Scatter correction of transmission near-infrared spectra by photon migration data: Quantitative analysis of solids

Abrahamsson, Christoffer; Lowgren, A; Stromdahl, B; Svensson, Tomas; Andersson-Engels, Stefan; Johansson, Jonas; Folestad, S

Published in:
Applied Spectroscopy

2005

Link to publication

Citation for published version (APA):

Total number of authors: 7

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Scatter Correction of Transmission Near-Infrared Spectra by Photon Migration Data: Quantitative Analysis of Solids

CHRISTOFFER ABRAHAMSSON,* 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\textsuperscript{1-4} and reflectance\textsuperscript{5,6} mode, and the development of fiber optical probes\textsuperscript{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.\textsuperscript{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,\textsuperscript{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,\textsuperscript{13} multiplicative scatter correction,\textsuperscript{14} and orthogonal signal correction,\textsuperscript{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,\textsuperscript{16} spatially resolved,\textsuperscript{17} and integrating sphere measurements.\textsuperscript{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\textsuperscript{19,20} as well as for particle size analysis.\textsuperscript{21} Measurements of the optical properties have also been used to make quantitative measurements of pharmaceutical powder blend homogeneity.\textsuperscript{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.\textsuperscript{23} A variety of different evaluation schemes have been developed for evaluating time-resolved data, ranging from simple evaluations like the final slope fitting\textsuperscript{24} to more complex schemes like diffusion\textsuperscript{25} and Monte Carlo models.\textsuperscript{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-
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, \(\mu_a\), 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, \(\mu_s\), 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. 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 \(\mu_s\) 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, 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. 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

\[
T(r, t) = \exp \left( -\mu_a c t - \frac{r^2}{4Dt} \right)
\]

where

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

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

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.

<table>
<thead>
<tr>
<th>Sieve fraction (μm)</th>
<th>&lt;150</th>
<th>150–400</th>
<th>&gt;400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal concentration (% weight)</td>
<td>28.5</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>31.8</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>34.9</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>

in phosphate buffer pH 3.0. 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. 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 and stimulated Raman scattering. 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.

**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 $\mu_s$ and $\mu'_s$ 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

![Diagram of instrumentation](image-url)
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.\(^{34}\)

\[
\mu_s' = a\lambda^b
\]  

(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. 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.

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

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

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
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.

**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 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.