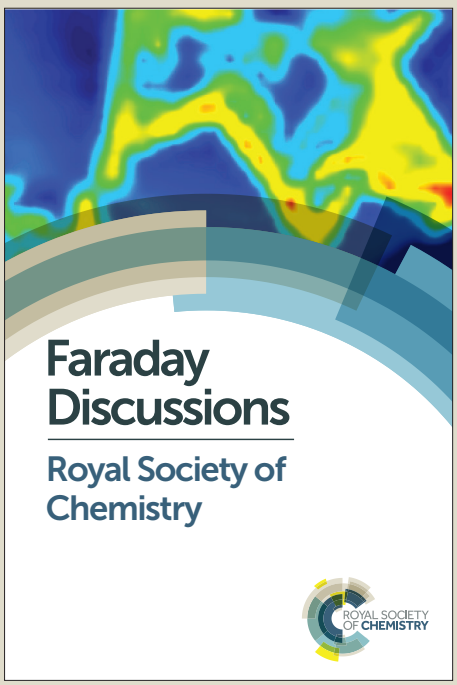
Faraday Discussions

Accepted Manuscript

This manuscript will be presented and discussed at a forthcoming Faraday Discussion meeting. All delegates can contribute to the discussion which will be included in the final volume.

Register now to attend! Full details of all upcoming meetings: http://rsc.li/fd-upcoming-meetings



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

Mie scatter corrections in single cell infrared microspectroscopy

Tatiana Konevskikh[1], Rozalia Lukacs[1], Reinhold Blümel[2], Arkadi Ponossov[1], Achim Kohler[1]

[1]Department of Mathematical Sciences and Technology (IMT), Norwegian University of Life Sciences, 1430 Ås, Norway

[2]Department of Physics, Wesleyan University, Middletown, Connecticut 06459-0155, USA

Key Words: Mie scattering, infrared microscopy, Extended Multiplicative Signal Correction (EMSC)

*Corresponding author: Tatiana Konevskikh

Tel: +47 67 23 15 84

Fax: +47 64 96 54 01

E-mail: tatiana.konevskikh@nmbu.no

Abbreviations: Extend multiplicative signal correction (EMSC), Multiplicative signal correction (MSC), Fast Fourier transform (FFT), Principal Component Analysis (PCA)

Abstract

Strong Mie scattering signatures hamper the chemical interpretation and multivariate analysis of infrared microscopy spectra of single cells and tissues. During recent years, several numerical Mie scatter correction algorithms for the infrared spectroscopy of single cells have been published. In the paper at hand, we critically reviewed existing algorithms for corrections of Mie scattering and suggest improvements. We developed an iterative algorithm based on Extended Multiplicative Scatter Correction (EMSC), for the retrieval of pure absorbance spectra from highly distorted infrared spectra of single cells. The new algorithm uses the Van de Hulst approximation formula for the extinction efficiency employing a complex refractive index. The iterative algorithm involves the establishment of an EMSC meta-model. While existing iterative algorithms for the correction of resonant Mie scattering employ three independent parameters for establishing a meta-model, we could decrease the number of parameters from three to two independent parameters, which reduced the calculation time for the Mie scattering curves for iterative EMSC meta-model by a factor 10. Moreover, by employing the Hilbert transform for evaluating the Kramers-Kronig relations based on an FFT algorithm in Matlab, we further improved the speed of the algorithm by a factor 100. For testing the algorithm we simulate distorted apparent absorbance spectra by utilizing the exact theory for the scattering of infrared light at absorbing spheres taking into account the high numerical aperture of infrared microscopes employed for the analysis of single cells and tissues. In addition, the algorithm was applied to measured absorbance spectra of single lung cancer cells.

..

Introduction

Since the invention of infrared microscopes in the 90s, infrared microscopy has become an attractive tool for the analysis of single cells and connective tissues in biological and medical sciences[1]. Via infrared microscopy, tissues and cells can be analysed chemically without destroying the chemical structure of the material. Unfortunately, strong scatter effects that hamper the chemical interpretation of spectra occur, which have been interpreted as Mie-type scattering[2]. According to Mie theory[3, 4], single cells are highly efficient scatterers in the infrared, since the wavelength of the infrared radiation is the same order as the size of the cells.

The correction of Mie type scattering in infrared spectra of single cells is difficult, since scattering and absorption are highly entangled. The reason for this is that due to the strong scattering, a large part of the scattered light does not reach the detector: only scattered light that is collected by the Schwarzschild optics reaches the detector, while a substantial part of the light is scattered into solid angles, which are not covered by the Schwarzschild optics. The scattered light that does not reach the detector leads to apparent absorption signatures in the measured absorbance spectrum, and the researcher is at a loss to decide which part of the measured absorbance spectrum is due to scattering and which part is due to chemical absorption. Thus, in FTIR microspectroscopy of single cells, the measured absorbance spectrum is also termed the apparent absorbance spectrum, since it contains both absorption and scattering signatures. Due to the difficulties in the interpretation of absorption peaks, scattering and absorption of infrared radiation at cells have been studied during recent years and algorithms have been established to separate scattering and absorption in apparent absorbance spectra in infrared spectroscopy of single cells[5-11]. All established algorithms, are based on Mie-theory, an exact theory for the scattering of infrared light at absorbing spheres developed by Gustav Mie in 1905[3]. The theory describes the scattering of electromagnetic radiation at spheres rigorously. Based on the exact Mie theory and approximation formulas thereof, a number of approaches have been developed, which aim at disentangling the scattering and absorption signatures in infrared spectra of single cells. In 2008, Kohler et al. developed an algorithm based on extended multiplicative signal correction (EMSC), which corrects Mie scattering in infrared spectra of cells[9]. The algorithm employs an approximation formula for the extinction efficiency and takes into account a constant refractive index. Since the exact size, morphology and refractive index of a cell are unknown in all practical situations, the algorithm employs a meta-model that comprises a range of cell sizes and refractive indices. The algorithm involves several approximations. (1) It

considers only the extinction efficiency in forward direction when calculating the absorbance, while infrared microscopes use Schwarzschild optics and collect light over a large numerical aperture. (2) The extinction in forward direction is calculated by an approximation formula developed by V. De Hulst[4]. (3) The algorithm considers the refractive index as constant and real in the employed wavelength range, while it is known that for absorbing scatterers, the refractive index has a non-zero imaginary part and a fluctuating real part, i.e., is dispersive (4). The algorithm approximates the apparent absorbance by the scattering efficiency. Despite the approximations employed, the algorithm corrects the broad Mie scatter oscillations in the apparent absorbance of single cell spectra, while so-called dispersive artefacts due to the resonant Mie effect remain uncorrected. The term 'dispersive artefact' circumscribes the fact that absorption resonances lead to fluctuations in the real part of the refractive index, which affect the extinction efficiency and thereby the measured absorbance spectrum. In 2010, Bassan et al. [6] developed an iterative algorithm based on EMSC, which allows correcting the dispersive artefact. The algorithm developed by Bassan et al. [6] involves several approximations. (1) As in Kohler's algorithm[9], the apparent absorbance is approximated by the extinction efficiency in forward direction developed by V. de Hulst[4]. (2) While Bassan's algorithm[6] takes into account a fluctuating real refractive index, it does not use the imaginary part of the refractive index. Recently, van de Dijk et al.[10] have introduced an iterative algorithm taking into account the high numerical aperture of infrared microscopes. The algorithm employs exact Mie formulas for the absorbance and treats the refractive index as complex. While it was shown that the algorithm corrects sharp dispersive Mie signatures in spectra of PMMA spheres[10], it has also been demonstrated that the algorithm is not applicable to biological samples, which have a complex composition and a shape that often deviates substantially from a sphere[11].

Although the algorithm of Bassan et al. [6] does not take into account a complex refractive index, it has been widely employed during recent years, since it provides a stable extraction of the pure absorbance spectrum. It is currently considered as the best existing method for correcting the Mie scatter distortions in infrared microspectroscopy of cells and tissues.

The aim of the present paper is to further develop the algorithm by Bassan et al. [6] by taking into account a complex refractive index according to Mie theory. The algorithm of Bassan et al. [6] requires the calculation of the Kramers-Kronig transform iteratively. Since this turns out to be a very time consuming process, we further aimed at improving the speed of the algorithm. In order to test the newly developed algorithm, we created a test set of apparent absorbance spectra. The test set consists of spectra that were distorted according

to exact Mie theory. For the creation of the test set, the optical set-up including the Schwarzschild optics with a focusing and collecting optics has been taken into account. The algorithm presented in this paper is based on the algorithm of Kohler et al.[9] and the algorithm of Bassan et al. [6] based on a meta-model taking into account a broad parameter range for parameters such as refractive index, size of the cell and effective sample thickness. While the resonant Mie scatter correction algorithm developed by Bassan et al. [6] uses three independent parameters for building the meta-model, we will show in the present paper that the meta-model can be set up with two independent parameters, which further improves the speed of the algorithm and the usage of memory in the modelling step.

Simulation of pure absorbance spectra

To validate the Mie scatter correction algorithm developed in this paper, a simulated data set of pure absorbance spectra was created. The spectra were simulated such that the obtained absorbance spectra resembled a matrigel spectrum of Bassan et al.[6], i.e. a spectrum of an artificial base membrane consisting mainly of proteins. The matrigel spectrum is considered as a nearly scatter-free pure absorbance spectrum. A random number generator was used to change of heights (±20%), and shift band positions (±1 cm^{-1}) of peaks by superimposing Lorentz lines for 50 spectra. This data set was divided into two data sets of 25 spectra. For data set one, we systematically changed the amplitudes at the peaks positions 1116, 1127, 1172, 1233, 1294, 1388, 1544 and 1648 cm^{-1} by making them either higher or smaller for each respective band. For data set two, we changed the amplitudes at the positions 1075, 1155, 1192, 1243, 1315, 1404, 1456, 1551 and 1656 cm^{-1} by again making them either higher or smaller for each respective band. When analysing these data sets by PCA, two clusters of samples were obtained according to the design of the simulation. The simulated spectra are shown in Fig. 1a. The corresponding score plot for the first two PCA components is shown in Fig. 1b. The first two components account in total for 78,3% of the variance in these data. From the simulated pure absorbance spectra, the imaginary parts of refractive indices $n'(\tilde{\nu})$ were calculated according to

$$n'(\tilde{\nu}) = \frac{A(\tilde{\nu})\ln(10)}{4\pi d_{eff}\tilde{\nu}}, \qquad (1)$$

where d_{eff} is the effective thickness of the cell[11] and $\tilde{\nu}$ is the wavenumber. The real part of the fluctuating part of the refractive index was calculated by using the Kramers-Kronig transform

$$n(\tilde{\nu}) = n_0 + \frac{2}{\pi}\mathrm{P}\int_0^{\infty}\frac{s\cdot n'(s)}{s^2-\tilde{\nu}^2}\,ds, \qquad (2)$$

where n_0 is the constant part of the real refractive index, P denotes the Cauchy principal value integral and n' is the imaginary part of the refractive index[12].

Simulation of apparent absorbance spectra according to exact Mie theory

In order to simulate apparent absorbance spectra, exact Mie theory was used. In the simulations, the optical setup of an infrared microscope with a numerical aperture *NA* was taken into account. The apparent absorbance was calculated according to

$$A_{app} = -\log_{10}\left(1 - \frac{\pi a^2}{G}Q_{ext} + \frac{1}{4\pi G\tilde{\nu}^2}\int_0^{\theta_{NA}}[i_1(\theta) + i_2(\theta)]\sin(\theta)\,d\theta\right), \quad (3)$$

where a is the radius of the spherical scatterer, G is the size of the aperture and $\tilde{\nu}$ is the wavenumber. The integration is performed over the numerical aperture θ_{NA} with integrand functions $i_{1,2}(\theta)$, which are calculated from the scattering amplitudes $S_{1,2}(\theta)$. Details are given in the Supplementary Material S.2 of paper Lukacs et. al[11]. For each simulated pure absorbance spectrum (see description in the previous section), imaginary and real parts of the refractive index were calculated according to Eqs. 1 and 2 and displayed in Fig. 1c and 1d, respectively. For each simulated apparent absorbance spectrum A_{app}, random values were chosen for n_0 and a from the intervals $1.1 < n_0 < 1.4$ and $2m\mu < a < 5.5m\mu$. Following this procedure, a set of 50 simulated apparent absorbance spectra was obtained. The apparent absorbance spectra are shown in Fig. 1e. The score plot of the first two components of the corresponding PCA analysis is shown in Fig. 1f. We can see that due to the scatter distortions, the two groups that were observed in the score plot of the first two components of the PCA of the pure absorbance spectra (Fig.1b) are now mixed. It is important to note that visualization of score plots of higher components reveals the grouping also in the apparent absorbance spectra. This is expected, since the scatter distortions are simply leading to a distortion of the main variation pattern, but the information related to the grouping according to the chemical difference is still obtained in the simulated apparent absorbance spectra. It is important to mention that the simulated apparent absorbance spectra as shown in Fig. 1c include the so-called dispersive artefact. When the apparent absorbance spectra in Fig. 1c are corrected according to the EMSC algorithm developed by Kohler et al. [9], the dispersive artefact can be clearly seen (results not shown). The dispersive artefact is due to the fluctuations of the real refractive index caused by the absorption resonances and is not corrected by the algorithm developed by Kohler et al. [9], since the model assumes a real and constant refractive index.

We further would like to mention that the apparent absorbance spectra in Fig. 1e contain ripples which are higher frequency oscillations. Ripples can be seen clearly in the region from 2800cm^{-1} to 1800cm^{-1}. These ripples are also appearing in other spectral regions, but are less visible since they overlap with chemical absorbance bands. While we have observed ripples in spectra of pollen[11], ripples are usually not dominant in spectra of human and animal cells and tissues.

Extended multiplicative signal correction and meta-modelling

For the extraction of the pure absorbance spectra from simulated apparent absorbance spectra, an iterative algorithm based on EMSC was developed, which is a further extension of the algorithms presented in Kohler et al.[9] and Bassan et al.[7]. Multiplicative signal correction (MSC) and EMSC have been introduced for pre-processing of near-infrared spectra[13, 14]. It has been shown in the past that both methods are versatile tools for correcting infrared spectra of biological materials[6, 9, 15-17]. When EMSC is used for estimating and correcting Mie scattering, a measured absorbance spectrum $A_{app}(\tilde{\nu})$ is approximated by a reference spectrum $Z_{ref}(\tilde{\nu})$ times a multiplicative effect b, plus deviations from this reference spectrum expressed by a constant baseline c plus a sum of components $p_i(\tilde{\nu})$ times respective parameters g_i

$$A_{app}(\tilde{\nu}) = bZ_{ref}(\tilde{\nu}) + c + \sum_{i=1}^{A_{opt}} g_i p_i(\tilde{\nu}) + \varepsilon(\tilde{\nu}). \qquad (4)$$

The un-modelled part is captured by the residual term $\varepsilon(\tilde{\nu})$. Further baseline effects may be included in the model by adding polynomials to the EMSC model[15, 16]. Since EMSC models an apparent absorbance spectrum around a reference spectrum, the estimation of the model parameter has turned out to be a very stable process. This is because infrared spectra of biological materials have very similar spectral signatures deriving from protein, fat and carbohydrate absorptions, leading to a visually very similar overall shape of the spectrum. When the parameters are estimated according to Eq. 4, the apparent absorbance spectrum is corrected according to

$$A_{corr}(\tilde{\nu}) = \frac{\left(A_{app}(\tilde{\nu}) - c - \sum_{i=1}^{A_{opt}} g_i p_i(\tilde{\nu})\right)}{b}. \qquad (5)$$

In an EMSC Mie model, the component spectra $p_i(\tilde{\nu})$ are obtained from a meta-model based on Mie theory. In the meta-model used by Kohler et al.[9] and Bassan et al. [7], the scatter extinction was approximated by the formula derived by Van de Hulst[4], which was originally developed for a constant and real refractive index and writes as

$$Q_{ext}(\tilde{\nu}) \approx 2 - \frac{4}{\rho} sin\,\rho + \left(\frac{4}{\rho^2}\right)(1 - cos\,\rho), \quad (6)$$

where ρ is given by

$$\rho = 4\pi a\tilde{\nu}(n-1) \quad (7)$$

and a and n are the radius of the spherical particle and the real refractive index, respectively. In order to allow for the correction of the dispersive artefact, Bassan et al. [7] introduced in Eq. 6 for n a non-constant real refractive index, which was calculated from an estimate of the pure absorbance according to the Kramers-

Kronig transform (Eq. 2), while the imaginary part $n'(\tilde{\nu})$ of the refractive index may be calculated from an estimate of the pure absorbance spectrum $A(\tilde{\nu})$ according to Eq. 1. This relation was simplified in Bassan et al. [7] assuming that the imaginary part $n'(\tilde{\nu})$ of the refractive index is proportional to the absorbance $A(\tilde{\nu})$, i.e.

$$n'(\tilde{\nu}) \approx s \cdot A(\tilde{\nu}) \qquad (8)$$

where s is a proportionality factor. The proportionality is only a rough approximation since the wavenumber varies on the range under consideration. When the imaginary part of the refractive index, $n'(\tilde{\nu})$, is known, the real part, $n(\tilde{\nu})$, can be calculated according to the Kramers-Kronig relation in Eq. 2.

The iterative algorithm developed by Bassan et al. works as follows:

The algorithm of Bassan corrects a scatter distorted apparent absorbance spectrum $A_{app}(\tilde{\nu})$ according the following interative procedure, which is based on the EMSC model in Eq. 4.

Initialization: Eq. 4 requires the initialization of a reference spectrum $Z_{ref}(\tilde{\nu})$, which ideally should be closed to the pure absorbance spectrum. To initialize the algorithm, a matrigel spectrum[6] was used as pure absorbance spectrum $A(\tilde{\nu})$.

Iterative algorithm:

1. The reference spectrum $Z_{ref}(\tilde{\nu})$ in Eq. 4 is replaced by an estimate of the pure absorbance spectrum.
2. The estimate of the pure absorbance spectrum is further used to calculate the imaginary part of the refractive index $n'(\tilde{\nu})$ according to Eq. 1., From the imaginary part of the refractive index $n'(\tilde{\nu})$, the real part of the refractive index, $n(\tilde{\nu})$, is calculated by the Kramers-Kronig transform according to Eq. 2. The real part of the refractive index is then used to calculate the extinction efficiency $Q_{ext}(\tilde{\nu})$ according to Eq. 3 for a wide range of parameters a, n_0 and the proportionality factor of Eq. 8, denoted by s. For each parameter typically its range is covered by 10 different values. The apparent absorbance $A_{app}(\tilde{\nu})$ is then calculated for the range of parameters a, n_0 and s by assuming that $A_{app}(\tilde{\nu}) \approx Q_{ext}(\tilde{\nu})$, resulting in a set of 10x10x10=1000 apparent absorbance spectra.

3. The set of apparent absorbance spectra $A_{app}(\tilde{\nu})$ is approximated by a meta-model using PCA, resulting in a set of principal components $p_i(\tilde{\nu})$ that are used as components in Eq. 4. The EMSC parameters in Eq. 4 are estimated by ordinary least square fits. After estimation of the EMSC parameters, the spectrum is corrected according to Eq. 5, resulting in a corrected spectrum $A_{corr}(\tilde{\nu})$.
4. The estimate of the pure absorbance spectrum is replaced by the corrected spectrum $A_{corr}(\tilde{\nu})$ and the algorithm is reiterated starting with step 1.

The iterative algorithm by Bassan et al.[6] deserves some comments. It is important to mention that the assumed porportionality $A_{app}(\tilde{\nu}) \sim Q_{ext}(\tilde{\nu})$ that is employed in step 1, is a rough approximation of Eq. 3 [11]. It involves neglecting the third term in Eq. 3 resulting in

$$A_{app} = -\log_{10}(1 - \frac{\pi a^2}{G} Q_{ext}) \qquad (9)$$

where a is the radius of the sphere and G is the detector area, and further the expansion of the logarithm ot Eq. 9 up to linear order, resulting in

$$A_{app} \approx \frac{\pi a^2}{G} Q_{ext}, \qquad (10)$$

explaining the proportionality between $A_{app}(\tilde{\nu})$ and $Q_{ext}(\tilde{\nu})$.

Improved algorithm suggested in this paper.

The iterative algorithm suggested in this paper involves several improvements. First, the iterative algorithm in this paper is based on a meta-model, which involves a complex refractive index according to the Mie theory. We used an approximation formula for the extinction efficiency, that has been derived by Van de Hulst[4]

$$Q_{ext}(\tilde{\nu}) \approx 2 - 4e^{-\rho \tan\beta} \frac{\cos\beta}{\rho} \sin(\rho - \beta) - 4e^{-\rho \tan\beta} \left(\frac{\cos\beta}{\rho}\right)^2 \cos(\rho - 2\beta) + 4\left(\frac{\cos\beta}{\rho}\right)^2 \cos(2\beta) \quad (11)$$

with

$$\rho = 4\pi a \tilde{\nu}(n-1) \quad \text{and} \quad \tan\beta = n'/(n-1) \qquad (12)$$

where n and n' are the real and the imaginary parts of the refractive index, respectively. The differences between Eq. 6 and Eq. 11 are illustrated in[11]. The differences are significant, both with respect to the position of the band and the absolute values of the estimated extinction and absorbance.

The second improvement relates to the complexity of the meta-model. While in Bassan et al. [6] a parameter model was used including ranges of three parameters (3 dimensions), it can be shown that a 2-dimensional parameter model is sufficient for the parameter estimation. Details about the reduction from the 3-parameter model to the 2-parameter model are given in the appendix A. The application of a 2-dimensional parameter model decreases the computation time and required of computer memory. This can be illustrated by an example. When the three parameter ranges represented by the size of the cell, the background refractive index and the scaling parameter s of Eq. 8 are covered and 10 values are used to cover each parameter range, 1000 Mie extinction curves need to be simulated and used for the establishment of the meta-model. Employing a two-parameter model, 100 spectra are sufficient to cover the same range. Thus, memory usage is reduced by a factor 10 and the computation time for establishment of the meta-model is reduced as well.

A third improvement of the algorithm was achieved by speeding up the Kramers-Kronig transform of Eq. 2. Kramers-Kronig relations are mainly used in optical spectroscopy to determine the complex refractive index $\hat{n} = n + in'$ of the medium from the measured absorption, transmission or reflection spectrum. The refractive index is an important quantity when considering the scattering and absorption of light at biological materials. The real part n of this index describes the refractive properties of the material; the imaginary part n' of it, determines the absorptive properties of the material. The most employed method for obtaining n from n' and vice versa is the Kramers-Kronig transform, which expresses the real part n in terms of the imaginary part n' according to

$$n(\tilde{\nu}) = \frac{2}{\pi} P \int_0^{+\infty} \frac{s \cdot n'(s)}{s^2 - \tilde{\nu}^2} ds, (13a)$$

and the imaginary part n' in terms of the real part n by

$$n'(\tilde{\nu}) = -\frac{2\tilde{\nu}}{\pi} P \int_0^{+\infty} \frac{n(s)}{s^2 - \tilde{\nu}^2} ds. (13b)$$

In Appendix B, we show that these relations are equivalent to the Hilbert transform, since the real part of the refractive index is an even function of the wavenumber and the imaginary part of the refractive index is an odd function of wavenumber, i.e. $n(\tilde{\nu}) = n(-\tilde{\nu})$ and $n'(\tilde{\nu}) = -n'(-\tilde{\nu})$. This can be seen from the Lorentz model. Details are given in Supplementary Material S.1 of reference[11].The Hilbert transform writes as

$$n(\tilde{\nu}) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{n'(s)}{s - \tilde{\nu}} ds = -\frac{1}{\pi\tilde{\nu}} * n'(\tilde{\nu}) \ (14a)$$

and

$$n'(\tilde{\nu}) = -\frac{1}{\pi} P \int_{0}^{+\infty} \frac{n(s)}{s - \tilde{\nu}} ds = \frac{1}{\pi\tilde{\nu}} * n(\tilde{\nu}) \ (14b)$$

respectively, where $*$ denotes convolution. The Hilbert transform can be calculated via Fast Fourier Transform (FFT). The FFT method is based on the fact that both real and imaginary parts of the complex refractive index defining the Hilbert transform are proportional to the convolution product between $n(\tilde{\nu})$ or $n'(\tilde{\nu})$ and the convolution kernel function $\frac{1}{\pi\tilde{\nu}}$. In Eqs. 14 there is a singularity, when s is equal to $\tilde{\nu}$. The singularity problem is theoretically bypassed by introducing the Cauchy Principal Value[18].

The new iterative algorithm works as follows:

The algorithm corrects a scatter distorted apparent absorbance spectrum $A_{app}(\tilde{\nu})$ according to the following interative procedure which is based on the EMSC model in Eq. 4.

Initialization: The reference spectrum $Z_{ref}(\tilde{\nu})$ in Eq. 4 is initialized by a matrigel spectrum[6] or another appropriate non-distorted spectrum.

Iterative algorithm:

1. The reference spectrum $Z_{ref}(\tilde{\nu})$ in Eq. 4 may be either updated[6] or the same reference spectrum may be used in each iteration.
2. The estimate of the pure absorbance spectrum is further used to calculate the imaginary part of the refractive index $n'(\tilde{\nu})$ according to Eq. 1. The real part of the refractive index, $n(\tilde{\nu})$, is calculated from the imaginary part of the refractive index, $n'(\tilde{\nu})$, by the Hilbert transform (function Hilbert in Matlab) according to Eq. 14a. The calculations of the Hilbert function are based on the FFT algorithm. The real part of the refractive index is then used to calculate the extinction efficiency $Q_{ext}(\tilde{\nu})$ according to Eq. 11 involving a complex refractive index. In order to cover the relevant parameter range, the scaled parameters γ and α_0 according to Eqs. A20 and A21 are used (see Appendix A) with the ranges $\alpha_0 \in [0.2\,\mu m, 2.2\,\mu m] \cdot 4\pi$ and $\gamma \in \left[5 \cdot 10^4 \frac{1}{m},\ 6 \cdot 10^5 \frac{1}{m}\right]$. The corresponding ranges of n_0, a and f are given in A13. A set of 50 apparent absorbance $A_{app}(\tilde{\nu})$ is

then calculated according to Eq. 9 uniformly covering the parameter ranges for α_0 and γ, and stored as rows in a matrix M.

3. The set of apparent absorbance spectra $A_{app}(\tilde{\nu})$, i.e. the matrix M is approximated by a meta-model using PCA resulting in a set of principal components $p_i(\tilde{\nu})$ that are used as components in Eq. 4. Prior to approximation by PCA, the matrix M is orthogonalized with respect to the reference spectrum $Z_{ref}(\tilde{\nu})$. This is to avoid competition between the parameters b and g_i (see Eq. 4) in the subsequent parameter estimation. The EMSC parameters in Eq. 4 are estimated by ordinary least squares fits. After estimation of the EMSC parameters, the spectrum is corrected according to Eq. 5 resulting in a corrected spectrum $A_{corr}(\tilde{\nu})$.
4. The estimate of the pure absorbance spectrum is replaced by the corrected spectrum $A_{corr}(\tilde{\nu})$ and the algorithm is reiterated starting with step 1.

This iterative algorithm is described schematically in Fig. 2. We suggest to call the algorithm fast resonant Mie scatter correction.

EMSC correction of simulated and measured apparent absorbance spectra

The correction of an apparent absorbance spectrum according to the algorithm suggested in this paper is demonstrated in Fig. 3. In Fig. 3*a* an example of a simulated apparent absorbance spectrum is shown. The spectrum is simulated employing Eq. 3. As parameters we used $n_0 = 1.35$ and $a = 3.81\ \mu m$. Corrected spectra that were obtained by the first three iterations of the correction algorithm proposed in this paper are shown in Fig. 3*b* together with the reference spectrum and the pure absorbance spectrum. As a reference spectrum we used the average spectrum of all simulated pure absorbance spectra. Alternatively, the matrigel spectrum could be used. It can be seen that the algorithm converges quickly and that the corrected spectra are close to the pure absorbance spectra used for the simulation. Ripples that can be observed in the region between 4000 cm^{-1} and 3500 cm^{-1} were not corrected successfully. This is expected, since the Van de Hulst formula used for the correction (see Eq. 11) does not describe the Mie ripples. In Fig. 4*a*, a simulated apparent absorbance spectrum containing strong ripples is shown. As parameters, we used $n_0 = 1.36$ and $a = 5.12\ \mu m$. Ripples with these intensities are usually not observed in infrared microspectroscopy of single cells. An explanation for this will be given elsewhere. In the corrected spectra in Fig. 4*b* it can be seen clearly that the Van De Hulst algorithm is not capable of correcting the ripples.

The complete set of corrected spectra for the simulated apparent absorbance spectra from Fig.1e is shown in Fig. 5. The spectra that correspond to the red scores of the pure absorbance spectra in Fig. 1*b* are plotted red in Fig. 5, while the spectra that correspond to the blue scores of the pure absorbance spectra in Fig. 1 *b* are plotted blue in Fig. 5. The reference spectrum is shown in green. It can be seen that the correction of the spectra worked well. The two groups of spectra can be separated visually after correction. As mentioned before, ripples are not well corrected, which does not pose a major problem, since strong ripples are usually not observed in the infrared microspctroscopy of single cells. The computation time for the correction of 50 spectra (one iteration in the iterative algorithm) using a 4th generation Intel ®Core™i7-4702HQ quad core processor,16GB memory and Matlab 2015 is 77 seconds. This is a substantial improvement to the algorithm developed by Bassan et al.[6] which uses 765 seconds on the same computer (employing numerical integration and a parameter model with 3 parameters). No speed optimization of the algorithm by vectorization or parallel programming was done so far. It is important to mention that the Kramers-Kronig part of of the algorithm is improved by a factor of approximately 200 when using the algorithm based on the Hilbert transform compared to the numerical integration. The improvement of the Kramers-Kronig part shows especially advantages when more than 1 iterations are applied in the iterative algorithm.

In order to demonstrate that chemical information that was distorded in the apparent absorbance spectra can be restored, we investigated peak ratios before and after correction. In Fig. 6, the peak ratios for the simulated pure absorbance spectra of the band at 1546 cm^{-1} and the band at 1387 cm^{-1} are shown in blue. The peak ratio of this band is simulated such that the 50 spectra can be clearly separated into two groups of 25 spectra each. The spectra 1-25 have a lower peak ratio than the spectra 26-50. The corresponding peak ratios for the apparent absorbance spectra are shown in magenta. It is obvious that the scatter distorted apparent absorbance spectra do not allow to separate these two groups with the help of the peak ratio of the band at 1546 cm^{-1} and the band at 1387 cm^{-1}. The apparent absorbance spectra were then corrected by the iterative algorithm. Corrected spectra are shown in red (one iteration), green (two iterations) and black (three iterations). The ratio obtained from the average spectrum of all simulated pure absorbance spectra that was used as the referense spectrum for the EMSC model is shown as the blue dashed line for comparison. It can be seen that the iterative algorithm retreives a good estimation of the ratio of the simulated pure absorbance spectrum.

We further tested the algorithm on a set of measured infrared microspectroscopy spectra of lung cancer cells. Details about the data set can be found in Kohler et al [9]. The measured spectra are shown in Fig. 7 (red) together with the corrected spectra (green). For the correction one iteration step was used. As reference spectrum the matrigel spectrum was used[6]. In Fig. 7, it can be seen that correction works visually very well. The resonant Mie effect that was not corrected by the algorithm presented in Kohler et al[9], is now successfully corrected.

The importance of the reference spectrum for the iterative algorithm

An important comment relates to the updating of the EMSC reference spectrum in the iterative algorithm. The dashed line in Fig. 2 indicates, that there are two options: Either the reference spectrum is updated by the new estimate of the pure absorbance spectrum obtained after each iteration or the initial reference spectrum is kept and only the imaginary part of the refractive index and, after Kramers-Kronig transformation, the real part of the refractive index are updated. We have observed that an update of the reference spectrum may lead to instabilities in the iterative algorithm and the result obtained may completely depend on the number of iterations used, i.e. the iterative algorithm may drift completely apart from reasonable solutions. Thus, we do not suggest to update the reference spectrum for EMSC in the iterative algorithm.

The use of a good reference spectrum is crutial for the success of the correction algorithm. In Fig.7 it is obvious that the corrected spectra become in their overall shape very close to the reference spectrum. This can be clearly seen when comparing the reference spectrum (for example in Fig. 4b) with the corrected spectra in Fig. 7. The same observation can be made in the paper of Bassan et al. [6], where the same reference spectrum was used as in the current paper. In reference[6], after correction all spectra are in their overall shape similar to the matrigel spectrum used as a reference in EMSC. This is due to the high flexibility of the meta-model used for the correction. Notwithstanding, although the corrected spectra tend overally in their shape towards the reference spectrum, we could clearly show that chemical information can be restored by the suggested algorithm.

Conclusions

In this work we presented an improved iterative EMSC algorithm for correcting Mie scattering in infrared microspectroscopy of single cells and tissues. The iterative EMSC algorithm employs a meta-model based on

an approximate formula by Van De Hulst, taking into account a complex refractive index for correcting Mie scattering. The new iterative algorithm was tested using a simulated set of apparent absorbance spectra and a set of measured apparent absorbance spectra. The simulated apparent absorbance spectra were obtained by first simulating pure absorbance spectra and then generating scatter distorted apparent absorbance spectra by full Mie theory. For the simulations, the full optical set up of an infrared microscope including the collecting and focusing Schwarzschild optics were taken into account. The set of simulated spectra used in this paper is thus more difficult to correct than the spectra simulated in the reference[6], where spectra were simulated by using exactly the same approximation formula that was afterwards used for correction algorithm. In addition, the spectra simulated in reference[6] were obtained by addition of pure absorbance spectra to simulated scatter spectra. This facilitated the correction of the spectra in reference[6], since scatter distortions where only added on top of the pure absorbance spectra. In our simulations the scatter distorted spectra where obtained according to Eq. 3 and do not contain additive contributions of pure absorbance spectra. It is important to mention that according to the Mie theory, apparent absorbance spectra are not obtained by adding pure absorbance spectra to scatter contributions. Thus, the retrieval of chemical information from the simulated apparent absorbance spectra taking into account the full optical setup as employed in the current paper, demonstrates well the capability of the algorithm presented in this paper.

The measured spectra used for correction in the current paper were obtained from single lung cancer cells[9]. The correction of the measured spectra was successful, while it is obvious that the corrected spectra tend in their overall shape towards the matrigel spectrum employed. The high flexibility of the meta-model results in corrected spectra that are in their overall shape very similar to the reference spectrum used in the EMSC model, while we showed by employing a simulated data set that chemical characteristics of the pure absorbance spectra could be restored.

The iterative algorithm developed by Bassan et al. [6] involves a numerical integration in order to perform a Kramers-Kronig transform. In the algorithm presented in this paper, we replaced the integral of the Kramers-Kronig by a fast Fourier transform (FFT) algorithm. This reduced the computation time of the Kramers-Kronig transform approximately by a factor 100. Moreover we have shown that two independent parameters α_0 and γ (each parameter contains 10 equidistant values in its respective range) are sufficient for compressing 100 Mie scattering curves into a small number of principal components loading spectra to estimate the scattering contributions in the EMSC meta-model.

While we have shown that the new algorithm retrieves the pure absorbance spectra from highly distorted apparent absorbance spectra, ripples that are present in simulated apparent absorbance spectra could not be corrected, since the Van de Hulst approximation does not account for ripples. The ripples that are visible in the simulated apparent absorbance spectra can be explained by diffractive surface waves. While we have observed that the appearance of ripples is not present or suppressed in measured spectra, we believe that the correction of ripples is in general not required for the correction of infrared microspectroscopic spectra of single cells and tissues. Ripples may be absent in measured spectra because the apertures used in infrared microspectroscopy are comparable to the size of the cells, while the exact Mie theory which assumes an incoming plane wave assumes apertures that are much bigger than the probed cell. Thus, diffractive surface waves that are causing ripples in exact Mie theory may be suppressed in practical measurement situations.

Acknowledgements

The authors wish to thank Peter Gardener and Paul Bassan for providing the matrigel and Josep Sule Suso for providing the set of measured lung cancer cell spectra.

Appendix A: Complexity of meta model

For building a meta-model we take into account a suitable parameter space for the approximation formula employed for the extinction efficiency. The approximation formula given by Eq. 11 contains the parameters ρ and β. We define

$$\alpha = 4\pi a(n-1), \qquad \text{(A1)}$$

where n is the real refractive index and a is the size of the sphere. It follows that $\rho = \tilde{\nu}\alpha$.

When the pure absorbance spectrum $A(\tilde{\nu})$ is known, the imaginary part of the refractive index can be calculated according to

$$n'(\tilde{\nu}) = \frac{A(\tilde{\nu})\ln(10)}{4\pi d_{eff}\tilde{\nu}}, \qquad \text{(A2)}$$

where d_{eff} is the effective thickness of the cell[11]. The real part of the refractive index $n(\tilde{\nu})$ can then be calculated by the Kramers-Kronig relation according to

$$n(\tilde{\nu}) = n_0 + \frac{2}{\pi}\mathbf{P}\int_0^\infty \frac{s \cdot n'(s)}{s^2 - \tilde{\nu}^2} ds, \qquad \text{(A3)}$$

where n_0 is the constant part of the real refractive index and the integral term

$$n_{kk}(\tilde{\nu}) = \frac{2}{\pi} P \int_0^\infty \frac{s \cdot n'(s)}{s^2 - \tilde{\nu}^2} ds \qquad \text{(A4)}$$

is the fluctuating part of the real refractive index.

It follows that

$$\alpha = 4\pi a(n_0 + n_{kk} - 1). \qquad \text{(A5)}$$

If n_{kk} is calculated from the pure absorbance spectrum $A(\tilde{\nu})$ according to Eqs. A2 and A4, the effective thickness d_{eff} of the cell is in general not known. We therefore define a scaled imaginary part of the refractive index, $n_s'(\tilde{\nu})$, according to

$$n_s'(\tilde{\nu}) = \frac{A(\tilde{\nu})}{\tilde{\nu}}, \quad \text{(A6)}$$

where

$$n'(\tilde{\nu}) = f n_s'(\tilde{\nu}), \qquad \text{(A7)}$$

with the scaling constant

$$f = \frac{\ln(10)}{4\pi d_{eff}}. \quad \text{(A8)}$$

The fluctuating part is consequently calculated according to

$$n_{kk}(\tilde{\nu}) = f n_{kk,s}(\tilde{\nu}) \qquad \text{(A9)}$$

where

$$n_{kk,s}(\tilde{\nu}) = \frac{2}{\pi} P \int_0^\infty \frac{s \cdot n_s'}{s^2 - \tilde{\nu}^2} ds. \quad \text{(A10)}$$

Thus, we obtain

$$\alpha = 4\pi a\left(n_0 + f n_{kk,s} - 1\right) \quad \text{(A11)}$$

and

$$tan\beta = \frac{fn_s'}{n_0 + fn_{kk,s} - 1}. \qquad \text{(A12)}$$

Generally the exact numbers for a, n_0 and f are not known. Only ranges for these values can be given. These parameters have to be estimated in the modeling process. The following ranges were found to be optimal

$$n_0 \in [1.1, 1.4], \qquad \text{(A13}a\text{)}$$

$$f \in \left[10^4 \frac{1}{m}, 3 \cdot 10^4 \frac{1}{m}\right], \qquad \text{(A13}b\text{)}$$

$$a \in [2\mu m, 5.5\mu m]. \qquad \text{(A13}c\text{)}$$

Since the parameters a and n_0 are not independent (see A5), these parameter ranges are not absolute. We include for example considerably higher values of a than indicated by the range in (A13c), when the refractive index in an apparent absorbance spectrum is below the maximum value of (A13a), since the two parameters enter Eq. A5 as a product. The parameters in Eqs. A13 are not independent and one parameter can be omitted by rescaling as we will show in the following. We can write Eq. A11 as

$$\alpha = 4\pi a(n_0 - 1)\left(1 + \frac{fn_{kk,s}}{(n_0 - 1)}\right), \qquad \text{(A14)}$$

and Eq. A12 as

$$tan\beta = \frac{n_s'}{\frac{n_0 - 1}{f} + n_{kk,s}}. \qquad \text{(A15)}$$

By defining

$$\alpha_0 = 4\pi a(n_0 - 1), \qquad \text{(A16)}$$

and

$$\gamma = \frac{f}{(n_0 - 1)}, \qquad \text{(A17)}$$

we obtain

$$\alpha = \alpha_0\left(1 + \gamma n_{kk,s}\right), \qquad \text{(A18)}$$

and

$$tan\beta = \frac{n_s'}{1/\gamma + n_{kk,s}} \qquad \text{(A19)}$$

with the respective ranges that correspond to the values in Eqs.A13:

$$\gamma \in \left[5 \cdot 10^4 \frac{1}{m},\ 6 \cdot 10^5 \frac{1}{m}\right] \qquad \text{(A20)}$$

$$\alpha_0 \in [0.2\mu m, 2.2\mu m] \cdot 4\pi \qquad \text{(A21)}$$

It follows that the obtained model contains two independent parameters γ and α_0. The parameter γ is the new scaling value for the non-constant refractive index. A correct estimate of this parameter is important for the biochemical interpretation of the FTIR spectra. The parameter α_0 corresponds to the α-value of the non-resonant case used in Kohler et al[9]. A good estimation of the α_0-value is important for the estimation of the physical parameters as size of cells a and constant part of the real refractive index n_0.

Appendix B: Hilbert transform and Kramers-Kronig relations for a calculation of the complex refractive index

In order to define the complex refractive index $\hat{n}(\tilde{\nu}) = n(\tilde{\nu}) + in'(\tilde{\nu})$ of materials from measured absorption, transmission or reflection spectra in optical spectroscopy, Kramers-Kronig relations are used. The complex refractive index is an important physical quantity when considering the scattering and absorption of infrared light at biological materials. The real part $n(\tilde{\nu})$ of this index describes the refractive properties of the material; the imaginary part $n'(\tilde{\nu})$ of it determines the absorptive properties of the material. A method for obtaining $n(\tilde{\nu})$and $n'(\tilde{\nu})$ is the Kramers-Kronig transform, which relates the real part $n(\tilde{\nu})$ to the imaginary part $n'(\tilde{\nu})$ of the complex refractive index:

$$n(\tilde{\nu}) = \frac{2}{\pi} P \int_0^{+\infty} \frac{s \cdot n'(s)}{s^2 - \tilde{\nu}^2} ds, \qquad \text{(B1}a\text{)}$$

$$n'(\tilde{\nu}) = -\frac{2\tilde{\nu}}{\pi} P \int_0^{+\infty} \frac{n(s)}{s^2 - \tilde{\nu}^2} ds, \qquad \text{(B1}b\text{)}$$

where *P* denotes Cauchy's Principal Value.

These relations are equivalent to the Hilbert transform, provided that the real part is an even function of wavenumber and the imaginary part is an odd function of wavenumber, i.e. $n(\tilde{\nu}) = n(-\tilde{\nu})$ and $n'(\tilde{\nu}) = -n'(-\tilde{\nu})$:

$$n(\tilde{\nu}) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{n'(s)}{s-\tilde{\nu}} ds = -\frac{1}{\pi\tilde{\nu}} * n'(\tilde{\nu}), \qquad \text{(B2}a\text{)}$$

$$n'(\tilde{\nu}) = -\frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{n(s)}{s-\tilde{\nu}} ds = \frac{1}{\pi\tilde{\nu}} * n(\tilde{\nu}). \ \text{(B2}b\text{)}$$

Let us consider the connections between Eqs. B1 and B2. Because of $n'(\tilde{\nu})$ is an odd function of wavenumber, this property permits the conversion of the Hilbert transform pair (Eq.B2*a*) into the Kramers-Kronig relations. The Hilbert transform pair (Eq.B2*a*) is expanded as

$$n(\tilde{\nu}) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{n'(s)}{s-\tilde{\nu}} ds = \frac{1}{\pi} P \int_{-\infty}^{0} \frac{n'(s)}{s-\tilde{\nu}} ds + \frac{1}{\pi} P \int_{0}^{+\infty} \frac{n'(s)}{s-\tilde{\nu}} ds$$

with $n'(\tilde{\nu}) = -n'(-\tilde{\nu})$ we get

$$n(\tilde{\nu}) = \frac{1}{\pi} P \int_{0}^{+\infty} \frac{n'(s)}{s+\tilde{\nu}} ds + \frac{1}{\pi} P \int_{0}^{+\infty} \frac{n'(s)}{s-\tilde{\nu}} ds = \frac{2}{\pi} P \int_{0}^{+\infty} \frac{s \cdot n'(s)}{s^2-\tilde{\nu}^2} ds.$$

Because of $n(\tilde{\nu})$ is an even function of wavenumber, this property permits the conversion of the Hilbert transform pair (Eq.B2*b*) into the Kramers-Kronig relations. The Hilbert transform pair (Eq.B2*b*) is expanded as

$$n'(\tilde{\nu}) = -\frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{n(s)}{s-\tilde{\nu}} ds = -\frac{1}{\pi} P \int_{-\infty}^{0} \frac{n(s)}{s-\tilde{\nu}} ds - \frac{1}{\pi} P \int_{0}^{+\infty} \frac{n(s)}{s-\tilde{\nu}} ds$$

with $n(\tilde{\nu}) = n(-\tilde{\nu})$

$$n'(\tilde{\nu}) = \frac{1}{\pi} P \int_{0}^{+\infty} \frac{n(s)}{s+\tilde{\nu}} ds - \frac{1}{\pi} P \int_{0}^{+\infty} \frac{n(s)}{s-\tilde{\nu}} ds = -\frac{2\tilde{\nu}}{\pi} P \int_{0}^{+\infty} \frac{n(s)}{s^2-\tilde{\nu}^2} ds.$$

The Hilbert transform can be calculated via the Fast Fourier Transform (FFT). The FFT method is based on the fact that both real and imaginary parts of the complex refractive index defining the Hilbert transform are proportional to the convolution product between $n(\tilde{\nu})$ or $n'(\tilde{\nu})$ and the convolution kernel function $\frac{1}{\pi\tilde{\nu}}$.

In Eqs. B2 there is a singularity, when s is equal to $\tilde{\nu}$. The singularity problem is theoretically bypassed by introducing the Cauchy Principal Value.

Derivations of Hilbert transform and Kramers-Kronig relations

If $\hat{n}(\tilde{\nu})$ is an analytic function, then $\frac{\hat{n}(s)}{s-\tilde{\nu}}$ is an analytic function too except at the pole $s = \tilde{\nu}$. Cauchy's theorem states that

$$\oint \mathrm{f}(s)\mathrm{d}s = 0,$$

provided that the closed contour encloses no poles of the analytic function $f(s)$. Let us apply Cauchy's theorem to the function $\frac{\hat{n}(s)}{s-\tilde{\nu}}$, where $\tilde{\nu}$ is a point on the real axis, and the contour, shown in Fig. 8, is the union of four curves with parametric representations

$$\mathrm{C}_1\colon s = \Omega, \text{ where } -R \leq \Omega \leq \tilde{\nu} - \rho;$$

$$\mathrm{C}_2\colon s = \tilde{\nu} + \rho e^{i\varphi}, \text{ where } 0 \leq \varphi \leq \pi;$$

$$\mathrm{C}_3\colon s = \Omega, \text{ where } \tilde{\nu} + \rho \leq \Omega \leq R;$$

$$\mathrm{C}_4\colon s = Re^{i\varphi}, \text{ where } 0 \leq \varphi \leq \pi.$$

From Cauchy's theorem we have

$$\int_{-R}^{\tilde{\nu}-\rho} \frac{\hat{n}(s)}{s-\tilde{\nu}} ds + \int_{\pi}^{0} \frac{\hat{n}(\tilde{\nu}+\rho e^{i\varphi})}{\tilde{\nu}+\rho e^{i\varphi}-\tilde{\nu}} i\rho e^{i\varphi} d\varphi + \int_{\tilde{\nu}+\rho}^{R} \frac{\hat{n}(s)}{s-\tilde{\nu}} ds + \int_{0}^{\pi} \frac{\hat{n}(Re^{i\varphi})}{Re^{i\varphi}-\tilde{\nu}} iRe^{i\varphi} d\varphi = 0.$$

Therefore

$$\int_{-R}^{\tilde{\nu}-\rho} \frac{\hat{n}(s)}{s-\tilde{\nu}} ds + i\int_{\pi}^{0} \hat{n}(\tilde{\nu}+\rho e^{i\varphi}) d\varphi + \int_{\tilde{\nu}+\rho}^{R} \frac{\hat{n}(s)}{s-\tilde{\nu}} ds + i\int_{0}^{\pi} \frac{\hat{n}(Re^{i\varphi})}{Re^{i\varphi}-\tilde{\nu}} Re^{i\varphi} d\varphi = 0.$$

Putting $\rho \to 0$ and $R \to \infty$, we have

$$\lim_{\substack{\rho\to 0,\\ R\to\infty}} \left(\int_{-R}^{\tilde{\nu}-\rho} \frac{\hat{n}(s)}{s-\tilde{\nu}} ds + \int_{\tilde{\nu}+\rho}^{R} \frac{\hat{n}(s)}{s-\tilde{\nu}} ds + i\int_{\pi}^{0} \hat{n}(\tilde{\nu}+\rho e^{i\varphi}) d\varphi + i\int_{0}^{\pi} \frac{\hat{n}(Re^{i\varphi})}{Re^{i\varphi}-\tilde{\nu}} Re^{i\varphi} d\varphi \right) = 0.$$

$$\lim_{\rho\to 0} \left(\int_{-\infty}^{\tilde{\nu}-\rho} \frac{\hat{n}(s)}{s-\tilde{\nu}} ds + \int_{\tilde{\nu}+\rho}^{+\infty} \frac{\hat{n}(s)}{s-\tilde{\nu}} ds \right) - i\pi\hat{n}(\tilde{\nu}) = 0.$$

We obtain the last line since the fourth term vanishes as $R \to \infty$ if $lim_{|s|\to\infty}\, \hat{n}(s) = 0$.

According to the definition of the Cauchy Principal Value of an integral we get

$$\lim_{\rho\to+0}\left(\int_{-\infty}^{\tilde{\nu}-\rho}\frac{\hat{n}(s)}{s-\tilde{\nu}}ds+\int_{\tilde{\nu}+\rho}^{+\infty}\frac{\hat{n}(s)}{s-\tilde{\nu}}ds\right)=P\int_{-\infty}^{+\infty}\frac{\hat{n}(s)}{s-\tilde{\nu}}ds\,,(1)$$

and therefore, from (1), we have

$$\frac{1}{\pi i}P\int_{-\infty}^{+\infty}\frac{\hat{n}(s)}{s-\tilde{\nu}}ds=\hat{n}(\tilde{\nu}).$$

Taking the real and imaginary parts of $\hat{n}(\tilde{\nu})=n(\tilde{\nu})+in'(\tilde{\nu})$, we get

$$H[n(\tilde{\nu})]=n'(\tilde{\nu})=-\frac{1}{\pi}P\int_{-\infty}^{+\infty}\frac{n(s)}{s-\tilde{\nu}}ds,$$

$$H^{-1}[n'(\tilde{\nu})]=n(\tilde{\nu})=\frac{1}{\pi}P\int_{-\infty}^{+\infty}\frac{n'(s)}{s-\tilde{\nu}}ds\,.$$

The last two expressions are called the *Hilbert transform pair*.

Now we focus on the numerical solution of the Hilbert transform pair.

By means of the Fourier transform definition an analytic signal $x(t)$ can be represented in the following way

$$x(t)=F^{-1}[X(f)]=\int_{-\infty}^{+\infty}X(f)\exp(i2\pi ft)df,$$

where

$$X(f)=F[x(t)]=\int_{-\infty}^{+\infty}x(t)\exp(-i2\pi ft)\,dt.$$

According to the definition of the Hilbert transform for a function $x(t)$ we have

$$\tilde{x}(t)=H\big(x(t)\big)=\frac{1}{\pi}P\int_{-\infty}^{+\infty}\frac{x(s)}{s-t}ds=\frac{1}{\pi t}*x(t).$$

Since

$$F\left(\frac{1}{\pi t}\right)=-i[sign(f)]=\begin{cases}-i, & f>0;\\ 0, & f=0;\\ i, & f<0.\end{cases}$$

Then

$$F[\tilde{x}(t)]=\tilde{X}(f)=F\left[\frac{1}{\pi t}*x(t)\right]=F[x(t)]\cdot F\left(\frac{1}{\pi t}\right)=-i[sign(f)]\cdot X(f).$$

Multiplying both parts of the last expression by $i[sign(f)]$, we get

$$i[sign(f)]\cdot\tilde{X}(f)=X(f).$$

This equation is used for the numerical solution of the Hilbert transform pair. The algorithm of calculating the Hilbert transform consists of three steps:

1) to calculate the spectrum $X(f)$:

$$X(f) = F[x(t)];$$

2) to apply the expression $i[sign(f)] \cdot \tilde{X}(f) = X(f)$ for calculating $\tilde{X}(f)$:

$$\tilde{X}(f) = -i[sign(f)] \cdot X(f)$$

3) to calculate $\tilde{x}(t)$:

$$\tilde{x}(t) = F^{-1}[\tilde{X}(f)].$$

These three steps are used in Matlab for the calculation of Eqs.B2 via the function *Hilbert*.

More details about numerical solution of the Kramers-Kronig transforms based on the FFT algorithm are presented in P.Bruzzoni et[19].

References

1. D. Wetzel and J. Reffner, *Cereal Foods World*, 1993, **38**, 9-20.
2. B. Mohlenhoff, M. Romeo, M. Diem and B. R. Woody, *Biophys J*, 2005, **88**, 3635-3640.
3. G. Mie, *Ann. Phys.(Leipzig)*, 1908, **25**, 377-452.
4. H. C. Van De Hulst, *Light scattering by small particles*, Courier Corporation, 1957.
5. P. Bassan, H. J. Byrne, F. Bonnier, J. Lee, P. Dumas and P. Gardner, *Analyst*, 2009, **134**, 1586-1593.
6. P. Bassan, A. Kohler, H. Martens, J. Lee, H. J. Byrne, P. Dumas, E. Gazi, M. Brown, N. Clarke and P. Gardner, *Analyst*, 2010, **135**, 268-277.
7. P. Bassan, A. Kohler, H. Martens, J. Lee, E. Jackson, N. Lockyer, P. Dumas, M. Brown, N. Clarke and P. Gardner, *Journal of Biophotonics*, 2010, **3**, 609-620.
8. P. Bassan, A. Sachdeva, A. Kohler, C. Hughes, A. Henderson, J. Boyle, J. H. Shanks, M. Brown, N. W. Clarke and P. Gardner, *Analyst*, 2012, **137**, 1370-1377.
9. A. Kohler, J. Sulé-Suso, G. D. Sockalingum, M. Tobin, F. Bahrami, Y. Yang, J. Pijanka, P. Dumas, M. Cotte, D. G. van Pittius, G. Parkes and H. Martens, *Appl Spectrosc*, 2008, **62**, 259-266.
10. T. van Dijk, D. Mayerich, P. S. Carney and R. Bhargava, *Appl Spectrosc*, 2013, **67**, 546-552.
11. R. Lukacs, R. Blümel, B. Zimmerman, M. Bağcıoğlu and A. Kohler, *Analyst*, 2015, **140**, 3273-3284.
12. L. D. Landau, J. Bell, M. Kearsley, L. Pitaevskii, E. Lifshitz and J. Sykes, *Electrodynamics of continuous media*, Elsevier, 1984.
13. J. L. Ilari, H. Martens and T. Isaksson, *Appl Spectrosc*, 1988, **42**, 722-728.
14. H. Martens and E. Stark, *Journal of Pharmaceutical and Biomedical Analysis*, 1991, **9**, 625-635.
15. A. Kohler, C. Kirschner, A. Oust and H. Martens, *Appl Spectrosc*, 2005, **59**, 707-716.
16. N. K. Afseth and A. Kohler, *Chemometrics and Intelligent Laboratory Systems*, 2012, **117**, 92-99.
17. B. Zimmermann and A. Kohler, *Appl Spectrosc*, 2013, **67**, 892-902.
18. R. P. Kanwal, *Linear integral equations*, Springer Science & Business Media, 2013.
19. P. Bruzzoni, R. Carranza, J. C. Lacoste and E. Crespo, *Electrochimica acta*, 2002, **48**, 341-347.

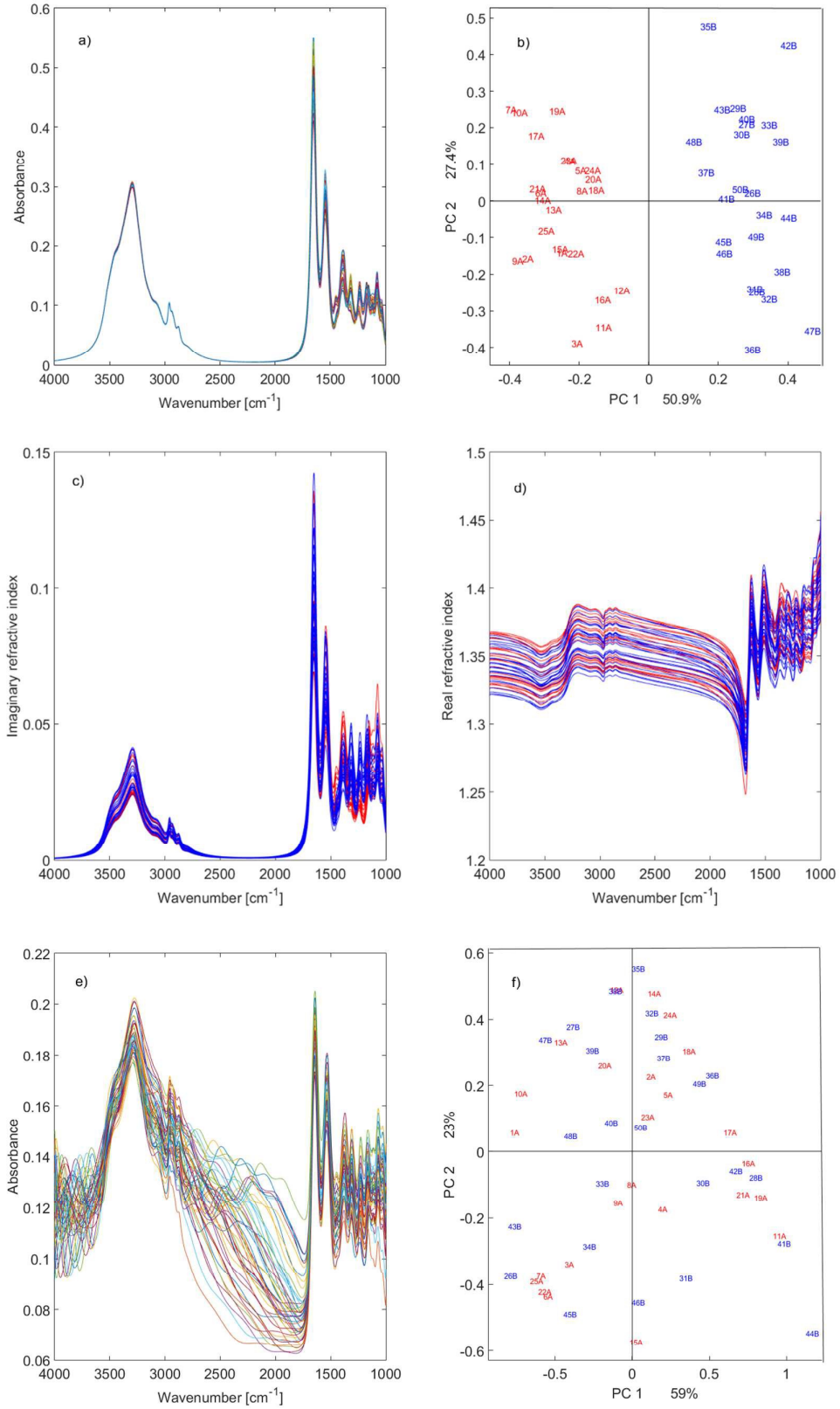


Figure 1. (a) A set of simulated pure absorbance spectra. (b) PCA of simulated pure absorbance spectra. (c) Imaginary part of the refractive index obtained from the simulated absorbance spectra (d) Real part of the refractive index obtained from the simulated absorbance spectra. (e) Apparent absorbance spectra obtained by employing exact Mie theory taking into account the optical setup of an infrared microscope. (f) PCA of simulated apparent absorbance spectra.

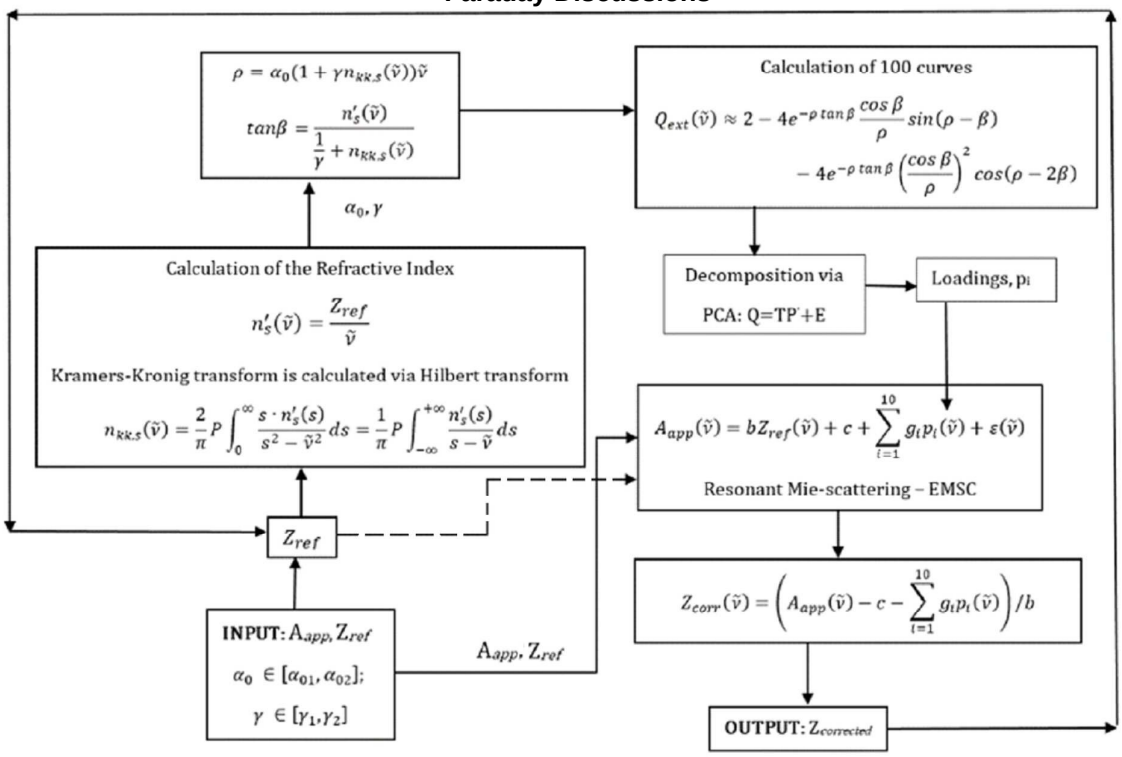


Figure 2: Schematic illustration of the iterative algorithm for the retrieval of the pure absorbance spectrum from the apparent absorbance spectrum. After each iteration, a new estimate of the pure absorbance spectrum is obtained. The next iteration is initialized by using the estimate of the pure absorbance spectrum for updating the real and imaginary parts of refractive index for the iterative algorithm.

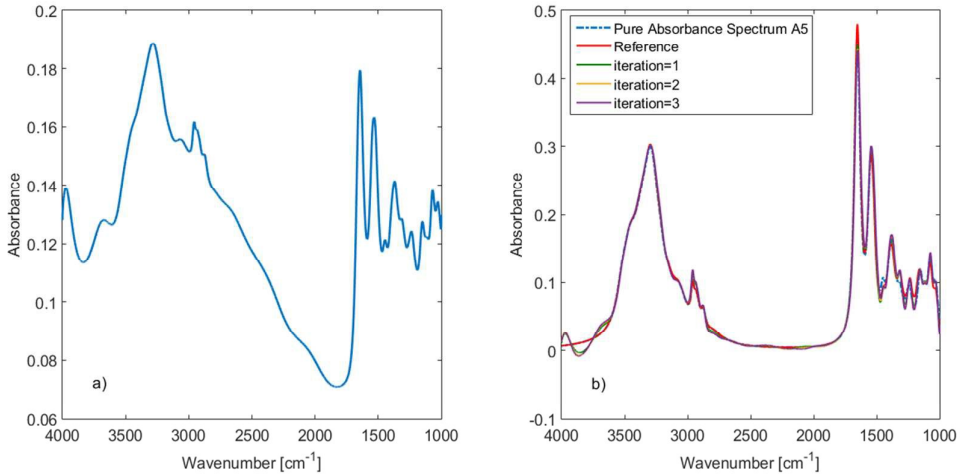


Figure 3. (a) An example of a simulated apparent absorbance spectrum is shown. (b) Spectra corrected by an iterative EMSC meta-model employing the Van De Hulst approximate formula with complex refractive index are shown. Corrected spectra for iteration 1, 2 and 3 of the correction algorithm (green, brown and violet, respectively) are close to the underlying pure absorbance spectrum (blue dashed line) and different from the reference spectrum (red line) employed. For simulation of the apparent absorbance spectrum in (a) we used the parameters $n_0 = 1.35$ and $a = 3.81\ \mu m$.

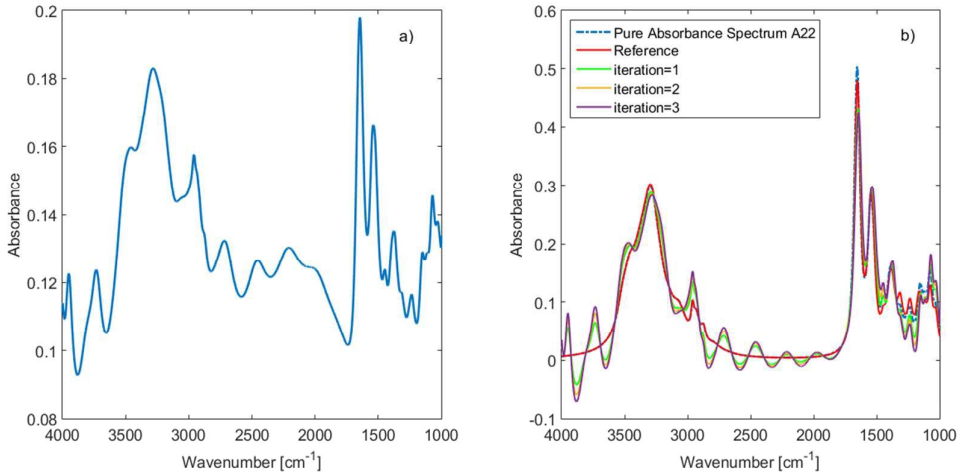


Figure 4. (a) Apparent absorbance spectrum containing ripples. For the simulation of the apparent absorbance spectrum the parameters $n_0 = 1.36$ and $a = 5.12\ \mu m$ were used. (b) The corrected spectra for iteration 1, 2 and 3 of the correction algorithm (green, brown and violet, respectively). They still contain the ripple structure. As expected, the EMSC meta-model employing the Van De Hulst approximate formula with complex refractive index does not correct the ripples.

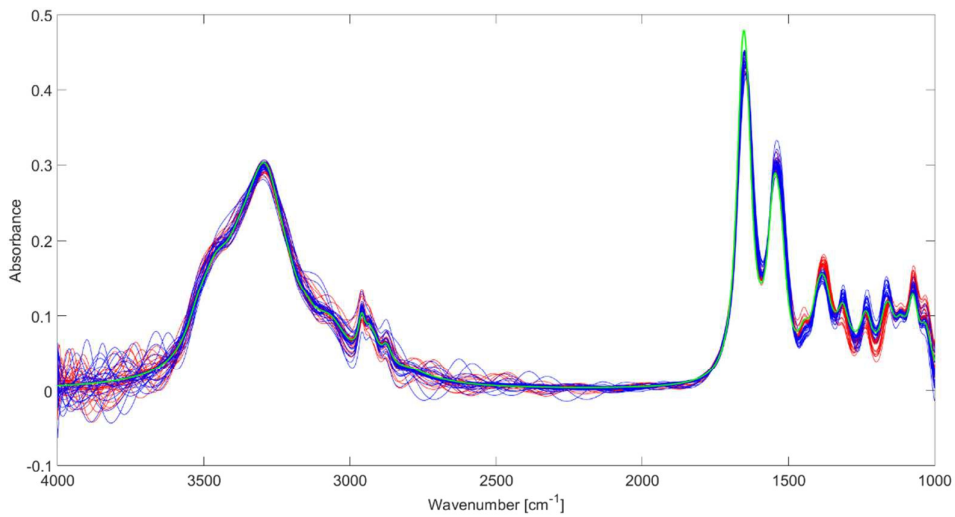


Figure 5. The complete set of simulated apparent absorbance spectra shown in Fig.1c is corrected by the algorithm presented in this paper. The spectra that correspond to red scores of pure absorbance spectra in Fig 1b are drawn in red; the spectra that correspond to blue scores of pure absorbance spectra in Fig. 1b are marked blue.The reference spectrum is shown in green. Some corrected absorbance spectra contain ripples. As expected, the EMSC meta-model employing the Van de Hulst formula with complex refractive index does not remove ripples.

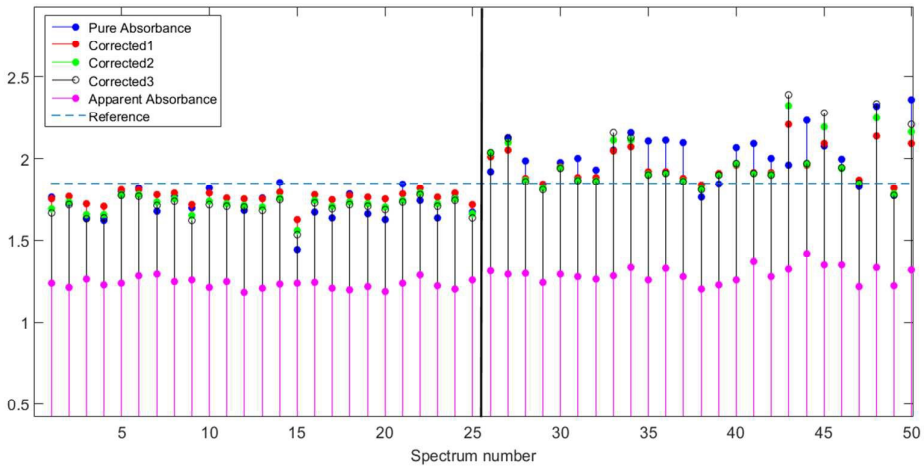


Figure 6: The peak ratio is shown for the bands at 1546 cm^{-1} and 1387 cm^{-1} for the simulated pure absorbance spectra (blue), for iteration 1, 2 and 3 of the correction algorithm (red, green and black, respectively) and for the apparent absorbance (magenta). The ratio for the reference spectra is plotted as the blue dashed line for comparison.

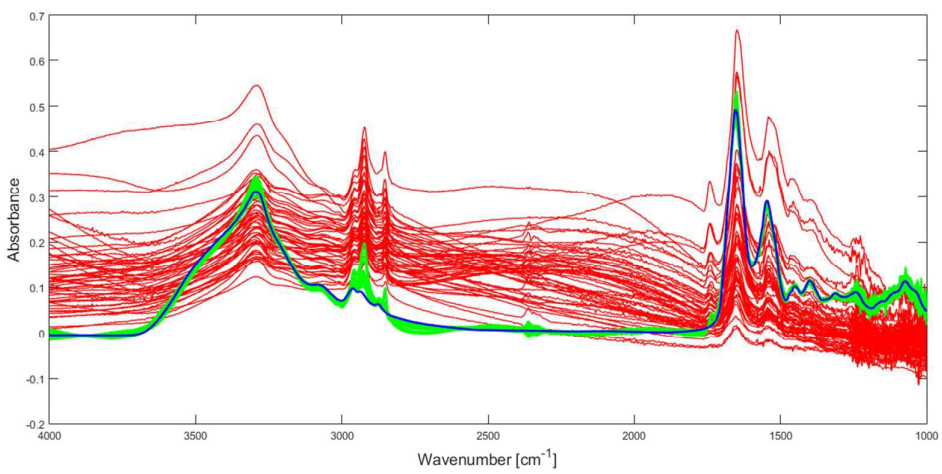


Figure 7. The measured infrared microspectroscopy spectra of lung cancer cells are shown in red together with the corrected spectra in green. For the correction one iteration step was used. As a reference spectrum (in blue) the matrigel spectrum was used.

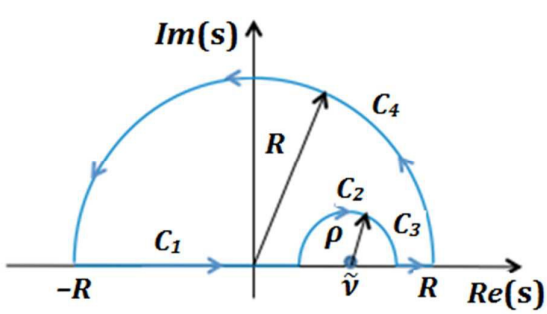


Figure 8. Contour in the complex plane used to derive the Kramers-Kronig relations. The integration path, which is skirting the singularity point $s = \tilde{\nu}$, is indicated by arrows. The radius R of the outer semicircle tends to infinity. The radius ρ of the small semicircle about $\tilde{\nu}$ tends to zero.