into the ICF. As a result, the spectral profile of the irradiating source was changing from one sample measurement to other. LS-SVM model uses the temporal signal at a given wavelength normalised in order to obtain a maximum value of 1. Hence, source intensity variation from one sample measurement to other does not act upon the model prediction efficiency.

The recorded signal from one apple is depicted in Fig. 3-b. The temporal dispersion is very high due to scattering phenomena inside the apple. Since the recorded signal-to-noise ratio was high enough in the region ranging from 800 nm to 1050 nm, this spectral window was selected for studying the optical properties of the apples.

4.B. Model tuning

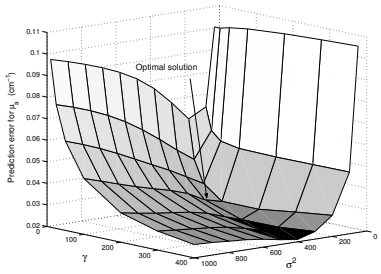


Fig. 4. Optimization surface for γ and σ^2 tuning for μ_a modelisation

The optimization responses surface for μ_a prediction is illustrated in Fig. 4. This surface represents the standard error of prediction (SEP) on the validation set B. The best prediction of μ_a was found for $\gamma = 50$ and $\sigma^2 = 500$. The μ'_s response surface (not presented here) gives the optimal solution for the same values. Since σ^2 values are the same for both μ_a and μ'_s, the kernel matrix is the same; this means that both models are built on the same subspace, allowing the same degree on non-linearities. Only the regression vectors are different for predicting μ_a and μ'_s. Low values of robustness criteria, γ, imply the regression vectors have a small norm which is necessary for a robust model.

4.C. Evaluation of scattering and absorption coefficients on experimental data

Fig. 5 compares μ_a and μ'_s values predicted by LMA and LS-SVM for one apple. The absorption coefficient curves are very similar, which proves LS-SVM prediction capabilities. In spite of the noise, the water peak is clearly visible at 970 nm as normally seen in conventional NIR spectra of fruits. Regarding the scattering coefficient the prediction values present an offset

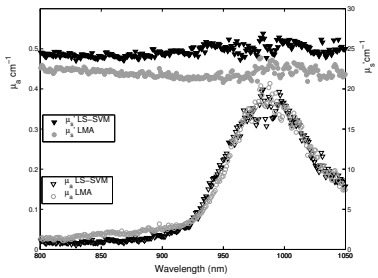


Fig. 5. Result of μ_a and μ'_s prediction on an apple for all wavelengths

compared to the LMA results. This can be explained by the temporal dispersion Fig. 6: LS-SVM model considers the irradiating peak as perfectly resolved in time (time width infinitely small), whereas LMA takes the instrumental response into account in the calculations. Since LMA is based on convolution the predicted TRS curves are closer to the measured signal. However LS-SVM produce acceptable results. LS-SVM curve are above LMA and slightly peak shifted, which explains the offset previously noticed between μ'_s values.

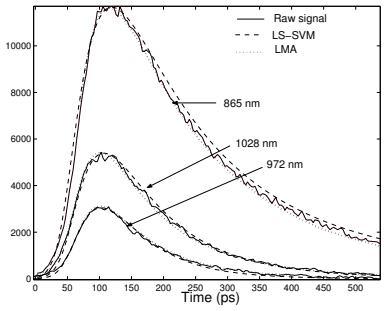


Fig. 6. Measured signal and fitted signals for 3 wavelengths

4.D. Prediction performances

Fig. 7 shows the LS-SVM predicted values versus LMA values of μ_a for the 15 apples (271 dispersion curves per sample). The determination coefficient of 0.96 is satisfactory, with a standard error of prediction of $0.02cm^{-1}$. It should be noted that there are no real reference values, but only reference values estimated by LMA. Fig. 8 shows a bias between LMA values and LS-SVM predicted values for μ'_s determination. As explained before, this difference comes from the convolution process which is not used in the LS-SVM. Since the determination

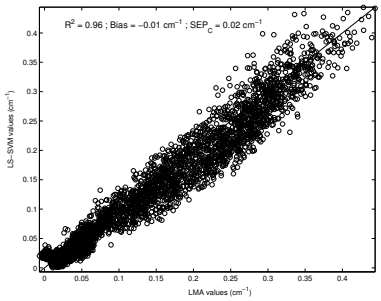


Fig. 7. Prediction performance of the μ_a prediction model

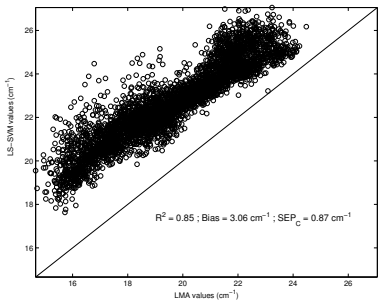


Fig. 8. Prediction performance of the μ_s' prediction model

coefficient is satisfactory (0.85), the model may be easily bias corrected by adding a constant ($-3.06cm^{-1}$). However, this approach would consider LMA values as real reference values, although LMA has also its drawbacks and inaccuracies. For this reason, it would be more interesting to follow a more sophisticated approach, integrating a convolution process in the database building. In this case, the model would be calibrated on theoretical curves obtained by convoluting the diffusion equation with the instrumental response function. Of course, this method is more time consuming since the model must be designed for each sample. When this approach is followed the prediction plot gives the results shown in Fig. 9. As assumed, the bias is reduced but is not small enough to be neglected. Furthermore, the correlation coefficient between LMA and LS-SVM values decreases to 0.75. The noise in the measured data acts differently on the two methods since they have different bases. Visual curve analysis (Fig. 10-a) is not accurate enough to judge differences between method performance. For this reason, the determination coefficient between the raw signal and the two estimated signals are presented for each wavelength in Fig. 10-b. The temporal curves

calculated using LS-SVM predicted coefficient have clearly high performance ($r^2 > 0.99$) very close to LMA one's. This tends to prove the accuracy of the proposed approach.

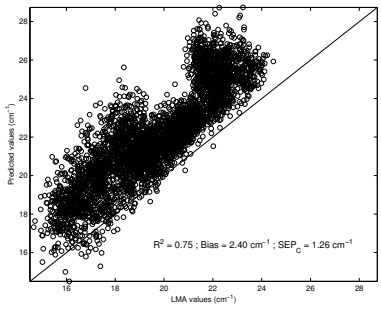


Fig. 9. Prediction performance of the μ_s' prediction model with convolution approach

4.E. Conclusion

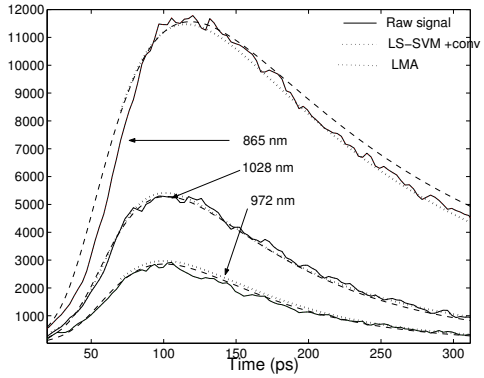
Thanks to its performance, the LS-SVM model can be applied to time-resolved data for extraction of absorption and scattering coefficients. The model proposed in this paper has two main advantages. The first one is that it can be used on any diffusing sample with $\mu_a < 0.08\text{cm}^{-1}$ and $1.5\text{cm}^{-1} < \mu_s' < 3\text{cm}^{-1}$ (but a larger model may be calibrated) such as for human tissues. The second one is that since the model uses only 41 time resolved curve for the model, it can easily be integrated into an embedded sensor for industrial use. Even if the model performance is already interesting, the method may be improved by integrating a convolution process into the database construction. In the perspective of optimization, a data smoothing may be applied on the raw data.

LS-SVM could also be used with a data base of Monte-Carlo data. This would be very interesting for measurement geometries where the diffusion approximation is not valid, e.g. where the source and the detection fibers are situated close to each other or when the boundary conditions are too complex to solve analytically.

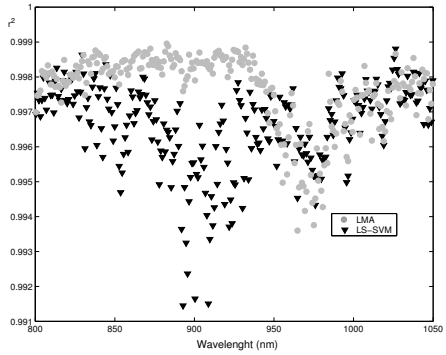
As TRS transmission measurements produce the same type of curves as the reflection geometry, LS-SVM model may also be derived and applied efficiently on transmission data (slab geometry). We also think that LS-SVM modeling would be of great interest for spatially resolved spectroscopy and phase modulation spectroscopy.

Acknowledgments

This work was supported by the Integrated Initiative of Infrastructure project LASERLAB-EUROPE, Contract No. RII3-CT-2003-506350.



a- Predicted signals for 3 wavelengths



b-r^2 between curve for each wavelength

Fig. 10. Measured signal and fitted signals with LS-SVM using convolution and determination coefficient for the spectral range

References

1. B. Chance, J. Leigh, H. Miyake, D. Smith, S. Nioka, R. Greenfeld, M. Finander, K. Kaufmann, W. Levy, and M. Young, "Comparison of time-resolved and unresolved measurements of deoxyhemoglobin in brain," Proc. Natl. Acad. Sci. USA **85**, 4971–4975 (1988).
2. S. Jacques, "Time-resolved propagation of ultrashort laser pulses whithin turbid tissue," Appl. Optics **28**, 2331–2336 (1989).
3. S. Andersson-Engels, R. Berg, A. Persson, S. Svanberg, and O. Jarlman, "Time-resolved transillumination for medical diagnostics," Opt.Lett **15**, 1179–1191 (1990).
4. J. Johansson, S. Folestad, M. Josefson, A. Sparen, C. Abrahamsson, S. Andersson-Engels, and S. Svanberg, "Time-resolved NIR/vis spectroscopy for analysis of solids: Pharmaceutical tablets," Applied Spectroscopy **56**, 725–731 (2002).

5. P. Zerbini, M. Grassi, R. Cubeddu, A. Pifferi, and A. Torricelli, "Nondestructive detection of brown heart in pears by time-resolved reflectance spectroscopy," Postharvest Biology and Technology **25**, 87–97 (2002).
6. J. Johansson, R. Berg, A. Pifferi, S. Svanberg, and L. Bjorn, "Time-resolved studies of light propagation in Crassula and Phaseolus Leaves," Photochemistry and Photobiology **62**, 242–247 (1999).
7. M. Patterson, B. Chance, and B. Wilson, "Time-resolved reflectance and transmittance for the non invasive measurement of tissue optical properties," Appl. Opt. **28**, 2331–2336 (1989).
8. S. Andersson-Engels, R. Berg, A. Persson, and S. Svanberg, "Multispectral tissue characterization with time resolved detection of diffusely scattered white light," Opt. Lett **18**, 1697–1699 (1993).
9. C. Abrahamsson, T. Svensson, S. Svanberg, and S. Andersson-Engels, "Time and wabvelength resolved spectroscopy of turbid media using light continuum generated in a crystal fiber," Optics Express **12**, 4103–4112 (2004).
10. T. Farrell, M. S. Patterson, and B. Wilson, "A diffusion theory of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo," Med. Phys. **19**, 879–888 (1992).
11. R. Cubeddu, A. Pifferi, P. Taroni, A. Torricelli, and G. Valentini, "Experimental test of theoretical models for time resolved reflectance," Med. Phys **23**, 1625–1633 (1996).
12. S. Madsen, B. Wilson, M. Patterson, Y. Park, S. Jacques, and Y. Hefetz, "Experimental tests of a simple diffusion model for the estimation of scattering and absorption coefficients of turbid media from timeresolved diffuse reflectance measure," Appl. Opt. **31**, 3509–3517 (1992).
13. L. Leonardi and D. H. Burns, "Quantitative constituent in scattering media from statistical analysis of photon time-of-flight distributions," Analytica Chimica Acta **348**, 543–551 (1997).
14. J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines* (World Scientific Publishing Co., Singapore, 2002).
15. A. Belousov, S. Verzakov, and J. von Frese, "Applicational aspect of support vector machine," J. of Chemometrics **16**, 482–489 (2002).
16. R. Goodacre, "Explanatory analysis of spectroscopic data using machine learning of simple, interpretable rules," Vibrational Spectroscopy **32**, 33–45 (2003).
17. F. Chauchard, R. Cogdill, S. Roussel, J. Roger, and V. Bellon-Maurel, "Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes," Chemometrics and Intelligent Laboratory Systems **71**, 141–150 (2004).

18. A. Ishimaru, *Wave propagation and scattering in random media* (Academic Press, New York, 1978).
19. V. Vapnik and A. Lerner, "Pattern recognition using generalized portrait method," Automation and Remote Control **24**, 774–780 (1963).
20. R. Cubeddu, C. Andrea, A. Pifferi, P. Taroni, A. Torricelli, G. Valentini, C. Dover, D. Johnson, M. Ruiz-Altisent, and C. Valero, "Nondestructive quantification of chemical and physical properties of fruits by time-resolved reflectance spectroscopy in the wavelength range 650-1000 nm," Applied Optics **40**, 538–543 (2001).

Time-Resolved NIR/Vis Spectroscopy for Analysis of Solids: Pharmaceutical Tablets

JONAS JOHANSSON,* STAFFAN FOLESTAD, MATS JOSEFSON, ANDERS SPARÉN,† CHRISTOFFER ABRAHAMSSON, STEFAN ANDERSSON-ENGELS, and SUNE SVANBERG

Analytical Development, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden (J.J., S.F., M.J., A.S.); and Department of Physics, Lund Institute of Technology, P.O. Box 118, SE-221 00 Lund, Sweden (C.A., S.A.-E., S.S.)

Time-resolved spectroscopy in the visible and near-infrared (NIR) regions was used in a feasibility study for analysis of solid pharmaceuticals. The objective of the experiments was to study the interaction of light with pharmaceutical solids and to investigate the usefulness of the method as an analytical tool for spectroscopic analysis. In these experiments, a pulsed Ti:sapphire laser and white light generation in water was utilized to form a pulsed light source in the visible/NIR region. The light was focused onto the surface of tablets, and the transmitted light was detected by a time-resolving streak camera. Two types of measurements were performed. First, a spectrometer was put in front of the streak camera for spectral resolution. Secondly, the signal originating from different locations of the sample was collected. Time-resolved and wavelength/spatially resolved data were generated and compared for a number of different samples. The most striking result from the experiments is that the typical optical path length through a 3.5-mm-thick tablet is about 20–25 cm. This indicates very strong multiple scattering in these samples. Monte Carlo simulations and comparison with experimental data support very high scattering coefficients on the order of 500 cm^{-1}. Furthermore, the data evaluation shows that photons with a particular propagation time through the sample contain a higher chemical contrast than other propagation times or than steady-state information. In conclusion, time-resolved NIR spectroscopy yields more information about solid pharmaceutical samples than conventional steady-state spectroscopy.

Index Headings: **Time-resolved spectroscopy; Near-infrared; Transmission; Diffuse reflectance.**

INTRODUCTION

Near-infrared (NIR) spectroscopy[1,2] has been shown to be a highly useful tool for analysis of pharmaceuticals.[3–6] NIR analysis has been used for many diverse applications in the pharmaceutical industry, such as for qualitative and quantitative analysis of powders, pellets, whole tablets, tablets in their blister packages, freeze-dried products, etc. Among the advantages of NIR analysis, it can be mentioned that NIR is nondestructive, fast, and can be performed remotely through optical fibers—factors that make NIR analysis particularly useful for process analysis.[7] In addition, spectroscopic analysis of solids may offer probing of solid state properties such as crystallinity and sample density, parameters that are entirely lost by chromatography and other wet-chemistry methods. The discussion here is confined to pharmaceutical tablets; however, the method reported in this paper is applicable for other types of turbid samples.

Received 14 November 2001; accepted 14 February 2002.

* Author to whom correspondence should be sent.

† Present address: Quality Control and Assurance, AstraZeneca Tablet Production Sweden, SE-151 85 Södertälje, Sweden.

Advances in NIR analysis have been substantial in instrumental development as well as in the development of data evaluation tools, mainly based on multivariate analysis. The instrumental development includes the use of different detection principles such as the Fourier transform spectrometer, the diode array detector, and the acousto-optic tunable filter approach.[8] Instrumental developments also include improvements in the sample presentation, such as fiber-optic probes and the use of transmission geometry rather than the traditional reflectance geometry.[9–12] Comparing the two geometries, transmission NIR has shown better precision than reflectance due to the fact that the entire tablet volume is probed by the light in the case of transmission geometry.[11,12]

There are, however, also some drawbacks to spectroscopic techniques in general and to NIR spectroscopy in particular. Perhaps the most important drawback stems from the fact that NIR analysis of turbid media does not strictly follow the Beer–Lambert law. In other words, the reflected or transmitted intensity is not a simple function of the absorptivity of the sample, but is also affected by multiple scattering of light in the sample volume. In fact, since in the red/NIR optical region scattering is more prominent than absorption, variations in scattering properties of a sample may alter a measured signal more than a corresponding change in absorption properties. The light scattering properties of the sample matrix are strongly affected by its physical parameters, such as tablet hardness and porosity. The more micro-cavities and variations in the refractive index of the matrix, the more prominent the light scattering will be, resulting in wavelength-dependent attenuation. Other physical parameters of relevance for tablets are the height and diameter. Small changes in the tablet diameter, shape, or thickness have strong effects on the transmission NIR spectrum. Similarly, engravings and scores on the tablet surface will have a strong effect on the reflectance NIR spectrum. The fact that NIR analysis is perhaps more sensitive to physical parameters than to chemical content limits the utility of the technique. The most common way to deal with these problems is the use of numerical transformations and spectral pretreatment before analyzing the spectroscopic data. In this context, chemometrics has evolved as a very powerful tool to extract minute variations from complex spectral data sets.

Another limitation of NIR spectroscopy of pharmaceuticals is that a strong light scattering in the matrix results in varying path lengths through the tablet. The path length is critical to how effectively the absorption will

0003-7028/02/5606-0725$2.00/0

attenuate the signal. In a situation where the chemical content is evaluated from the recorded transmission spectrum, chemometric methods are frequently used to extract the absorption properties of interest. It is thus extremely important that this evaluation algorithm is trained using representative tablets. The quality of NIR measurements is clearly limited by small variations in the physical properties of the tablets.

To be able to measure and evaluate the effects of both absorption and scattering and to obtain a better understanding of the light propagation and attenuation in the sample, more sophisticated measurement techniques are required. Such experimental tools have been developed mainly for studying biological tissue and can roughly be divided into time- or frequency-resolved and spatially resolved techniques.[13] Using one of these general approaches, the optical properties of the sample can be elucidated. Hence, time- or spatially resolved spectroscopy have been used to map, for instance, human breast,[14] brain,[15] and skin[16] tissues, as well as green leaves from plants.[17] The optical properties of pharmaceutical solids have also been measured, e.g., in connection with determination of absorption coefficients for powders[18] and effective sample size.[19] One interesting feature found in all these studies was the strong domination of scattering over absorption for some but not all wavelength regions. The optical properties of a turbid sample can be described by the absorption coefficient μ_a, the scattering coefficient μ_s, and the scattering anisotropy g. Together, these three parameters characterize the interaction of light with the sample. Often, μ_s and g are represented as the reduced scattering coefficient $\mu_s' = (1 - g)\mu_s$. In order to model the light propagation in the sample, either hard models, such as diffusion theory, or numerical methods, such as Monte Carlo modeling, are utilized and fitted against the experimental data. In addition, chemometric tools are often used in connection with numerical fitting algorithms for a better characterization of the sample.

In this paper we have explored the use of time-resolved transmission spectroscopy as an analytical tool for pharmaceutical tablets. Both wavelength-resolved and spatially resolved data have been obtained at a high time resolution by using ultra-short pulsed lasers and a streak camera. The main objectives of the investigation were to estimate the degree of light scattering in this type of tablet and to evaluate whether this more sophisticated measurement technique can improve NIR spectroscopy. The experiments were mainly carried out in the visible rather than in the NIR optical region, although the aim is to extend future measurements further into the NIR region, where characteristic absorption features may enhance the analytical quantitative information. The obtained data are discussed in relation to conventional NIR spectroscopy to indicate the potential advantages of time-resolved spectroscopy over steady-state techniques.

MATERIALS AND METHODS

Optical Set-Up. A system for ultra-short white light illumination and picosecond (ps) time-scale detection was utilized to measure the photon propagation time through the sample tablets. The general-purpose light source was a table-top tera-watt laser system located at the Lund Laser Centre.[20] It used an Ar-ion laser pumping a Ti:sapphire laser as a master oscillator and one or two Nd:YAG laser-pumped amplifier stages. The Ti:sapphire laser was passively mode-locked at a frequency of 76 MHz to yield 100 femtosecond (fs) pulses at a wavelength near 800 nm. Before amplification, the pulses were sent to a pulse stretcher consisting of two gratings yielding different path lengths for different wavelengths within the line profile, creating chirped pulses. After pulse stretching, the pulses had a pulse length of about 260 ps. A Pockels cell was used to select pulses for the amplifier stages at a repetition rate of 10 Hz. The amplifier stages employed Q-switched Nd:YAG laser pumped Ti:sapphire crystals. After amplification in two stages the pulse energy could be as high as 450 mJ. The amplified pulses were sent to a compressor, where another pair of gratings was used to compress the pulses back to about 200 fs. The compressor had a transmission efficiency of about 50%. These transmission studies required substantially lower pulse energies than those available, and a maximum of 50 mJ was utilized.

The experimental set-up used for the transmission studies is shown in Fig. 1. In the case of the wavelength-dispersed measurements, the light pulse train from the laser system was focused into a water-containing cell using a 15-cm-focal-length lens (L1) for white light generation. Self-phase modulation of the refractive index was induced, resulting in a spectral profile ranging from the UV to the NIR region. As a result, laser pulses of 200 fs duration were accessible at all the desired wavelengths. The light from the white light generator was made parallel with an achromatic camera lens (L2). The light was focused with a 100-mm camera lens (L3) onto the sample, which was fixed in an iris holder allowing light to pass through the sample. The average light power used at the sample surface was approximately 1 mW and the size of the laser spot on the sample was approximately 1 mm. The transmitted light was imaged onto the 120-μm entrance slit of a spectrometer (Oriel, Model 77480) using two 50-mm camera lenses (L4 and L5). The spectrometer was equipped with one of two different gratings optimized for the blue or NIR optical region. The output from the spectrometer was imaged onto a streak camera (Hamamatsu, C1587) using a pair of camera lenses (L6 and L7) providing an overall magnification of either 1 or 0.4. In the case of spatially resolved measurements, the white light generator was removed and the light transmitted through the sample was directly imaged onto the streak camera without passing though the spectrometer. The streak camera received a trigger signal from a photodiode close to the Ti:sapphire laser. The trigger signal was preamplified, stabilized by a constant fraction discriminator, and delayed in a delay generator before it was sent to the streak camera control unit. The signal from the streak camera was fed to a PC and the time-dispersed transmission curves were stored and displayed on the monitor. The signal from up to 100 laser pulses was integrated to allow a sufficient signal-to-noise ratio. The time resolution of the streak camera was about 20 ps and the time window was about 2 ns. The optical set-up allowed a fraction of the excitation light to be split off and sent directly to the spectrometer slit in order to have a fixed time reference in the recorded time-resolved spec-

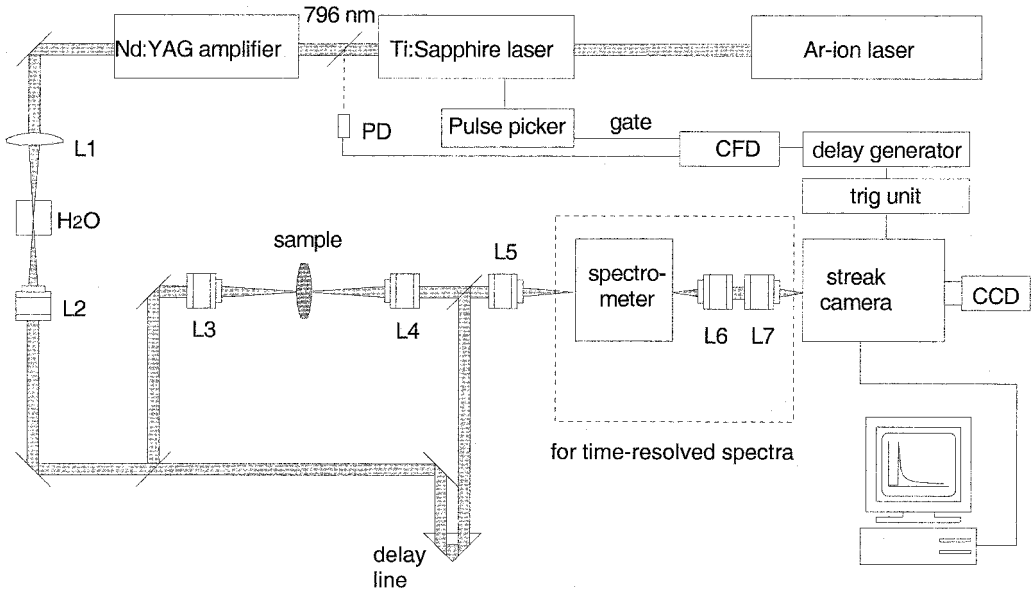


FIG. 1. Experimental set-up for time-resolved transmission of tablets. ***L1***: lens, ***L2–L7***: achromate camera lenses, ***PD***: photo diode, ***CFD***: constant fraction discriminator.

tra. Using the reference pulse it is possible to partly correct for the electronic time jitter that limits the time resolution.

Samples. ***Model Samples.*** Solutions of varying concentrations of intralipid were used as model substances. Different amounts of ink were added to the intralipid solution in order to obtain different absorption coefficients for the sample solutions. Detailed experimental conditions are given elsewhere.[21]

Tablets. The manufacturing and experimental design has been described in detail previously[12] and will only be described here in part. The tablets consisted of film-coated pellets (AstraZeneca R&D Mölndal) and excipients, and were manufactured according to an experimental design in which several tablet parameters were varied. The most abundant component in these tablets was microcrystalline cellulose. The heights of the tablets were 3.5 and 4 mm and the diameter was 9.0 mm. The tablets were stored at room temperature until the measurements were taken.

Data Evaluation. The transmission data obtained from the streak camera were primarily displayed as 3D matrices. In Fig. 2, a data matrix from a spatially resolved measurement is shown. The transmitted intensity is shown as a function of photon migration time (abscissa) and position of light exiting the sample (ordinate). The main information comes from the light area in the center of the image. The smaller white areas in the left part of

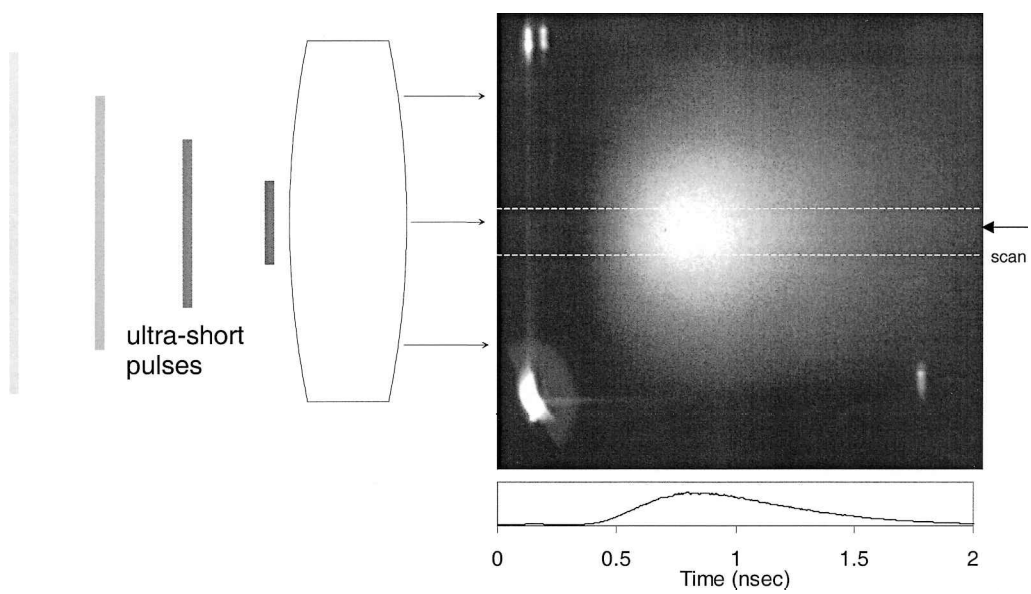


FIG. 2. Time-resolved transmission of a tablet using 200 fs pulses at 790 nm. The transmitted intensity is displayed as a function of time (abscissa) and position on tablet (ordinate). Below the picture is a scan across the picture at the position indicated between the dashed lines. The bright spots to the far left in the picture originate from light leaking beside the tablet, thus with very little time retardation.

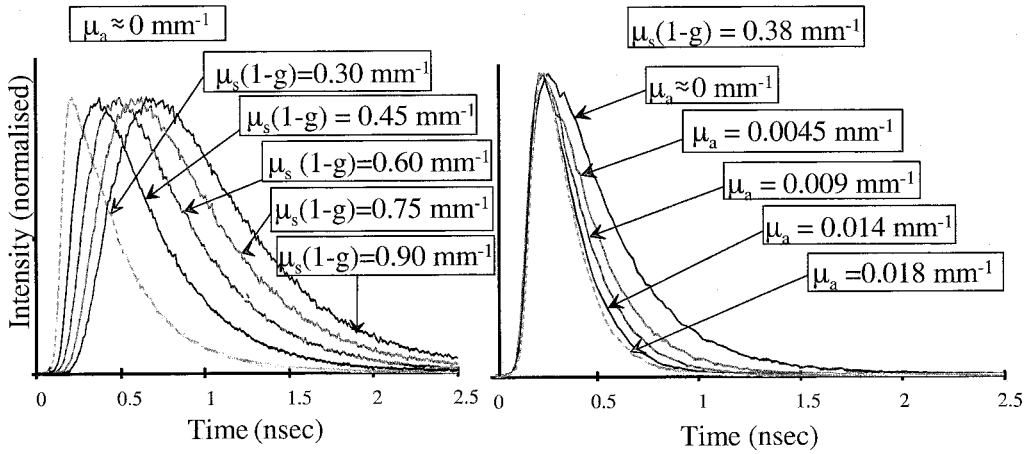


FIG. 3. Time-resolved transmission spectra of solutions of Intralipid and ink. The Intralipid concentration was varied with zero ink concentration (left graph) and held constant for varying concentrations of ink (right graph).

the image are due to light leakage between the sample holder and the sample. A horizontal scan through the image (between the dashed lines) results in a time profile at a given position on the sample, as is shown below the image in Fig. 2 for the center of the sample. In the same way, vertical scans through the image may be extracted (not shown). These are consequently interpreted as time-gated one-dimensional images (or spectra, if the spectrometer is used).

Monte Carlo Simulations. Monte Carlo simulations were utilized to model the light transport through the tablets. One of the most used and best validated implementations of Monte Carlo simulations of photon transport in multilayered tissues (MCML)[22] was used. The routine was altered for this application to allow a cylindrical geometry mimicking a tablet for which the radius and thickness were changed.

RESULTS AND DISCUSSION

Absorption and Scattering in Turbid Media. Time-resolved transmission experiments were performed on pharmaceutical tablets to roughly estimate the optics of the tablets. In addition to the streak camera measurements reported here, some measurements on tissue phantoms are discussed at this point in order to improve the understanding of the experimental data. The time-resolved transmission data of tissue phantoms were recorded with a time-correlated single-photon counting technique at 640 nm rather than with the streak camera and have been described in detail previously.[21] The tissue phantom consisted of a series of Intralipid water solutions with ink added as an absorber. By varying the particle concentration of the Intralipid and the ink concentration, both the scattering coefficient and absorption coefficient were varied. In Fig. 3 (left graph) it is observed that by increasing μ_s' the transmission profile is shifted towards longer times. On the other hand, when μ_a is increased (right graph) merely the late slope of the profile is affected, leaving the peak position almost unchanged. For very high μ_s' the profiles approach a symmetric diffusion profile, whereas for high μ_a the curves collapse to Beer–Lambert's law. Hence, from the analysis of time-resolved transmission profiles, it is possible to qualitatively understand the optical paths of the samples. An interesting observation from Fig. 3 is that at early gate times changes in μ_a have a very small effect on the measured light intensity, whereas μ_s' affects the early transmitted light to a much higher degree.

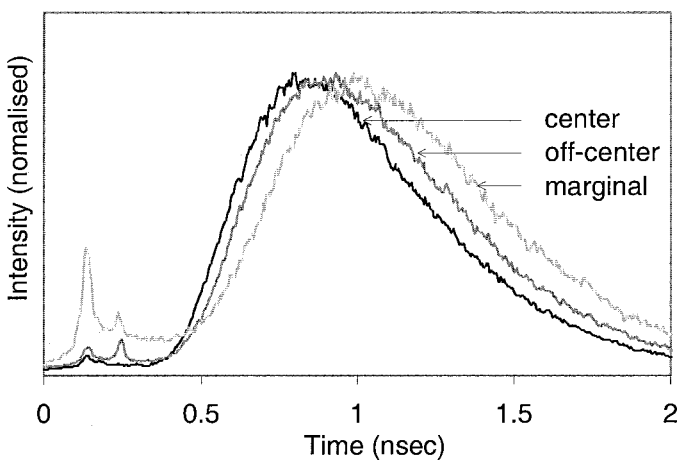


FIG. 4. Time-resolved transmission of a 3.5-mm-thick tablet using 200 fs pulses at 790 nm. The profiles correspond to light exiting the tablet at the center, 2 mm off the center, and close to the margin of the tablet, respectively. The curves are autoscaled to the same maximum intensity.

In Fig. 4, time-resolved transmission profiles extracted from the image in Fig. 2 are shown. The different time profiles correspond to horizontal scans through the center (region marked between dashed lines), the mid-upper, and the upper part of the data image. These in turn, can be interpreted as light exiting through the center part, the off-center part, and close to the margin of the back side of the tablet. The light was focused onto the central part of the front side of the tablet as illustrated in Fig. 2. A most striking observation is the high mean photon propagation time, approximately 1 ns. Assuming a refractive index of 1.4, this corresponds to a mean optical path length of more than 200 mm for a tablet with a thickness of 3.5 mm. This indicates extraordinarily strong light scattering, also supported by the symmetric diffusion-like time profile. Monte Carlo simulations were performed for a tablet-shaped object and compared with the experimen-

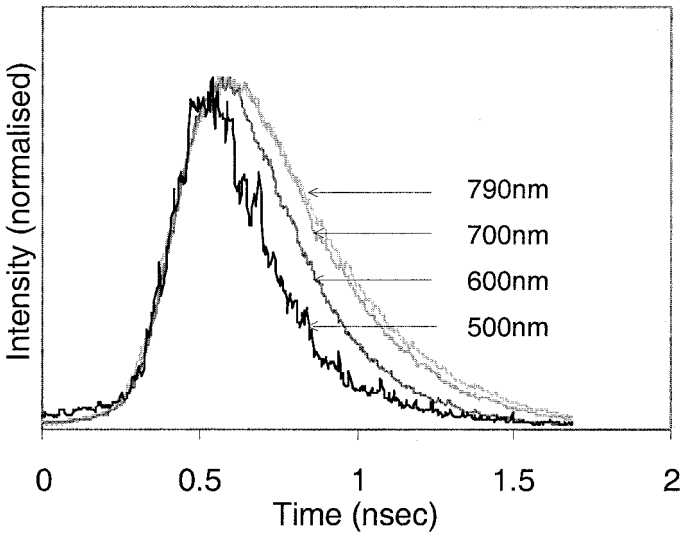


FIG. 5. Time-resolved transmission profiles of a tablet using ultra-short white light pulses. The profiles were extracted from the raw data matrix at 500, 600, 700, and 790 nm, respectively. The curves are autoscaled to the same maximum intensity.

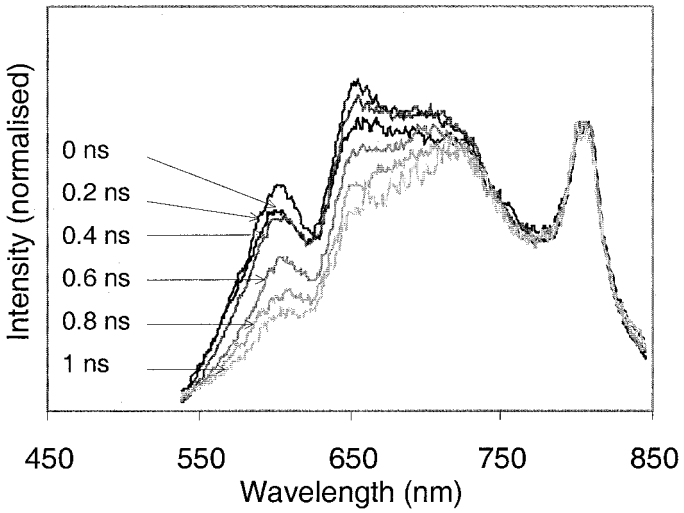


FIG. 6. Time-resolved transmission spectra; data extracted from same data matrix as the data in Fig. 5. The extracted spectra correspond to gating times indicated in the figure. Note that all spectra are uncorrected for the spectral instrument function. Each set of spectra is normalized at local maximum intensity around 800 nm.

tal data in order to estimate the optical properties of the sample. The flat cylinder used in the simulation had a diameter of 9 mm and a thickness of 3 mm. The best fit between the experimental data and the simulated was found when μ_s' was set to 500 cm^{-1} and μ_a was set to 0.01 cm^{-1}.

Comparing the three scans, it is noted that the light exiting from the margin of the tablet is shifted in time in relation to the centrally transmitted light. It can also be observed that the marginal profile is broadened compared to the center profile. With Fig. 3 in mind, these observations indicate that the light exiting further away from the center exhibit more light scattering than the central rays. This is consistent with a longer average path length for light recorded close to the margin as compared to that detected in the center.

Time-resolved data were collected with the spectrometer attached to the streak camera, enabling spectral in addition to temporal resolution. Thus, the data matrix here has a spectral rather than a lateral dimension. In these experiments, light from the center portion of the tablet backside was collected. In Fig. 5, a number of time profiles extracted from the data matrix at various wavelengths are shown. Note that the experiment was performed with a slightly modified optical set-up and the time scale is shifted compared to that in Fig. 4. As can be observed in Fig. 5, the time profile is more narrow for shorter wavelengths. In addition, the peak appears at slightly shorter times at shorter wavelengths. This suggests that the light transmission is more dominated by absorption at shorter wavelengths than at longer wavelengths. The effect of light scattering appears to be less significant since the peak intensity only shows a small shift upon the changing wavelength. In Fig. 6, the same data is displayed as time-gated spectra, i.e., transmission spectra for a number of different photon propagation times. The spectra are normalized to the same intensity to better illustrate the spectral differences. Comparing the spectra at different propagation times, it is again shown that at shorter wavelengths the light is attenuated more rapidly. The spectral features in Fig. 6 are not of particular importance in this study and originate not only from the sample but also from the detector response function and the spectrum of the white light source (including, e.g., inverse Raman effects[23]).

Photon Migration in Tablets. A series of measurements were performed using tablets with three different contents of the active substance. The experimental data was evaluated as shown in Fig. 7. In the lower graph (**e**), an example of a time profile is shown. In the upper four graphs (**a**–**d**), time-gated spectra are shown, one graph for each of four gate times analyzed. These gate times are approximately indicated in the time profile in the lower graph. In each of the four upper graphs, spectra from each of the three tablets are shown. In prompt light (**a**), light that has passed the tablet in a straight forward manner, very little difference between the curves can be observed: the three spectra have virtually the same shape (spectra autoscaled). However, for the spectra obtained at 0.25 ns photon propagation time (**b**), a substantial contrast between the 43-mg and the 52-mg tablets is observed. The behavior is similar at longer times (**c**, **d**), however, with a gradual loss of contrast as the gate time is increased. The important conclusion from this experiment is that the spectral contrast is different for different photon propagation times, and there appears to exist an optimal gating time where the highest contrast between tablets of different strengths is attained. Thus, by separating the spectral information in time, a higher analytical selectivity can be reached compared with the steady-state case of traditional transmission (NIR) spectroscopy.

It is interesting to compare the spectra in Fig. 7 in view of the effects shown in Fig. 3. As was discussed above, the absorption properties of the sample do not affect the early transmitted light, but the importance of absorption increases at longer times. For the tablet spectra in Fig. 7, it is observed that the most reliable spectral differences are found at medium propagation times. If the spectral

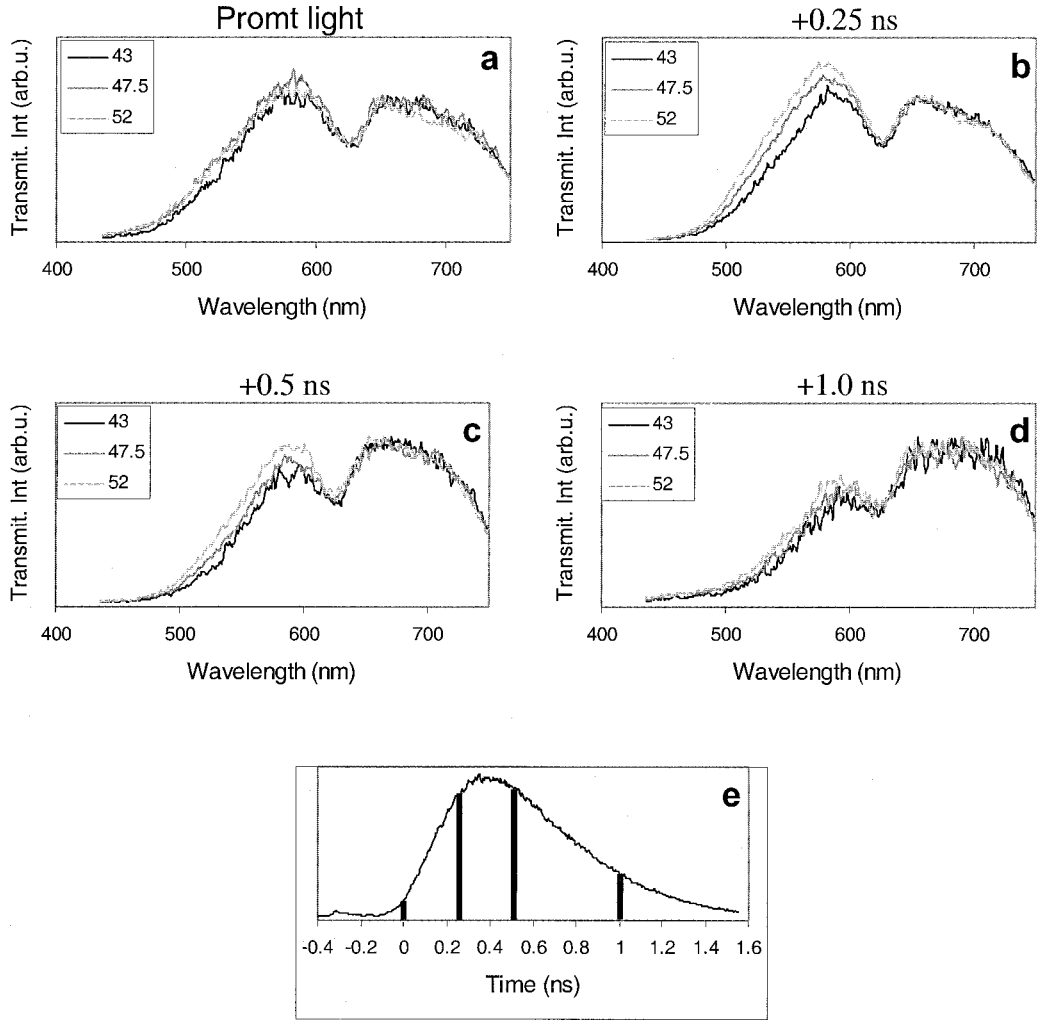


FIG. 7. Time-gated transmission spectra (**a**–**d**) at gate times indicated in the time profile (**e**). In each graph spectra from 3 tablets of different tablet strengths are shown. Each set of spectra is normalized at local maximum intensity around 650 nm.

differences were due only to absorption, Fig. 3 predicts that the contrast would increase with increased gate time, which is obviously not the case here. This suggests that not only the absorption but to a great extent also the scattering is important for the analysis of these samples.

Implications for NIR Spectroscopy. The main advantage of time-resolved spectroscopy is that the spectral contributions from both absorption and scattering effects can be experimentally determined. An attractive possibility should be to develop analytical methods for tablets of different shapes, thickness, and granular structure, based solely on their absorption characteristics, something that is impossible with today's NIR instruments. Also, calibration transfer between different systems should be simplified with a technique capable of measuring pure absorption features.

Future developments of time-resolved spectroscopy for quantitative analysis of solids, such as tablets, powders, etc., have to include both instrumental improvements to push the technique further into the NIR spectral region and development of new data evaluation techniques. The white light generation can be made much more compact and can be further optimized by harder focusing of the laser light and by using other nonlinear media, such as sapphire or photonic bandgap fibers. The collection optics can be improved by using optics coated for the desired wavelength region and in particular a more efficient spectrometer. The data collection can also be improved by increasing the repetition rate of the laser up to the kHz region for better data averaging. The main challenge in extending further into the NIR region is the lack of efficient photo cathode materials available for streak tubes; the quantum yield of the S1 material drops two orders of magnitude going from 800 to 1200 nm. Measurements in the 800–1200 nm spectral region were performed in addition to the experiments reported above. Results comparable to the above data matrices were obtained, however, with poor signal-to-noise ratios, preventing any meaningful evaluation.

An alternative approach is to use a different technique. Phase modulation spectroscopy is an attractive alternative technique in which the phase of the measured light is compared with the phase of the light source, a simple NIR-emitting lamp. From the phase difference, the photon propagation times can be calculated to yield the same information as with time-resolved techniques. The major advantage of this technique is the simplicity and low cost of the optical set-up. A possible disadvantage is the lim-

ited time resolution. Spatially resolved methods may also be an alternative to time-resolved techniques. The benefit of spatially resolved techniques is a much simplified experimental set-up; however, a homogeneous sample is required. Future work will also include improvements of the data evaluation. Three-way analysis such as the PARAFAC method[24] appears to be a useful alternative to Monte Carlo calculations. A more extensive study of simulation of photon propagation through tablets is currently being performed.

CONCLUSION

Time-resolved transmission measurements were performed on pharmaceutical tablets in a feasibility study to investigate the usefulness of the general technique for analytical work. The results show that the light scattering of the samples is very high, leading to mean optical path lengths of 20–25 cm. Monte Carlo simulations suggest a reduced scattering coefficient in the range of 500 cm^{-1} for a typical tablet. The spectral characteristics were found to depend on the delay of the detector gate time. For light with a long propagation time through the sample matrix, the spectrum was found to peak at higher wavelengths in relation to the early light. This was due to a greater attenuation of the blue/green light. Gated transmission spectra were recorded for tablets with different contents of the active substance, and it was found that the contrast between tablets was higher for some photon propagation times than for others. In particular, for these samples short times yielded no contrast, whereas medium times exhibit high spectral contrast. Thus, by gating the light transmission in time, a higher analytical contrast can be achieved than with conventional transmission spectroscopy. The technique seems promising and future work will be focused on extending the spectral range further into the NIR spectral region.

ACKNOWLEDGMENTS

The support by the Knut and Alice Wallenberg Foundation and the Swedish Council for Engineering Sciences, as well as the assistance by Dr. Anders Persson and Dr. Claes af Klinteberg, is gratefully acknowledged.

1. K. H. Norris, J. Near Infrared Spectrosc. **4,** 31 (1996).
2. *Handbook of Near-Infrared Analysis,* D. A. Burns and E. W. Ciurczak, Eds. (Marcel Dekker, New York, 2001), 2nd ed.
3. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, and C. de la Pezuela, Analyst (Cambridge, U.K.) **123,** 135R (1998).
4. M. A. Dempster, B. F. MacDonald, P. J. Gemperline, and N. R. Boyer, Anal. Chim. Acta **310,** 43 (1995).
5. P. R. Kahn, R. D. Jee, R. A. Watt, and A. C. Moffat, Pharm. Sci. **3,** 447 (1997).
6. P. K. Aldridge, R. F. Mushinsky, M. M. Andino, and C. L. Evans, Appl. Spectrosc. **48,** 1272 (1994).
7. J. Workman, Jr., D. J. Veltkamp, S. Doherty, B. B. Anderson, K. E. Creasy, M. Koch, J. F. Tatera, A. L. Robinson, L. Bond, L. W. Burgess, G. N. Bokerman, A. H. Ullman, G. P. Darsey, F. Mozayeni, J. A. Bamberger, and M. Stautberg Greenwood, Anal. Chem. **71,** 121 (1999).
8. W. F. McClure, Anal. Chem. **66,** 43A (1994).
9. A. Eustaquio, P. Graham, R. D. Jee, A. C. Moffat, and A. D. Trafford, Analyst (Cambridge, U.K.) **123,** 2303 (1998).
10. A. Eustaquio, M. Blanco, R. D. Jee, and A. C. Moffat, Anal. Chim. Acta **383,** 283 (1999).
11. P. Corti, G. Ceramelli, E. Dreassi, and S. Mattii, Analyst (Cambridge, U.K.) **124,** 755 (1999).
12. J. Gottfries, H. Depui, M. Fransson, M. Jongeneelen, M. Josefson, F. W. Langkilde, and D. T. Witte, J. Pharm. Biomed. Anal. **14,** 1495 (1996).
13. *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis,* V. Tuchin, Ed. (SPIE Press, Bellingham, 2000).
14. T. L. Troy, D. L. Page, and E. M. Sevick-Muraca, J. Biomed. Opt. **1,** 342 (1996).
15. F. Bevilacqua, D. Piguet, P. Marquet, J. D. Gross, B. J. Tromberg, and C. Depeursinge, Appl. Opt. **38,** 4939 (1999).
16. T. J. Farrell, M. S. Patterson, and B. Wilson, Med. Phys. **19,** 879 (1992).
17. J. Johansson, R. Berg, A. Pifferi, S. Svanberg, and L. O. Björn, Photochem. Photobiol. **69,** 242 (1999).
18. R. R. Shinde, G. V. Balgi, S. L. Nail, and E. M. Sevick-Muraca, J. Pharm. Sci. **88,** 959 (1999).
19. O. Berntsson, T. Burger, S. Folestad, L.-G. Danielsson, J. Kuhn, and J. Fricke, Anal. Chem. **71,** 617 (1999).
20. S. Svanberg, J. Larsson, A. Persson, and C.-G. Wahlström, Physica Scripta **49,** 187 (1994).
21. S. Andersson-Engels, R. Berg, and S. Svanberg, J. Photochem. Photobiol. **16,** 155 (1992).
22. L. Wang, S. L. Jacques, and L. Zheng, Computer Method. Prog. Biomed. **54,** 141 (1995).
23. W. J. Jones and B. P. Stoicheff, Phys. Rev. Lett. **13,** 657 (1964).
24. R. Bro, Chemom. Intel. Lab. Syst. **38,** 149 (1997).

Anal. Chem. **2005,** *77,* 1055−1059

Time-Resolved NIR Spectroscopy for Quantitative Analysis of Intact Pharmaceutical Tablets

Christoffer Abrahamsson,*,† Jonas Johansson,‡ Stefan Andersson-Engels,† Sune Svanberg,† and Staffan Folestad‡

Department of Physics, Lund Institute of Technology, P.O. Box 118, SE-221 00 Lund, Sweden, and AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden

Near-infrared (NIR) spectroscopy is a useful technique for quantitative measurements of intact tablets, but it suffers from limitations due to the fact that changes in the physical properties of a sample strongly affect the recorded spectrum. In this work, time-resolved transmission NIR spectroscopy was utilized to conduct quantitative measurements of intact tablets. The technique enables separation of the absorption properties of the sample from the scattering properties and can therefore handle changes of the physical parameters of the samples in a better way than conventional NIR transmission spectroscopy. The experiments were conducted using a pulsed Ti:sapphire laser coupled into a nonlinear photonic crystal fiber as light source. The light transmitted through the sample was measured by a time-resolving streak camera. A comparison of the results from the time-resolved technique with the results from conventional transmission NIR spectroscopy was made using tablets containing different concentrations of iron oxide and manufactured with different thicknesses. A PLS model made with data from the time-resolved technique predicted samples 5 times better than a PLS model made data from the conventional NIR transmission technique. Furthermore, an improvement to predict samples with physical properties outside those included in the calibration set was demonstrated.

In recent years, the application of transmission near-infrared (NIR) measurements of intact tablets has emerged as a useful technique for quantitative measurements.[1,2] It is faster than the traditionally used chromatographic methods, and it does not have the problems with subsampling that the reflectance NIR measurements have.[3] The ability to make transmission measurements of intact tablets relies on the relatively low absorption in this wavelength region. The main absorption bands in the NIR region are vibrational overtones and combination bands of hydrogen bonds, having their fundamental absorptions in the mid-infrared region. The transition probability of the overtones is much lower than that for the fundamental transitions, which makes the measured absorbance in a solid pharmaceutical sample in this region low, especially in the wavelength range 800−1400 nm, where the water absorption is low. Multiple wavelengths together with multivariate analysis are regularly employed in order to make quantitative calibrations, since the broad and often overlapping absorption bands often cannot be coupled to one specific component in the tablet. Previously, multiple linear regression was the preferred choice, but the development of new multivariate tools has made the partial least squares (PLS) the most frequently used calibration method for NIR spectroscopy today.[4] In addition to NIR spectroscopy, other thechniques, such as Raman spectroscopy, have proven useful for pharmaceutical analysis. However, Raman spectroscopy also suffers from limitations, in particular the small sampling volume associated with Raman spectroscopy of solids.[5,6]

Light scattering in intact tablets is due to the many microcavities causing rapid spatial changes in refractive index within the tablet. The scattering properties of a tablet are thus very dependent on the manufacturing process, e.g., on the compression force, grain size, etc. The scattering in a tablet is ∼3−4 orders of magnitude larger than the absorption, which leads to very long optical path lengths. For example, a transmission measurement through a 3.5-mm-thick tablet results in a mean optical path length of 20−30 cm.[7] There is, however, a huge span in effective path lengths in such a measurement. Some photons have path lengths as short as a few millimeters, while others have bounced back and forth within the tablet for as long as a meter before they can escape and reach the detector. From this it can be understood that variations in tablet density, radius, and thickness also affect the measurement and that small changes in the physical parameters during the manufacturing process can affect the NIR transmission signal more than the concentration variations to be measured.

Several data pretreatment methods have been developed to try to compensate for the scattering in the tablets, and in that way improve the quantitative calibrations. The standard normal

* To whom correspondence should be addressed. Phone: +46-46-222 31 20. Fax: +46-46-222 31 77. E-mail: christoffer.abrahamsson@fysik.lth.se.

† Lund Institute of Technology.

‡ AstraZeneca R&D Mölndal.

(1) Eustaquio, A.; Graham, P.; Jee, R. D.; Moffatt, A. C.; Trafford, A. D. *Analyst* **1998**, *123*, 2303−2306.

(2) Dyrby, M.; Engelsen, S. B.; Norgaard, L.; Bruhn, M.; Lundsberg-Nielsen, L. *Appl. Spectrosc.* **2002**, *56*, 579−585.

(3) Gottfries, J.; Depui, H.; Fransson, M.; Jongeneelen, M.; Josefson, M.; Langkilde, F. W.; Witte, D. T. *J. Pharm. Biomed. Anal.* **1996**, *14*, 1495−1503.

(4) Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; de la Pezuela, C. *Analyst* **1998**, *123*, 135R−150R.

(5) Mulvaney, S. P.; Keating, C. D. *Anal. Chem.* **2000**, *72*, 145R−157R.

(6) Pelletier, M. J. *Appl. Spectrosc.* **2003**, *57*, 20A−42A.

10.1021/ac0487754 CCC: $30.25
Published on Web 01/12/2005

deviate,[8] multiplicative scattering correction,[9] and orthogonal signal correction[10] are all examples of pretreatment methods used to enhance the ability to construct accurate PLS calibrations. Despite the many efforts, it has still been proven hard to incorporate tablets from different batches or tablets manufactured under different conditions into the same quantitative calibration model with a good result.

Similar problems, to accurately measure the absorption if the scattering is not well known, are faced in another field, biomedical optics. In that field, both diagnostic measurements and laser treatment dosimetry depend critically on the optical properties of tissue. It is thus important to be able to measure those. Several methods to measure these properties have been developed, including the integrating sphere,[11] spatially resolved,[12] and time-resolved measurements.[13]

Techniques to measure the optical properties developed primarly for medical applications have been used in some pharmaceutical applications. Scattering and absorption properties have been measured in order to calculate the effective sample size in diffuse reflectance NIR spectroscopy of powders[14,15] as well as for particle size analysis.[16] Measurements of the optical properties have also been used to make quantitative measurements of pharmaceutical powder blend homogeneity.[17]

When using time-resolved measurements, the optical properties of a sample can be assessed by analyzing the broadening of a short light pulse after the passage through the sample. Several methods and models have been developed to extract the optical measurements from such measurements at discrete wavelengths.[18,19] These models show that a change in the scattering coefficient mainly alters the peak position, while a change in the absorption coefficient will alter the final slope of the time dispersion curve.

In our previous work, we investigated the basic capabilities of time-resolved NIR spectroscopy as an analytical tool for spectroscopic analysis.[7] In this work, we report on measurements performed using a novel instrument for time-resolved broadband NIR spectroscopy. The results demonstrate the capability to separate absorption from scattering properties of pharmaceutical tablets using time-resolved spectroscopy and thereby allow recordings of pure absorption spectra. Improved quantitative assessments are thus possible and also for samples with scattering properties different from those covered by the calibration samples.

Table 1. Summary of the Batch ID of the Tablets Used in This Work

	concentration iron oxide (%)			
weight (mg)	0,1	0,08	0,065	0,05
300	A1	B1	C1	D1
400	A2	B2	C2	D2
500	A3	B3	C3	D3

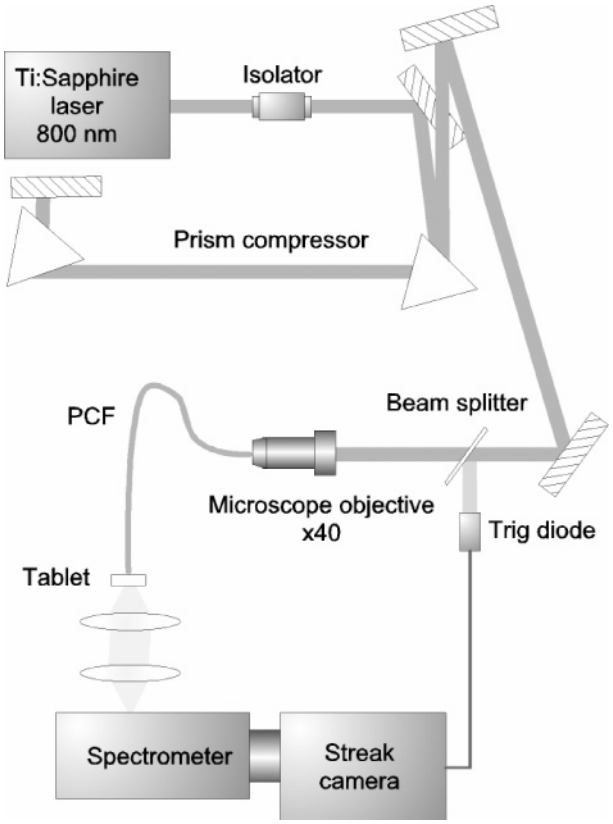


Figure 1. Overview of the instrumentation for time-resolved NIR measurements.

EXPERIMENTAL SECTION

Samples. The tablets used in this work consisted of red iron oxide and microcrystalline cellulose (MCC). The MCC and iron oxide were weighed and mixed in a mortar. The powder blend was pressed into tablets with a diameter of 9 mm using a manual tablet press. The compression force was held constant for all tablets. The tablet set consisted of four concentration levels (batches A−D) and three tablet thicknesses (batches 1−3). A total number of 27 tablets were produced with the characteristics given in Table 1. One sample from batch D2 was destroyed during the reference analysis and is therefore taken out of all evaluations.

Time-Resolved Measurements. The experimental arrangement used in the study is depicted in Figure 1.

The Ar ion laser-pumped mode-locked Ti:sapphire laser produced pulses shorter than 100 fs at a repetition rate of 80 MHz. The wavelength of the laser light was near 800 nm, and the energy of each pulse was 4 nJ. The light was focused into the PCF using

(7) Johansson, J.; Folestad, S.; Josefson, M.; Sparen, A.; Abrahamsson, C.; Andersson-Engels, S.; Svanberg, S. *Appl. Spectrosc.* **2002**, *56*, 725−731.
(8) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. *Appl. Spectrosc.* **1989**, *43*, 772−777.
(9) Geladi, P.; MacDougall, D.; Martens, H. *Appl. Spectrosc.* **1985**, *39*, 491−500.
(10) Wold, S.; Antii, H.; Lindgren, F.; Ohman, J. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175−185.
(11) Pickering, J. W.; Prahl, S. A.; van Wieringen, N.; Beek, J. F.; Sterenborg, H. J. C. M.; van Gemert, M. J. C. *Appl. Opt.* **1993**, *32*, 399−410.
(12) Doornbos, R. M. P.; Lang, R.; Aalders, M. C.; Cross, F. W.; Sterenborg, H. J. C. M. *Phys. Med. Biol.* **1999**, *44*, 967−981.
(13) Andersson-Engels, S.; Berg, R.; Jarlman, O.; Svanberg, S. *Opt. Lett.* **1990**, *15*, 1179−1181.
(14) Berntsson, O.; Burger, T.; Folestad, S.; Danielsson, L. G.; Kuhn, J.; Fricke, J. *Anal. Chem.* **1999**, *71*, 617−623.
(15) Pan, T.; Sevick-Muraca, E. M. *Anal. Chem.* **2002**, *74*, 4228−4234.
(16) Sun, Z.; Torrance, S.; McNeil-Watson, F. K.; Sevick-Muraca, E. M. *Anal. Chem.* **2003**, *75*, 1720−1725.
(17) Shinde, R. R.; Balgi, G. V.; Nail, S. L.; Sevick-Muraca, E. M. *J. Pharm. Sci.* **1999**, *88*, 959−966.
(18) Patterson, M. S.; Chance, B.; Wilson, B. C. *Appl. Opt.* **1989**, *28*, 2331−2336.
(19) Swartling, J.; Dam, J. S.; Andersson-Engels, S. *Appl. Opt.* **2003**, *42*, 4612−4620.

a standard ×40 microscope objective lens with a numeric aperture of 0.65. An optical isolator was used between the laser and the optics, to prevent optical feedback into the laser due to reflections. A prism compressor was also used in the setup to compensate for the time dispersion caused by the different optical components used. The PCF (Crystal Fibre A/S) was 0.5 m long with a core diameter of 2 μm, manufactured to have zero dispersion at 760 nm. The dispersion properties of the fiber combined with core diameter results in a high peak power of the light through the entire fiber, yielding a widely spectrally broadened light emission due to nonlinear effects. The main broadening effects in the PCF are identified to be self-phase modulation and stimulated Raman scattering.[20] As a result of this, light pulses with almost the same temporal width as the laser, and a spectral width spanning from 400 nm to at least 1200 nm was accessible. The light distribution was not flat though, but modulated with peaks with high intensities surrounded by wavelength regions with low intensities. To increase the signal-to-noise ratio, three measurements were made on each sample. The PCF was slightly realigned between each measurement in order to flatten the average light distribution in the range 860−1150 nm. This provides a good signal-to-noise ratio in the entire wavelength range. The output end of the PCF was put 2 mm from the face of the tablet held into place by an circular iris holder. The spot size on the tablet was ~2 mm. The light from the backside of the tablet was imaged onto the 250-μm slit of an imaging spectrometer (Chromex, model 250 IS), coupled to a streak camera (Hamamatsu, model C5680). The system measures a 600-nm broad wavelength region with a spectral resolution of 5 nm. The streak camera operated in synchro scan mode, allowing all light pulses to be collected. A small portion of the laser light was redirected by a beam splitter onto a photodiode that triggered the streak camera sweep. The system had a total temporal range of 2.1 ns with resolution of 4.5 ps. The instrumental response function was in the range of 30 ps when averaging over 50 s.

Conventional Transmission NIR Measurements. The conventional transmission NIR spectra were measured with a NIRSystems 6500 monochromator, with a NIRSystems InTact Multi-Tab Analyzer presenting the samples. The spectra were collected in the range of 600−1900 nm, but as in previous reports, the range 800-1350 nm was used when building calibration models.

Reference Analysis. For the reference analysis, the samples were heated in beakers with 40 mL of HNO_3 (70%) for 60 min and then transferred into 100-mL flasks and diluted to volume with water. Calibration samples were made using the same method, using concentrations between 1 and 7 ppm iron oxide. The samples were then analyzed using a Perkin-Elmer 3300 atomic absorption spectrometer. By analyzing reference samples, the relative error introduced by the method was measured to be smaller than 2%.

Evaluation of Time-Resolved Measurements. The evaluation of the data was based on the shape of the time dispersion curves, and it was made entirely independent for the different wavelengths. No white light correction was therefore necessary.

To combine the three measurements in the best possible way, optimizing the signal-to-noise ratio, a threshold procedure was employed. The signal level for an individual measurement had to exceed this threshold to be included in the analysis. The threshold was set so that all wavelengths from 860 to 1150 nm was covered in at least one of the three measurements. The wavelength range was chosen to incorporate important absorption features of both constituents. The final slopes of the time-resolved signals were calculated for each individual wavelength, one at the time. The time dispersion curve for each wavelength was first normalized to have intensities between zero and one and then logarithmic transformed. A line was fitted to a 210-ps-long region 105 ps from the peak; see Figure 2. The slopes calculated at the different wavelengths were extracted, plotted as a function of wavelength, and used in the analysis.

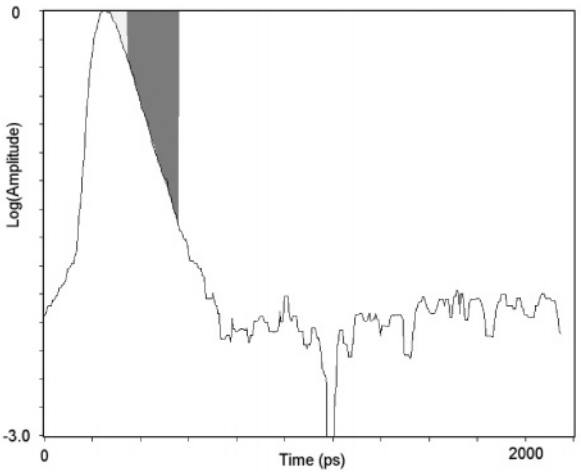


Figure 2. Evaluation of the time-resolved measurements was made by fitting a line to the final slope of the time dispersion curve for each individual wavelength. The fit was made in the shaded region of the curve.

Multivariate Analysis. All multivariate models were made in Simca-P 10.0 (Umetrics AB, Umeå, Sweden). All spectra were mean centered before calculations, and the number of principal components selected in the models were as many as Simca-P 10.0 found suitable. The program uses the cross-validated predicted fraction for both X and Y to find the optimal number of PLSCs. The software found the optimal number of principal components to be three for all models in this work.

RESULTS AND DISCUSSION

Design of Technical Device. The introduction of a photonic crystal fiber and a synchroscan streak camera unit into the instrumentation extended the measurement capability further out into the NIR region. It is important to reach the second overtone region starting at 1100 nm in order collect good absorption data from pharmaceutical samples, and the measurements in this work reached wavelengths as high as 1150 nm with a sufficient signal-to-noise ratio. The instrumentation is a complex laboratory system that is not easily adapted to simple applications.

The simple, but effective evaluation scheme used in this work had the big advantage of evaluating an intrinsic property at each wavelength independently. The evaluation was therefore independent of both the spectral distribution of the light source and the intensity of the light and why no normalization or white light correction was needed. The evaluation scheme had the limitation of extracting only the absorption properties, but new and more

(20) Genty, G.; Lehtonen, M.; Ludvigsen, H.; Broeng, J.; Kaivola, M. *Opt. Express* **2002**, *10*, 1083−1098.

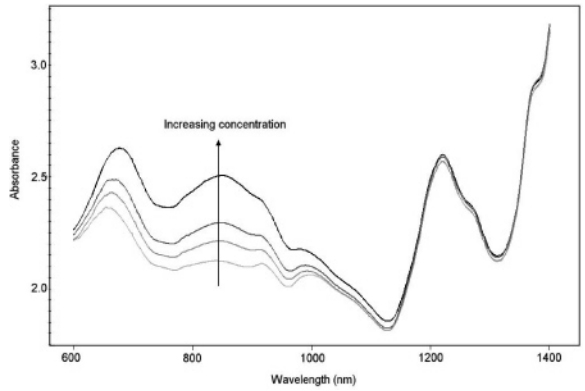


Figure 3. Conventional transmission NIR data from measurements of tablets with the same thickness but with different concentrations of iron oxide. The tablets all weighed 400 mg and contained 0.05, 0.065, 0.08, and 0.1% iron oxide, respectively.

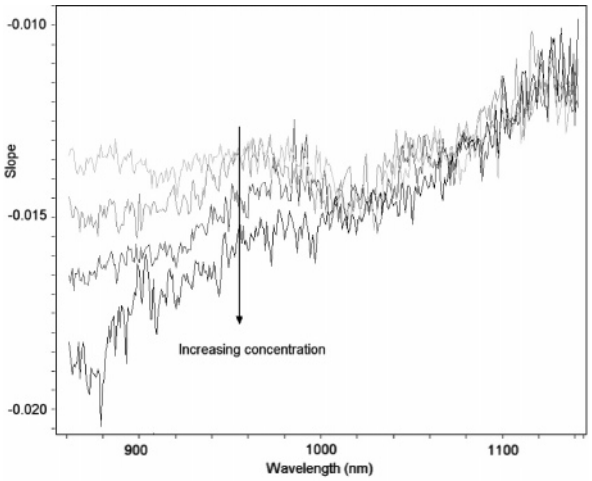


Figure 4. Decay slope versus wavelength from measurements of tablets with the same thickness, but with different concentrations of iron oxide. The tablets all weighed 400 mg and contained 0.05, 0.065, 0.08, and 0.1% iron oxide, respectively.

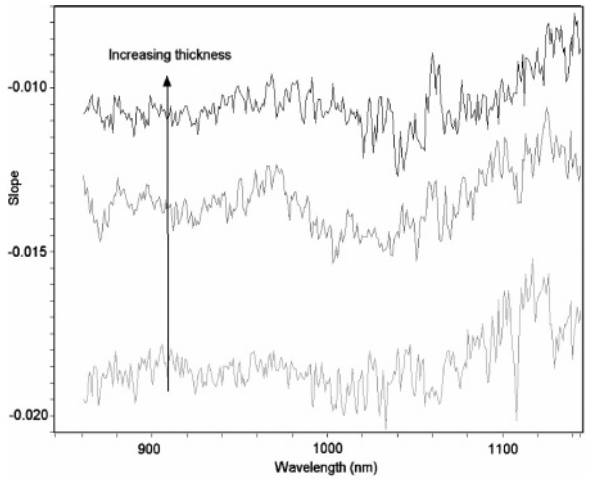


Figure 5. Decay slope versus wavelength from measurements of tablets with the same concentration of iron oxide, but with different thicknesses. The tablets contained 0.05% iron oxide and weighed 300, 400, and 500 mg, respectively.

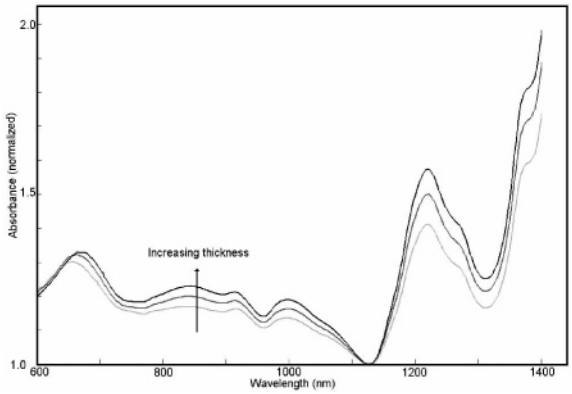


Figure 6. Conventional transmission NIR data from measurements of tablets with the same concentration of iron oxide but with different thicknesses. The tablets contained 0.05% iron oxide and weighed 300, 400, and 500 mg, respectively.

complex evaluation tools that allow the extraction of both the absorption and scattering properties of the sample are under development and will further enhance the technique.

Mechanistic Basis. Results from the conventional transmission NIR measurements are shown in Figure 3. The figure shows absorption spectra from four tablets with the same thickness but with different concentrations of iron oxide. The spectra exhibit a distinct absorption peak from MCC around 1200 nm. The main absorption band from iron oxide is prominent in the lowest wavelength region, below 1000 nm. The spectra from the different tablets exhibit different attenuation in this wavelength region, with the highest concentration of iron oxide having the highest attenuation. Figure 4 shows the extracted slope coefficients as a function of wavelength from time-resolved measurements of the same four tablets. The measurements were made in the wavelength region 860−1150 nm. The slope spectra are relatively noisy, partly due to the low signal in the time-resolved measurement, but also due to the fact that these spectra are evaluated from the slope of a curve rather than a plot of the signal level itself. This type of analysis is very sensitive to the noise in the original signal. When the absorption of a sample is high, the final slope of the time dispersion curve is steeper than the final slope of the time dispersion curve from a sample with low absorption. This is clearly seen in the figure as a steeper slope in the wavelength region below 1000 nm for tablets with higher concentration of iron oxide. Comparing the time-resolved slope coefficients with the conventional transmission NIR measurement yields clear similarities. The slope spectra are noisier than the conventional NIR spectra, but both show differences between the tablets in the region 850−1000 nm where the iron oxide absorbs the light.

Figure 5 shows the slopes as a function of wavelength from measurements of tablets with the same nominal iron oxide concentration but with different thicknesses. The slope spectra are noisy, and the change in thickness between the tablets introduces an offset in the slope values. Unlike conventional transmission NIR measurements on tablets with different thicknesses, see Figure 6, where the longer path length of the light will introduce not only an offset of the absorption profile but also an increased contrast, the differences in shape between the slope spectra are smaller than the noise level. This is because the slope in the slope spectra is proportional to the path length-independent absorptivity, while the conventional NIR spectra are proportional to the total attenuation, which is strongly dependent on the optical path length.

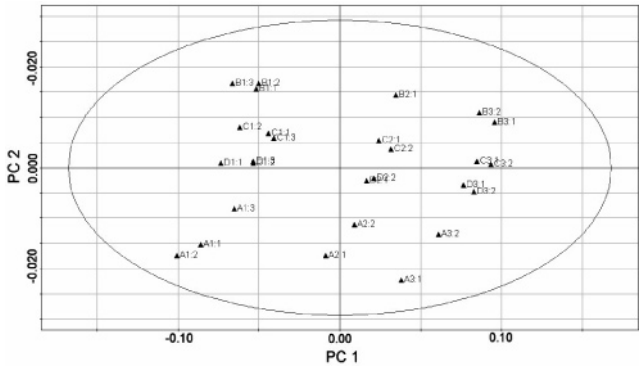


Figure 7. Score plot of a PCA model made from all 26 samples using the slopes calculated from the time-resolved measurements. The scores are from the first two principal components of the PCA model.

Assay of Chemical Content. A PCA model was made from all 26 samples using the slopes calculated from the time-resolved measurements. The wavelength range 850−1160 nm was used in the calculations. A score plot of the first two components was constructed to display the main contribution to the spectral information, see Figure 7. It was found that the samples are spread in two almost orthogonal directions. Principal component 1 divides the samples into the three different thickness groups, while principal component 2 spread the samples according to their concentration of iron oxide.

A PLS model was constructed by using 12 of the samples. This PLS model was used to predict the iron oxide concentration of the other samples. By doing this for the two different techniques, a quantitative comparison was done. The 12 calibration samples included samples from all thickness groups and all concentration levels.

The models were based on three PLS components and the RMSEP values for the conventional NIR measurements was 0.0080% (m/m) iron oxide, compared to 0.0019% (m/m) iron oxide for the time-resolved measurements. The relative prediction error in the time-resolved calibration was on the order of 2.5%, compared to 12% for the conventional NIR calibration. The latter number may seem high for a NIR method, but it is explained by the fact that the calibration model contains only 12 samples that span a wide range of thicknesses and concentrations of iron oxide. To achieve a good predictive ability of a PLS model based on conventional NIR spectra of samples with such divergent physical properties, far more than 12 samples are needed.

To compare the robustness of the two measurement techniques, calibration models were build by using all samples from two of the thickness groups and using the samples from the third thickness group as a prediction set.

The RMSEP values were in this case 0.0026% (m/m) iron oxide for the time-resolved data and 0.011% (m/m) iron oxide for the conventional NIR data. The relative error for the time-resolved data was in this case 3.5% compared to 15% for the conventional NIR data. Again, the time-resolved method is superior to the conventional NIR method and allows the analysis to be extrapolated outside the calibration range of the PLS model.

The fact that the PCA model yields two orthogonal components and that PLS calibrations are linear over a wide range and are even valid outside the calibration range support the conclusion that pure chemical information can be attained with time-resolved spectroscopy. By using the slope of the time-resolved curves, information attributed directly to light absorption and with no relation to light scattering can be attained. So far the analytical precision is limited by the determination of the slope. There should be a significant potential for improvements by using other data evaluation schemes.

CONCLUSIONS

In this work, we have used a newly developed instrumentation for time-resolved transmission broadband NIR spectroscopy to separate the absorption from the scattering of the samples. The time-resolved measurements were used to make quantitative assessments of intact tablets with different thicknesses, and the results were then compared with conventional transmission NIR spectroscopy.

The comparison shows that the time-resolved technique is better at handling physical variations of the tablets, in this case different thicknesses. Although the thickest tablets were more than 50% thicker than the thinnest, the relative prediction error was only 2.5% in a PLS model incorporating slope spectra from tablets with different thicknesses, compared to 12% for a PLS model based on the conventional transmission NIR technique. The slope spectra also showed better potential than the conventional NIR technique when predicting samples from a thickness group not included in the PLS calibration.

The time-resolved technique seems promising, for analysis of solid samples with varying physical properties, which would make the samples difficult to analyze with conventional NIR techniques. The future work will be focused on increasing the signal-to-noise ratio in the measurements and enabling measurements on real pharmaceutical tablets. Other evaluation schemes will also be developed, to get a better understanding of the interaction between the light and the sample.

Received for review August 17, 2004. Accepted November 22, 2004.

AC0487754

Scatter Correction of Transmission Near-Infrared Spectra by Photon Migration Data: Quantitative Analysis of Solids

CHRISTOFFER ÅBRAHAMSSON,* ALEXANDRA LÖWGREN, BIRGITTA STRÖMDAHL, TOMAS SVENSSON, STEFAN ANDERSSON-ENGELS, JONAS JOHANSSON, and STAFFAN FOLESTAD

Department of Physics, Lund Institute of Technology, P.O. Box 118, SE-221 00 Lund, Sweden (C.A., A.L., B.S., T.S., S.A.-E.); and AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden (J.J., S.F.)

The scope of this work is a new methodology to correct conventional near-infrared (NIR) data for scattering effects. The technique aims at measuring the absorption coefficient of the samples rather than the total attenuation measured in conventional NIR spectroscopy. The main advantage of this is that the absorption coefficient is independent of the path length of the light inside the sample and therefore independent of the scattering effects. The method is based on time-resolved spectroscopy and modeling of light transport by diffusion theory. This provides an independent measure of the scattering properties of the samples and therefore of the path length of light. This yields a clear advantage over other preprocessing techniques, where scattering effects are estimated and corrected for by using the shape of the measured spectrum only. Partial least squares (PLS) calibration models show that, by using the proposed evaluation scheme, the predictive ability is improved by 50% as compared to a model based on conventional NIR data alone. The method also makes it possible to predict the concentration of active substance in samples with other physical properties than the samples included in the calibration model.

Index Headings: **Scatter correction; Near-infrared spectroscopy; NIR spectroscopy; Partial least squares; PLS; Photon migration; Time-resolved spectroscopy; Diffusion.**

INTRODUCTION

Near-infrared (NIR) spectroscopy is an important tool for assessment of the chemical content of solid samples due to the fact that the samples can be analyzed directly in their native solid state. NIR spectroscopic measurements can be conducted both in transmission[1–4] and reflectance[5,6] mode, and the development of fiber optical probes[7–10] has enabled measurements directly in the reaction vessels, e.g., in a pharmaceutical process line.

Although the versatility and speed of NIR spectroscopic measurements has made it an important tool in process analytical chemistry, the technique has some limitations. One of the major drawbacks of NIR spectroscopy is its sensitivity to variations of the physical characteristics of the samples.[3,11] This is due to the fact that the measured absorbance follows the Beer–Lambert law and is therefore dependent on the concentration of the constituent to be quantified, but also on the path length of the light passing through the sample. The path length of the light passage between the light source and the detector is dependent on the physical parameters of the samples, e.g., sample thickness, particle size distribution, and sample compactness. In fact, when measuring on an intact tablet in the NIR wavelength range, the scattering is about 1000 times more prominent than the absorption,[12] which means that a small change in the physical parameters of the samples can alter the measured spectra to a larger extent than the alterations introduced by the variation in concentration of the sample constituents.

Several mathematical spectral pretreatment methods, e.g., standard normal deviation,[13] multiplicative scatter correction,[14] and orthogonal signal correction,[15] have all been proposed to correct NIR spectra in order to eliminate systematic variations unrelated to analyte concentrations. Despite the many efforts, it has still proven hard to incorporate samples from different batches or samples manufactured under different conditions into the same quantitative calibration model with acceptable results.

An alternative to mathematical pretreatment methods is to use a direct measurement of the scattering properties of the samples to correct conventional NIR spectra. Different measurement techniques have been developed to deconvolute the scattering and absorption properties of a sample. These techniques include time-resolved,[16] spatially resolved,[17] and integrating sphere measurements.[18] The techniques to measure the optical properties were developed primarily for biomedical applications but have also been used in some pharmaceutical applications. Scattering and absorption properties have been measured in order to calculate the effective sample size in diffuse reflectance NIR spectroscopy of powders[19,20] as well as for particle size analysis.[21] Measurements of the optical properties have also been used to make quantitative measurements of pharmaceutical powder blend homogeneity.[22]

When conducting time-resolved measurements a temporally very short light pulse is sent through the sample to be analyzed. The temporal shape of the pulse is altered when passing through the sample due to the dispersion of the light inside the sample. By analyzing the modified temporal shape of the pulse, the optical properties of that sample can be deduced.[23] A variety of different evaluation schemes have been developed for evaluating time-resolved data, ranging from simple evaluations like the final slope fitting[24] to more complex schemes like diffusion[25] and Monte Carlo models.[26]

The aim of this work was to introduce a methodology to improve quantitative assessments made from conventional NIR transmission data by using a scatter correction scheme based on the measurements of the actual scattering properties of the samples. To measure the scattering properties of the tablets in this work a novel broad-band time-resolved system was used in combination with dif-

Received 13 July 2005; accepted 24 August 2005.

* Author to whom correspondence should be sent. E-mail: christoffer.abrahamsson@fysik.lth.se.

0003-7028/05/5911-1381$2.00/0

fusion modeling of light transport. The results demonstrate the capability to deconvolute the absorption and scattering properties of pharmaceutical tablets using time-resolved spectroscopy. The quality of the quantitative assessments after the scatter correction was greatly improved, compared to assessments made directly from conventional NIR data. The improvements were especially large for samples with physical properties different from those covered by the calibration samples. The work also points out one possible direction for the development of NIR spectrometers, aiming at a system consisting of a standard NIR spectrometer in combination with a time-resolved diode-laser-based system at a few discrete wavelengths. Such a system would enable measurements of the absorption of samples without any contribution from scattering effects.

THEORY

Optical Properties of Turbid Media. The interaction between light and a turbid medium is governed by the optical properties of that medium. In this work the light will be assumed to be diffusely scattered and light transport will be modeled by the diffusion approximation.

The optical properties can be divided into absorption, primarily a measure of the chemical content of the sample, and scattering, dependent on the physical characteristics of the sample. The parameter used to describe the absorption of light is the absorption coefficient, μ_a, which is defined as the probability for absorption per unit length. The scattering of light is caused by variations of refractive index within the sample and is in the diffusion approximation described by the reduced scattering coefficient, μ_s', which is defined as the probability for an isotropic scattering event per unit length.

Diffusions Models. The measured time-resolved dispersion curves were analyzed using a solution of the radiative transport equation, under the diffusion approximation, for a semi-infinite slab.[27] The solution is based on the introduction of an isotropic point source in the sample at a distance z_0, equal to the inverse of μ_s' from the surface. This is applicable for many types of geometries as long as the solution is calculated for points far away from the source. Another restriction is that the reduced scattering coefficient must be much larger than the absorption coefficient for the diffusion approximation to be valid. Although single scattering events may not be isotropic, but rather be more prominent in specific directions, the validity of the diffusion model is dependent on the fact that the light is so vastly scattered that it loses its directionality and can be treated as isotropic. The diffusion approximation is valid when the distance between the light source and detector is larger than 10 times the mean free path of the photons in the sample,[28] which is greatly exceeded by the samples used in this work.

Since the refractive index changes at the surfaces of the slab, reflections will occur, and hence extrapolated boundaries, where the fluence rate equals zero, are introduced at a distance z_e from the real surface. Mirror sources are introduced around the extrapolated boundaries to fulfill the boundary condition.[29] In this study 30 mirror sources were used. At a time t and a radial distance r from the injection point, the transmittance through a slab is given by

$$T(r,t) = \frac{\exp\left(-\mu_a ct - \frac{r^2}{4Dct}\right)}{2(4\pi Dc)^{3/2}t^{5/2}} \times \sum_{m=-\infty}^{\infty}\left[z_{1,m}\exp\left(-\frac{z_{1,m}^2}{4Dct}\right) - z_{2,m}\exp\left(-\frac{z_{2,m}^2}{4Dct}\right)\right] \quad (1)$$

where

$$z_{1,m} = d(1-2m) - 4mz_e - z_0 \quad \text{for positive sources,}$$

$$z_{2,m} = d(1-2m) - (4m-2)z_e + z_0 \quad \text{for negative sources.}$$

where c is the speed of light, m is the number of the source, d is the thickness of the slab, and D is the diffusion coefficient given by

$$D = \frac{1}{3(\mu_a + \mu_s')} \quad (2)$$

An expression for steady-state transmission can be calculated by integrating the time-resolved expression over t, which gives

$$T(r) = \frac{1}{4\pi}\sum_{m=-\infty}^{\infty}\left(z_{1,m}(r^2+z_{1,m}^2)^{-3/2}\left\{1+\left[\frac{\mu_a(r^2+z_{1,m}^2)}{D}\right]^{1/2}\right\} \times \exp\left\{-\left[\frac{\mu_a(r^2+z_{1,m}^2)}{D}\right]^{1/2}\right\} - z_{2,m}(r^2+z_{2,m}^2)^{-3/2}\left\{1+\left[\frac{\mu_a(r^2+z_{2,m}^2)}{D}\right]^{1/2}\right\} \times \exp\left\{-\left[\frac{\mu_a(r^2+z_{2,m}^2)}{D}\right]^{1/2}\right\}\right) \quad (3)$$

EXPERIMENTAL

Samples. The tablets used in this work were produced in a cylindrical shape with flat end surfaces. The tablets had a diameter of 10 mm and thicknesses varied between 1.85 and 2.75 mm. All tablets had the same weight, and the thickness was varied by varying the compression force during the manufacturing process.

Three granulated materials with different concentration of active substance were used. The three granulated materials were sieved so that each tablet contained only particles of a certain size fraction. Two sieves were used, giving three different size fractions. The population investigated consisted of 82 tablets with approximately 9 tablets of each combination of particle size and concentration. The number of samples in the different batches is summarized in Table I. The different size fractions differed somewhat in concentration, but these differences were revealed by the reference analysis and therefore only made the concentration span of the samples larger.

As reference analysis, ultraviolet-absorption measurements were made on the tablets after they were dissolved

TABLE I. Overview of the number of measured tablets from the different batches.

		Sieve fraction (μm)		
		<150	150–400	>400
Nominal concentration (% weight)	28.5	9	10	9
	31.8	10	9	9
	34.9	9	8	9

in phosphate buffer pH 3.0.[30] The absorption was measured at 274 nm and the background at 550 nm on an HP 8453 UV/vis spectrometer (Agilent Technologies Sweden AB, Spånga, Sweden). From calibration samples, the content of active substance in the samples was calculated using the Beer–Lambert law. These values were used as reference values in the multivariate calibration models.

Time-Resolved Measurements. The time-resolved system used in this work has previously been described in detail.[31] Briefly, the experimental arrangement is depicted in Fig. 1. An Ar-ion laser pumped mode-locked Ti:Sapphire laser produced pulses shorter than 100 fs at a repetition rate of 80 MHz. The wavelength of the laser light was centered around 800 nm, and the energy of each pulse was 4 nJ. The light was focused into an index guiding crystal fiber (ICF) using a standard 40× microscope objective lens with a numeric aperture of 0.65. An optical isolator was used between the laser and the optics to prevent optical feedback into the laser due to reflections. A prism compressor was also used in the setup to compensate for the time dispersion caused by the different optical components. The ICF (Crystal Fibre A/S, Birkerod, Denmark) was 1 m long with a core diameter of 2 μm, manufactured to have zero dispersion at 760 nm. The dispersion properties of the fiber combined with the small core diameter resulted in a high peak power of the light through the entire fiber, yielding a widely spectrally broadened light emission due to nonlinear effects. The main broadening effects in the ICF were identified to be self-phase-modulation[32] and stimulated Raman scattering.[33] As a result of this, light pulses with almost the same temporal width as the laser, and with a spectral width spanning from 400 nm to at least 1200 nm, were accessible. However, the light distribution was not flat, but modulated with peaks with high intensities surrounded by wavelength regions with low intensities. The light from the output end of the ICF was focused by a lens onto the face of the tablet held into place by a circular iris holder, preventing stray light from reaching the detection system. The spot size on the tablet was approximately 2 mm. The light from the backside of the tablet was imaged onto the 250 μm slit of an imaging spectrometer, Chromex 250 IS (Bruker Optics Scandinavia AB, Taby, Sweden) coupled to a streak camera, Hamamatsu C5680 (Hamamatsu Photonics Norden AB, Solna, Sweden). The system measures a 600 nm broad wavelength region with a spectral resolution of 5 nm. The streak camera operated in synchro scan mode, allowing all light pulses to be collected. A small portion of the laser light was redirected by a beam splitter onto a photodiode that triggered the streak camera sweep. The system had a total temporal range of 2.1 ns with resolution of 4.5 ps. The instrumental response function was in the range of 30 ps when averaging over 300 s.

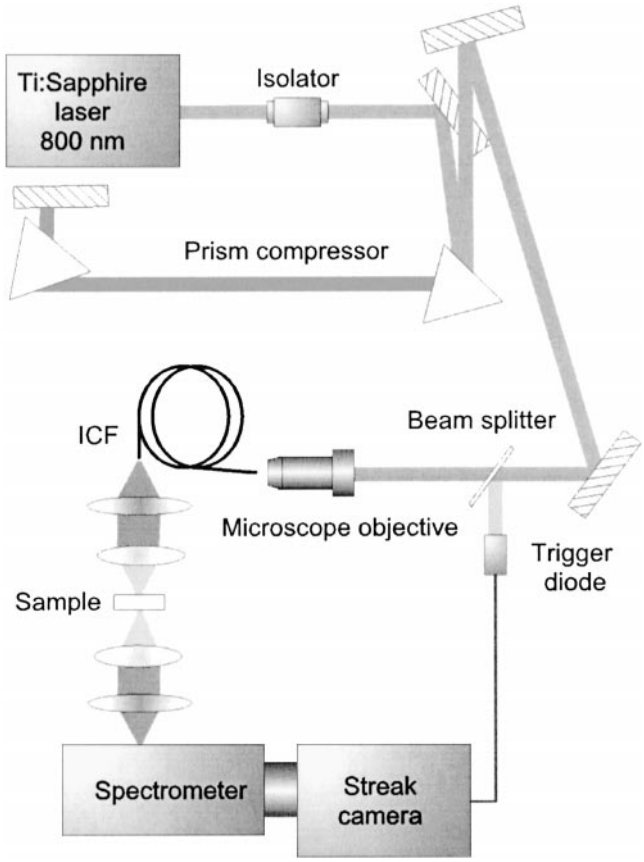


FIG. 1. Overview of the instrumentation used for the time-resolved measurements.

Conventional Transmission Near-Infrared Measurements. The conventional transmission NIR measurements were conducted on a Bomem MB 160 PH Fourier transform spectrometer (ABB Automation Technologies AB, Sollentuna, Sweden). The spectrometer was equipped with a tablet sampler making transmission measurements possible. The measurements were made in the wavelength range from 800 nm to 1500 nm with a resolution of 16 cm^{-1} in the entire range.

Deconvolution of Scattering and Absorption Properties of Samples Using Time-Resolved Measurements. The time-resolved data was evaluated for each wavelength individually in the wavelength region ranging from 800 to 1100 nm. The evaluation was made by fitting the measured time dispersion curves to the time-resolved diffusion model (Eq. 1), convolved with the instrumental response function (see Fig. 2). The dip at 275 ps in the photon migration data is due to detector sensitivity variations, but the effect is corrected before the evaluation. The data points included in the calculation were determined by two thresholds set to include all points with higher intensity than 20% of the peak intensity on the rising edge and higher than 10% on the falling edge. The evaluation algorithm used a Levenberg–Marquardt iterative procedure to extract μ_a and μ_s' from the data.

Scatter Correction of Conventional Transmission Near-Infrared Data. An overview of the complete scatter correction scheme is depicted in Fig. 3. The scattering

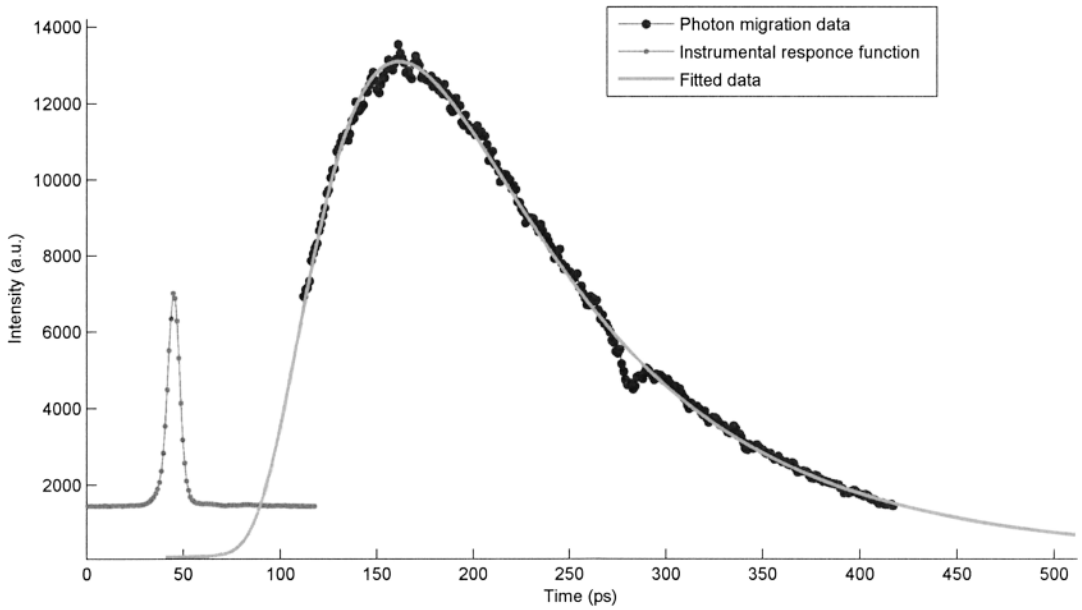


FIG. 2. An example of the evaluation of the time-resolved measurements, which was made by fitting the measured data to the time-resolved diffusion model. The measured data marked by the thicker line was used in the evaluation at this particular wavelength.

coefficients calculated from the time-resolved measurements at five wavelengths (855, 905, 955, 1005, and 1075 nm) were used in the scatter correction procedure. Using only five of all the available wavelengths had two objectives. First of all to mimic a simplified laser diode based system for *in situ* measurements, but also to facilitate the use of the other measured wavelengths to verify the correctness of the following steps of the evaluation scheme.

The scattering coefficients were calculated as an average over a 10 nm wide window to increase the signal-to-noise ratio. These values were used to calculate the scattering dependence on wavelength. The calculation was done by fitting the points to Eq. 4, which approximately describes the wavelength dependence of Mie scattering:[34]

$$\mu_s' = a\lambda^b \quad (4)$$

This approximation made it also possible to extrapolate the scattering coefficients into wavelength ranges not measurable by the present time-resolved system.

The extracted scattering coefficients from the Mie approximation were combined with the conventional NIR data and the steady-state diffusion model (Eq. 3) to extract the absorption coefficients in the entire wavelength range covered by the conventional NIR instrument (see Fig. 4). This calculation was also conducted using a Levenberg–Marquardt iterative procedure. The resulting absorption coefficients were independent of the path length of the light through the sample and therefore independent of the scattering properties of the sample.

Multivariate Calibrations. All multivariate calibration models were made in Simca-P 10.0 (Umetrics AB).

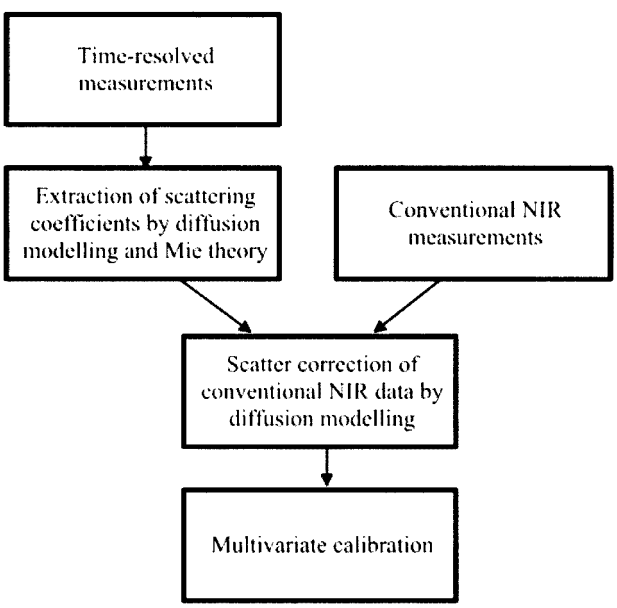


FIG. 3. Overview of the scatter correction procedure.

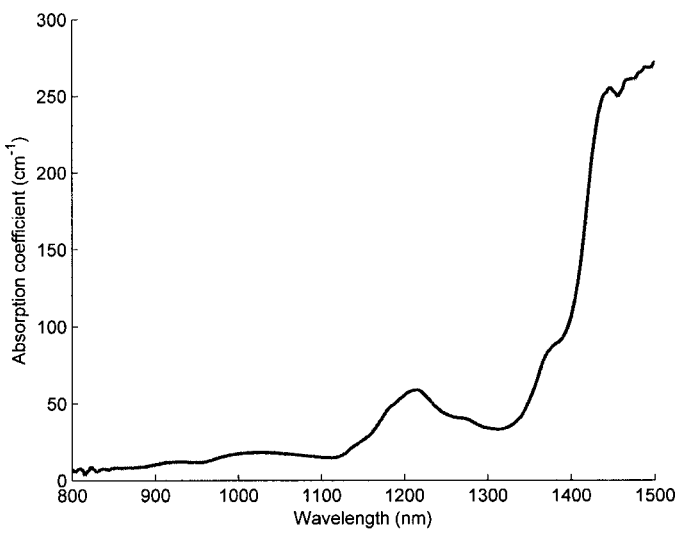


FIG. 4. The calculated absorption coefficients of a tablet from the batch with the highest nominal content of active substance and the medium sieve fraction.

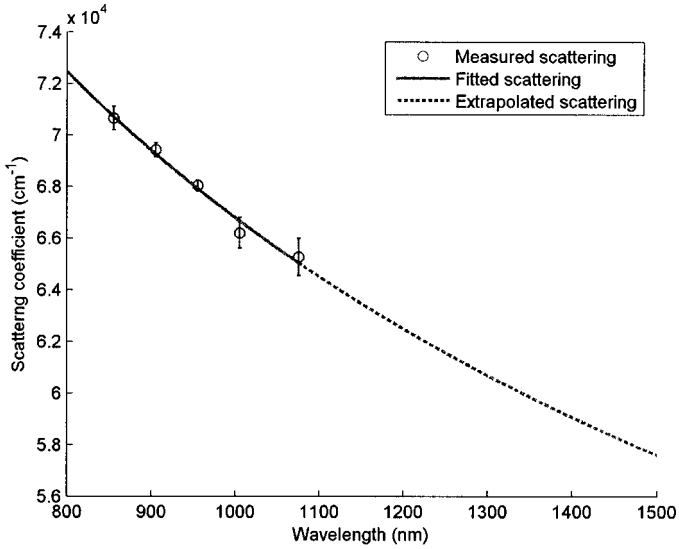


FIG. 5. A typical fit of the measured scattering to the equation defined by Mie theory. The measured scattering values are plotted to show the average and the standard deviation of ten measurements. The dotted line shows the extrapolation into longer wavelengths.

All spectra were mean centered before calculations and the number of principal components (PLSCs) selected in the models were as many as Simca-P 10.0 found suitable. The program uses the cross-validated predicted fraction for both X and Y to find the optimal number of PLSCs. In all models the samples not used in the calibration were used as a validation set, and the root mean square error of prediction value (RMSEP) (Eq. 5) was used to evaluate the performance of the different models.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad (5)$$

where $\hat{y}$ is the concentration of active substance predicted by the PLS model, y is the concentration of active substance measured by the reference analysis, and n is the number of samples. To get a better idea of the quality of the models, the RMSEP was divided by the mean concentration of the samples, giving the relative error in %.

RESULTS AND DISCUSSION

Deconvolution of Scattering and Absorption Properties of Samples Using Time-Resolved Data. The fitting of the time-resolved diffusion model to the time-resolved data was generally very good. The fit shown in Fig. 2 is typical for this step of the evaluations. A fit of the calculated scattering coefficients from the five wavelengths to the equation given by Mie theory is seen in Fig. 5. Values of b (see Eq. 4) were in the range from -0.25 to -0.5 for different samples.

After the scattering had been combined with the conventional NIR measurement and the steady-state diffusion model, the resulting absorption coefficients were compared to the absorption coefficients extracted from the time-resolved measurements alone. This comparison revealed that 70% of the samples showed a good agreement, with residuals below 10%, as seen in the left part of Fig. 6. The resulting 30% of the samples exhibit absorption coefficients that deviated from the absorption coefficients calculated directly from the time-resolved measurements. The deviations could be rather small, occurring just in limited wavelength regions, but some of the samples disagree completely, as seen in the right part of Fig. 6. The main source of error for this sometimes large deviation is thought to be errors introduced by an estimated time delay between the instrumental response function and the sample measurement, but also the signal-to-noise ratio in the evaluation of the scattering coefficients at the five wavelengths is crucial in order to obtain good results. The delay between the instrumental response function and sample measurements was calculated to be 15 ps. The delay occurs due to the insertion of a filter when measuring the instrumental response function, which was a necessity in order not to over-expose the detection system. Using this calculated delay gave unrealistic values of the absorption coefficients. Previous measurements on tissue phantoms, with known optical properties, have shown that by adding an extra 20

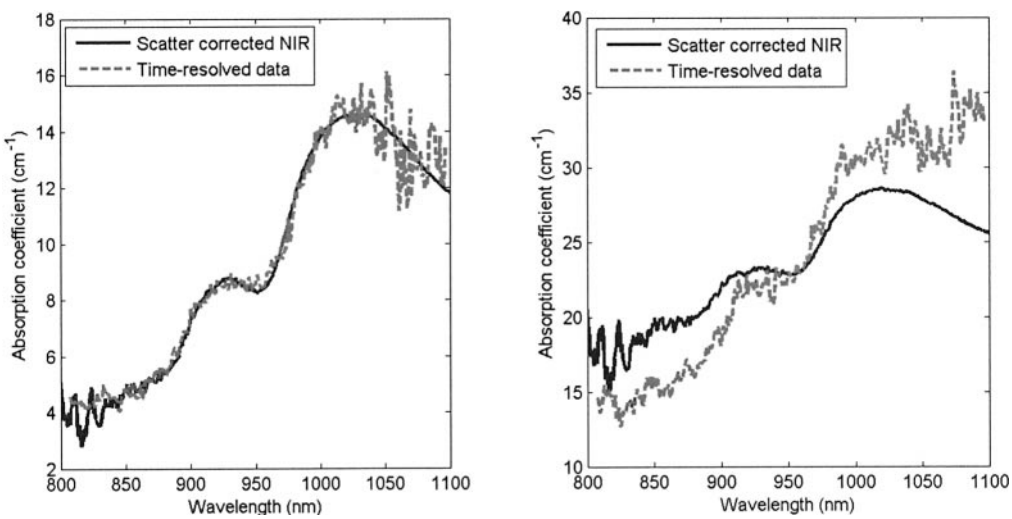


FIG. 6. Comparison between scatter corrected NIR spectra and absorption coefficient spectra calculated from time-resolved measurements.

ps to the calculated time delay, correct values of the absorption coefficient were gained. Therefore, a time delay of 35 ps was used in all evaluations. At this point, the reason for the extra time delay is not fully understood, but work to increase the understanding of the evaluation scheme is planned.

Quantitative Analysis Using Scatter Corrected Near-Infrared Data. To evaluate the data from the scatter correction scheme described above, a comparison with conventional NIR measurements was performed. The data from the scatter correction scheme will further on be referred to as the scatter corrected data. This to separate it from evaluations based on uncorrected conventional NIR data alone.

Basic Model. In order to compare the precision of the two methods, models based on half the data set were constructed and used to predict the other half. Both the calibration and validation sets included tablets from all nine batches. The model based on conventional NIR data resulted in an RMSEP value of 4.1% using five PLS components, while scatter corrected data resulted in an RMSEP value of 1.8% using six PLS components.

This evaluation shows that by correcting the conventional NIR data using time-resolved spectroscopy, the predictive ability of the constructed PLS models improved by more than 50%.

Models Based on Different Tablet Thicknesses. By building two different models, one only including the 17 thinnest tablets and one including only the 12 thickest tablets, a comparison of the robustness of the two methods was made. The validation sets contained the rest of the tablets, 65 and 70 samples respectively. To build a calibration model with that few samples, and to use it to predict samples with physical characteristics lying outside the parameter space spanned by the calibration samples, is troublesome when using conventional NIR data. The two models based on conventional NIR data showed the presumed quite poor predictive abilities, with RMSEP values of 5.2% and 11.2% for the models based on the thinnest and thickest tablets, respectively. The results from the scatter corrected data did not show the same drastic deterioration when compared to the basic model as the results from the conventional NIR data. When using scatter corrected data the RMSEP values for the two models were found to be 2.4% and 3.6% for the model based on the thinnest and thickest tablets, respectively.

This clearly shows that correcting conventional NIR data with time-resolved measurements at five wavelengths increases the robustness of the calibration models and makes it possible to predict samples with different physical dimensions than the tablets included in the calibration model (see Fig. 7).

Models Based on Tablets with Different Particle Size Distributions. To further compare the ability of the two techniques, all tablets manufactured from the largest particle size fraction (27 samples) were used as calibration set. When predicting the tablets made from the other two particle size groups the same kind of pattern as in the previous models was seen. The conventional NIR model gave prediction errors of 5.3% while the scatter corrected model showed a RMSEP value of 2.8%, further proving the robustness of the models based on scatter corrected data.

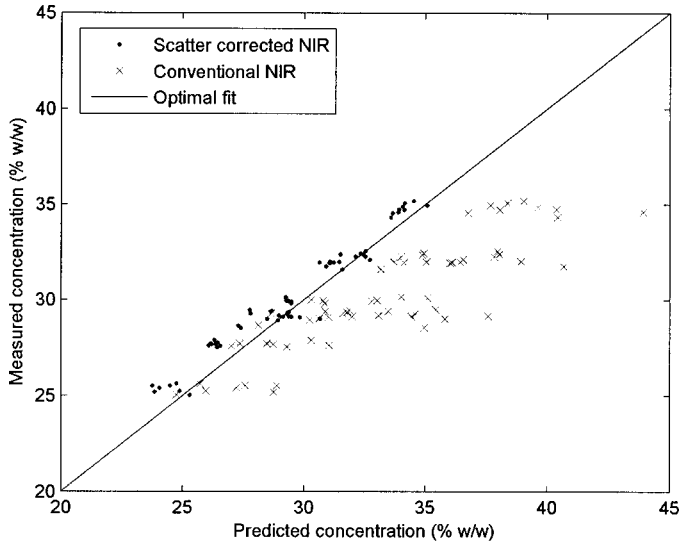


FIG. 7. Observed versus predicted values when predicting the 65 thickest tablets using a PLS models based on the 17 thinnest tablets. The scatter corrected data lies much closer to the line of optimal fit than the conventional NIR data.

Future Prospects. Although the instrumental setup for the time-resolved measurements used in this study only works in a research environment, the measurements and evaluations are conducted in a way that mimics a simplified laser diode based system. Such a system could be small and robust enough to be used for laboratory use as well as for on-line or at-line measurements in a process environment.

Combining a conventional NIR spectrometer with a simple time-resolved system could be an important step in making NIR spectroscopy more robust, making it possible to measure absorption spectra without any contribution from scattering effects. The technique might also be used for calibration transfer schemes[35,36] or other applications where additional information about the scattering properties of samples can complement conventional NIR data.

CONCLUSION

The scope of this work is a new methodology to correct conventional NIR data for scattering effects. The technique aims at measuring the absorption coefficient of the samples rather than the total attenuation, measured in conventional NIR spectroscopy. The main advantage of this is that the absorption coefficient is independent of the path length of the light inside the sample and therefore independent of the scattering effects.

The method is based on time-resolved spectroscopy and modeling of light transport by diffusion theory. This provides an independent measure of the scattering properties of the samples and therefore the path length of light. This yields a clear advantage over other preprocessing techniques, where scattering effects are estimated and corrected for by using the shape of the measured spectrum only.

Partial least squares calibration models show that, by using the proposed evaluation scheme, the predictive ability is improved by 50% as compared to a model based on conventional NIR data only. The method also makes

it possible to predict the concentration of active substance in samples with other physical properties than the samples included in the calibration model.

1. A. Eustaquio, P. Graham, R. D. Jee, A. C. Moffatt, and A. D. Trafford, Analyst (Cambridge, U.K.) **123,** 2303 (1998).
2. M. Dyrby, S. B. Engelsen, L. Norgaard, M. Bruhn, and L. Lundsberg-Nielsen, Appl. Spectrosc. **56,** 579 (2002).
3. P. Corti, G. Ceramelli, E. Dreassi, and S. Mattii, Analyst (Cambridge, U.K.) **124,** 755 (1999).
4. J. Gottfries, H. Depui, M. Fransson, M. Jongeneelen, M. Josefson, F. W. Langkilde, and D. T. Witte, J. Pharm. Biomed. Anal. **14,** 1495 (1996).
5. K. A. Martin, Appl. Spectrosc. Rev. **27,** 325 (1992).
6. R. B. Bruce, A. B. Mark, C. Show, Q. Xue-Zhi, and A. R. Priscilla, Pharm. Res. **13,** 616 (1996).
7. B. F. MacDonald and K. A. Prebble, J. Pharm. Biomed. Anal. **11,** 1077 (1993).
8. M. Andersson, O. Svensson, S. Folestad, M. Josefson, and K.-G. Wahlund, Chemom. Intell. Lab. Syst. **75,** 1 (2005).
9. J. Rantanen, H. Wikstrom, R. Turner, and L. S. Taylor, Anal. Chem. **77,** 556 (2005).
10. M. Blanco, J. Coello, A. Eustaquio, H. Iturriaga, and S. Maspoch, Anal. Chim. Acta **392,** 237 (1999).
11. M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, and C. de la Pezuela, Analyst (Cambridge, U.K.) **123,** 135R (1998).
12. J. Johansson, S. Folestad, M. Josefson, A. Sparen, C. Abrahamsson, S. Andersson-Engels, and S. Svanberg, Appl. Spectrosc. **56,** 725 (2002).
13. R. J. Barnes, M. S. Dhanoa, and S. J. Lister, Appl. Spectrosc. **43,** 772 (1989).
14. P. Geladi, D. MacDougall, and H. Martens, Appl. Spectrosc. **39,** 491 (1985).
15. S. Wold, H. Antii, F. Lindgren, and J. Ohman, Chemom. Intell. Lab. Syst. **44,** 175 (1998).
16. S. Andersson-Engels, R. Berg, O. Jarlman, and S. Svanberg, Opt. Lett. **15,** 1179 (1990).
17. R. M. P. Doornbos, R. Lang, M. C. Aalders, F. W. Cross, and H. J. C. M. Sterenborg, Phys. Med. Biol. **44,** 967 (1999).
18. J. W. Pickering, S. A. Prahl, N. van Wieringen, J. F. Beek, H. J. C. M. Sterenborg, and M. J. C. van Gemert, Appl. Opt. **32,** 399 (1993).
19. O. Berntsson, T. Burger, S. Folestad, L. G. Danielsson, J. Kuhn, and J. Fricke, Anal. Chem. **71,** 617 (1999).
20. T. Pan and E. M. Sevick-Muraca, Anal. Chem. **74,** 4228 (2002).
21. Z. Sun, S. Torrance, F. K. McNeil-Watson, and E. M. Sevick-Muraca, Anal. Chem. **75,** 1720 (2003).
22. R. R. Shinde, G. V. Balgi, S. L. Nail, and E. M. Sevick-Muraca, J. Pharm. Sci. **88,** 959 (1999).
23. A. Pifferi, A. Torricelli, A. Bassi, P. Taroni, R. Cubeddu, H. Wabnitz, D. Grosenick, M. Möller, R. Macdonald, J. Swartling, T. Svensson, S. Andersson-Engels, R. van Veen, H. J. C. M. Sterenborg, J.-M. Tualle, H. L. Nghiem, S. Avrillier, M. Whelan, and H. Stamm, Appl. Opt. **44,** 2104 (2004).
24. S. K. Wan, Z. X. Guo, S. Kumar, J. Aber, and B. A. Garetz, J. Quant. Spectrosc. Radiat. Transfer **84,** 493 (2004).
25. S. J. Madsen, B. C. Wilson, M. S. Patterson, Y. D. Park, S. L. Jacques, and Y. Hefetz, Appl. Opt. **31,** 3509 (1992).
26. A. Pifferi, R. Berg, P. Taroni, and S. Andersson-Engels, Opt. Soc. Am. **32767,** 311 (1996).
27. M. S. Patterson, B. Chance, and B. C. Wilson, Appl. Opt. **28,** 2331 (1989).
28. K. M. Yoo, F. Liu, and R. R. Alfano, Phys. Rev. Lett. **64,** 2647 (1990).
29. R. C. Haskell, L. O. Svaasand, T.-T. Tsay, T.-C. Feng, M. S. McAdams, and B. J. Tromberg, J. Opt. Soc. Am. A **11,** 2727 (1994).
30. C. Abrahamsson, J. Johansson, A. Sparén, and F. Lindgren, Chemom. Intell. Lab. Syst. **69,** 3 (2003).
31. C. Abrahamsson, T. Svensson, S. Svanberg, S. Andersson-Engels, J. Johansson, and S. Folestad, Opt. Express **12,** 4103 (2004).
32. R. R. Alfano and S. L. Shapiro, Phys. Rev. Lett. **24,** 592 (1970).
33. S. Coen, J. D. Harvey, R. Leonhart, J. C. Knight, W. J. Wadsworth, and P. S. J. Russell, Opt. Lett. **26,** 1356 (2001).
34. F. Bevilacqua, A. J. Berger, A. E. Cerussi, D. Jakubowski, and B. J. Tromberg, Appl. Opt. **39,** 6498 (2000).
35. R. N. Feudale, N. A. Woody, H. Tan, A. J. Myles, S. D. Brown, and J. Ferre, Chemom. Intell. Lab. Syst. **64,** 181 (2002).
36. T. Fearn, J. Near Infrared Spectrosc. **9,** 229 (2001).