






RESEARCH ARTICLE

An automated approach for fringe frequency estimation and removal in infrared spectroscopy and hyperspectral imaging of biological samples

Johanne Heitmann Solheim^{1,2*}  | Ferenc Borondics²  |
Boris Zimmermann¹  | Christophe Sandt²  | Florian Muthreich³  |
Achim Kohler¹

¹Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway

²Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin-BP48, Gif-sur-Yvette CEDEX, France

³Department of Biological Sciences, University of Bergen, Bergen, Norway

*Correspondence

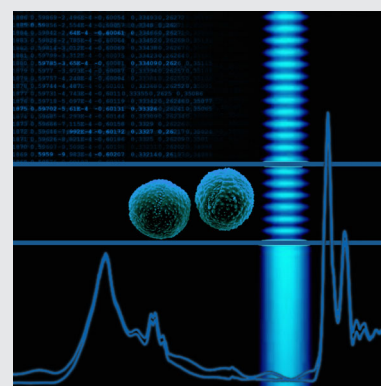
Johanne Heitmann Solheim, Faculty of Science and Technology, Norwegian University of Life Sciences, Drøbakveien 31, 1432 Ås, Norway.
Email: johanne.heitmann.solheim@mbu.u.no

Funding information

PollChem project, Grant/Award Number: 249844; DeepHyperSpec project, Grant/Award Number: 289518

Abstract

In infrared spectroscopy of thin film samples, interference introduces distortions in spectra, commonly referred to as fringes. Fringes may alter absorbance peak ratios, which hampers the spectral analysis. We have previously introduced *extended multiplicative signal correction* (EMSC) for fringes correction. In the current article, we provide a robust open-source algorithm for fringe correction in infrared spectroscopy and propose several improvements to the Fringe EMSC model. The suggested algorithm achieves a more precise fringe frequency estimation by mean centering of the measured spectrum and applying a window function prior to the Fourier transform. It selects two frequencies from a user defined number of maxima in the Fourier domain. The improved Fringe EMSC algorithm is validated on two experimental datasets, one of them being a hyperspectral image. Techniques for separating sample spectra from background spectra in hyperspectral images, and techniques to identify spectra affected by fringes are also provided.



KEYWORDS

extended multiplicative signal correction, fringes, infrared spectroscopy, preprocessing

1 | INTRODUCTION

Fourier-transform infrared spectroscopy (FTIR) is one of the most widely used analytical techniques. The

technique has broad application in natural sciences, industrial research and development, and medical diagnostics. Examples includes material characterization and analysis, quality control, process monitoring, and

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Journal of Biophotonics* published by Wiley-VCH GmbH.

identification and classification of various organic, inorganic and biological samples [1–4].

In infrared spectroscopy, one of the main objectives is to acquire chemical information about a sample. In a transmission experiment, this is achieved by illuminating a sample with infrared radiation and recording radiation that is transmitted through the sample to a detector. By taking the intensity ratio of the incident radiation and the intensity of the transmitted radiation, the unit-free quantity transmittance is calculated. From the transmittance we obtain the so-called absorbance spectrum, which as the name implies is expected to reveal signatures from loss of radiation through absorption. However, in any real measurement, radiation is always lost due to scattering off the sample as well. Scattering comprises all effects that lead to deviations from rectilinear wave propagation, including reflection and diffraction. In some cases the loss due to scattering is considerable and leads to strong spectral distortions.

In an ideal case, one can assume that the interaction between the infrared radiation with matter follows Beer–Lambert behavior in which case the absorbance is proportional to the optical thickness of the sample [5–7]. A type of absorbance spectrum that contains only signatures of chemical absorbance, and is devoid of physical features such as scattering, we call a *pure absorbance spectrum*. In contrast, spectra which are affected by physical phenomena we call *apparent absorbance spectra*. Variation in source intensity or in sample thickness, which lead to constant baseline shifts and scaling of the spectra, respectively, can be handled using the standard extended multiplicative signal correction (EMSC) [8]. EMSC is based on the Beer–Lambert model, and can be extended to handle moderate scattering. With moderate scattering we refer to situations where scattering effects do not strongly deform absorption bands and absorption bands do not strongly affect scattering properties of the sample [7]. With EMSC, the chemical and physical features in the absorbance spectra are estimated and separated, and the pure absorbance spectrum is retrieved.

For highly sophisticated scattering signatures, the standard EMSC is not sufficient. When the size of the sample is on the same order as the wavelength of the employed infrared radiation, complex non-Beer–Lambert behavior is readily observed. For these samples, the wave nature of the radiation dominates, leading to intricate relations between scattering and absorption. An example is Mie scattering, the impact of which on infrared microspectroscopy has been studied extensively since Mohlenhoff et al. described the phenomenon in 2005 [9–14].

Another example of strong spectral distortions caused by scattering is the case of thin film samples. For thin films, with two nearly perfectly parallel surfaces, multiple internal reflections inside the film lead to signatures of

constructive and destructive interference. Films which give rise to multiple internal reflections are in photonics also referred to as etalons. While multiple internal reflection is a special case of scattering, we will also refer to it as reflection. This effect can for example be observed for thin oil films on top of water displaying a whole spectrum of colors due to constructive and destructive interference of the incoming sun light. In the spectral domain, the constructive and destructive interference manifests itself in the absorbance spectra as sine wave signatures. The signatures are commonly referred to as *interference fringes*.

Fringes are observed in the infrared spectra of many different samples, such as thin tissue sections [7] and paraffin embeddings [15], as well as in measurements with for example microfluidic chambers [16] or diamond anvil cells used in high pressure measurements [17]. Different techniques have been proposed to deal with interference fringes in spectroscopy. The methods can be divided into three categories; (1) eliminating multiple internal reflections during the measurements, (2) modeling and removing the fringes by a wave optics approach, and (3) suppressing the fringes in the interferogram or the spectrum without a physical model. Modeling and removing the fringes according to a physical model, as opposed to eliminating interference during the measurements, is advantageous since the fringe problem is often discovered after the experiment has been conducted.

Approaches that attempt to remove fringes without a physical model, the third category, include altering the interferogram, fitting and removing sine waves from the spectra, SVD denoising and deconvolution. In FTIR measurements, the sinusoidal fringes appear in the interferogram as spikes, located outside of the central peak. Both zeroing out these spikes, or replacing them with a patch from an interferogram belonging to a fringe-free spectrum, have been proposed [18,19]. However, both of these methods can lead to modifications of absorbance peaks in the spectra [7], called aliasing effects. In 1978, Clark and Moffatt suggested to remove interference fringes by manually fitting sine waves, by changing the frequency, amplitude, phase and offset, until the fringes were minimized [20]. This method is tedious and inapplicable to large datasets where the parameters vary between spectra. Iwata et al. suggested in 1994 a method for removing noise from spectra by use of a singular value decomposition (SVD) on a cyclic matrix constructed from the interferogram [21]. This method can be used to remove interference fringes, however, it requires the fringe signal to be of relatively high frequency and low amplitude, such that it resembles noise. Further, the method is inapplicable to absorbance spectra from biological samples, as it fails when several absorbance bands are present in one spectrum. A deconvolution method

has also been proposed by Pistorius et al. [22] However, this method may also lead to aliasing effects. Approaches based on wave optics models, the second category, includes the method developed by King and Milosevic in 2012, which aims at retrieving the absorption coefficient of a sample, provided that the refractive index is known at one wavelength and the thickness is approximately known [23,24]. The method has proven useful for correcting FTIR transmission spectra from non-biological material [25]. Further, Mayerhöfer et al. developed in 2020 an iterative algorithm which aims to retrieve the refractive index spectrum and the thickness of the sample in a wave optics correction scheme. They show that they are able to correct for fringes, as well as the dispersive effect, in measured spectra of polymethyl methacrylate films [26].

A promising method proposed recently is to fit and remove sinusoidal waves in a model-based, automated manner. Konevskikh et al. suggested in 2015 a model based method and showed that interference fringes, as predicted by the Fresnel equations, can be approximated to a single sinusoid, for samples with refractive index below 2.4 [7]. In his method, sine waves are fitted to the measured spectra in an extended multiplicative signal correction (EMSC) model. The method is similar to the manual fitting procedure of Clark and Moffatt, but the frequency, amplitude, phase and offset are all estimated automatically. The frequency of the sine waves is found through a Fourier transform of the fringe signal, while adding both sine and cosine curves to the EMSC model allows an automated estimation of the phase. In EMSC, the amplitude and constant offset are estimated as well. The method, we call it Fringe EMSC, was revised in 2018 by Azafar et al., with the suggestion to make the fringe removal an iterative process, as well as including sine waves of multiple frequencies as model functions in the EMSC [16]. The EMSC based method has the advantage that it can correct many spectra in a fast and robust manner. This is one of the reasons why EMSC based methods are often preferred to iterative methods.

Here, we propose a robust and automated approach for precise estimation fringe frequencies, which is needed as an input for the EMSC correction, and provide it as an open source code for the infrared community. The code is available at the BioSpec Norway GitHub page, as a part of the biospectools toolbox (<https://github.com/BioSpecNorway/biospectools>). The proposed Fringe EMSC is validated by simulated data and two experimentally obtained datasets, one from pollen samples embedded in a polyethylene and paraffin film, and one from thin hair cross sections. The developed code can be used for infrared spectra and infrared hyperspectral images. We demonstrate how the Fringe EMSC can be used on a hyperspectral image. This procedure includes steps to

separate sample spectra from the image background, and to identify spectra which need fringe correction from the ones which do not.

2 | EXPERIMENTAL DATA

2.1 | Pollen grains dataset

The pollen dataset consist of single spectra of pollen grains from four Oak species, collected in Australia and Portugal in the time period 2017–2018. The pollen samples were embedded, without any sample pretreatment, in a matrix medium, consisting of a soft paraffin (Vaseline-type petroleum jelly) between two sheets of polyethylene foils. The amount of paraffin was approximately 1 mg per 25 cm². An X-ray fluorescence (XRF) sample cup was used to stretch the foils and to minimize the appearance of air bubbles in the sample [27].

Microscopic transmission measurements of pollen were performed using a Vertex 70 FTIR spectrometer with a Hyperion 3000 IR microscope (Bruker Optik, Ettlingen, Germany), equipped with a globar as mid-IR source and a liquid nitrogen-cooled mercury cadmium telluride (MCT) detector. The spectra were recorded with a total of 128 scans in the 7000–600 cm⁻¹ spectral range, with a spectral resolution of 2 cm⁻¹, and digital spacing of 0.9630 cm⁻¹. The samples were measured using a ×15 objective, with a 30 × 30 μm² aperture. Background spectra were recorded at the start of each measurement (one per sample) by measuring a sample-free setting. The microscope was equipped with a computer-controlled x/y/z stage. The spectroscopic system was controlled with OPUS 7.5 software (Bruker Optik, Ettlingen, Germany). 50 spectra per sample were obtained, each corresponding to different single pollen grains, resulting in 200 spectra in total. Only a selection of these 200 spectra was used in this study.

2.2 | Hair cross section dataset

The hair dataset consists of a hyperspectral image of thin cross sections of human hair. The samples were prepared by inserting the hair in a plastic tubing of pharmaceutical grade, filled with water. Subsequently, the samples were frozen and cryosectioned with a cryotome at negative temperature (−5 to −10°C). The thicknesses of the samples were approximately 10 μm. The sections were then deposited on a slide which is transparent in the infrared.

The infrared image was collected in 2016 at the Synchrotron SOLEIL, using a synchrotron source on a Continuum XL microscope (ThermoFisher Scientific, Courtaboeuf, France) with a MCT detector. The spectral resolution was

8 cm^{-1} and the number of averaged scans was 50 for the sample and 512 for the background. The aperture size was set to $6 \times 6 \text{ }\mu\text{m}^2$ in single aperture mode.

3 | THEORETICAL MODEL

3.1 | The electromagnetic theory of fringes

In infrared spectroscopy, the aim is to obtain a scatter-free signal, the pure absorbance spectrum. The pure absorbance spectrum allows us to estimate concentration of constituents via absorption bands that are proportional to the concentrations. The absorbance spectrum $Z(\tilde{\nu})$ can be obtained from measured intensities as

$$Z(\tilde{\nu}) = -\log_{10} \frac{I(\tilde{\nu})}{I_0(\tilde{\nu})} = -\log_{10} T(\tilde{\nu}) \quad (1)$$

where $I_0(\tilde{\nu})$ refers to the incident radiation, $I(\tilde{\nu})$ the transmitted radiation, and $T(\tilde{\nu})$ is the transmittance.

According to the Beer–Lambert law, a pure absorbance spectrum can be approximated as the product of the characteristic absorptivity $e_j(\tilde{\nu})$ of an absorbing species, j , the concentration, c_j , of the absorbing species and the optical thickness, d , of the sample, summed over all absorbing species, J [28,29];

$$Z_{\text{pure}}(\tilde{\nu}) \approx \left[\sum_{j=1}^J c_j \cdot e_j(\tilde{\nu}) \right] \cdot d. \quad (2)$$

From this equation, the absorbance spectrum is considered to be proportional to the optical thickness of the sample, d .

For the ideal case, when the sample does not scatter, and the expression above is valid, the absorbance signals are solely related to absorption. In many measurements, however, some of the incident radiation is scattered off the sample. In the absorbance spectrum, scattering may lead to broad, underlying baseline distortions as well as distortions in chemical bands. When the measured spectrum contains scatter features, we refer to it as the apparent absorbance, $Z_{\text{app}}(\tilde{\nu})$. In many cases scattering features can be estimated by standard extended multiplicative signal correction (EMSC). EMSC is a model based preprocessing technique which was developed first as multiplicative signal correction (MSC) by Martens et al. in the 80's [30–32], and later as the extended version called EMSC [8]. EMSC was shown to be an excellent tool for correcting infrared absorbance spectra containing moderate scattering features [29]. It is based on the aforementioned Beer–Lambert model for absorption, where

changes in the optical thickness of a sample causes scaling of the absorbance spectra. Physical effects are treated as additive contributions, and accounted for by different so-called model spectra. The EMSC model has been developed and combined with meta-modeling of scattering phenomena to handle sophisticated physical phenomena, such as Mie scattering [10,11,13,14] and multiple internal reflection [7,16].

Interference fringes, arising from multiple internal reflections, are apparent in absorbance spectra as sine waves, and are especially noticeable in the “silent” regions where they do not overlap with chemical absorbance signals. At wavelengths corresponding to destructive interference, a peak is found in the spectrum, while constructive interference corresponds to the troughs. A typical example of an infrared spectrum containing fringes is shown in black in Figure 1. Figure 1 also shows an example of a so-called pure absorbance spectrum, that is, a spectrum without physical signatures, plotted in gray.

The interference pattern can be described using the Fresnel equations. We start out by expressing the transmitted intensity

$$I(\tilde{\nu}) = I_0(\tilde{\nu}) \cdot |t(\tilde{\nu})|^2 \quad (3)$$

where $t(\tilde{\nu})$ is the transmission coefficient at normal incidence for a film in vacuum, given by

$$t(\tilde{\nu}) = \frac{2ime^{-2\pi\tilde{\nu}d}}{(1+m^2)\sin(2\pi m\tilde{\nu}d) + 2im\cos(2\pi m\tilde{\nu}d)} \quad (4)$$

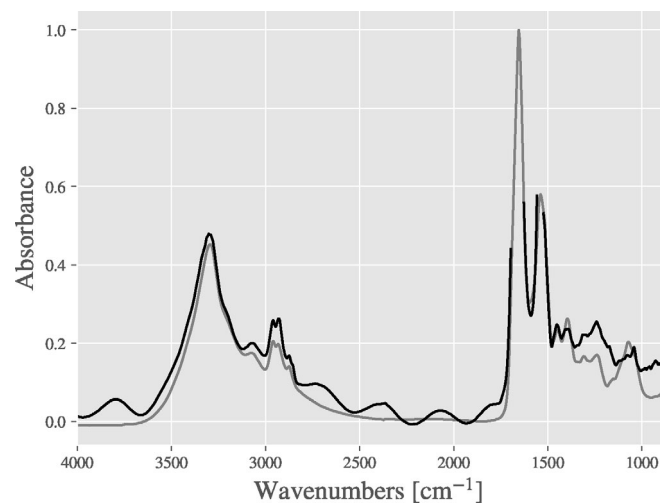


FIGURE 1 Example of a pure absorbance spectrum in gray (Matrigel [14]), and an apparent absorbance spectrum with interference fringes in black (from a thin film hair cross section). The amide I and II peaks are cut in the hair cross section spectrum due to over saturation in the measurements. Interference fringes are clearly present in the silent region

where m is the complex refractive index, d is the thickness of the film and $\tilde{\nu}$ is the wavenumber. The complex refractive index consists of a real part n , which relates to the wave propagation dynamics of the material, and an imaginary part n' , which relates to the dissipation of energy. The real and imaginary part of the refractive index are related by

$$n(\tilde{\nu}) = n_0 + \frac{2}{\pi} \text{P} \int_0^{\infty} \frac{sn'(s)}{s^2 - \tilde{\nu}^2} ds \quad (5)$$

where n_0 is the constant (or slowly varying) offset of the real part of the refractive index, and the second term constitutes the fluctuating part, given by the Kramers-Kronig relation. P is the Cauchy principal value of the integral. We denote the fluctuating part of the real part of the refractive index n_{kk} .

It can be shown that the imaginary part of the refractive index is related to the pure absorbance spectrum [33] by

$$n'(\tilde{\nu}) = \frac{Z_{\text{pure}}(\tilde{\nu}) \ln 10}{4\pi d\tilde{\nu}}, \quad (6)$$

assuming that the reflection of the incident radiation by the film is negligible, such that the Beer-Lambert law holds. For further details we refer the reader to references [5–7]:

3.1.1 | Dispersion

The real part of the refractive index may undergo two types of dispersion in a spectral region. The first is a slow dispersion, which is determined by the overall absorption properties of the material outside the spectral region. This type of dispersion is well known from the refraction of visible electromagnetic radiation at a prism. Dispersion can occur as well as a rapid fluctuation of the real part of the refractive index in the neighborhood of absorption bands. The rapid fluctuations of the real part of the refractive index is related to absorption and is governed by the Kramers-Kronig relation.

The wavelength-dependent real part and the imaginary part of the refractive index determine together the absorption efficiency of the material [34], and may cause band shape changes in infrared microspectroscopy of cells and tissues, which have been interpreted as Mie scattering [35]. Similar phenomena are observed for thin films of materials with a high refractive index [36]. Theoretically, this causal connection becomes evident from the fact that both n and n' enter Equation (4) through m . Band shape changes in absorbance spectra which result from the dispersive real part of the refractive index are

known in the Mie literature as the dispersive effect. Initially, the effect was named “the dispersive artefact” [11,35]. The effect is however completely described by theory, and is therefore not an artifact. Therefore, the literature terms this effect today “dispersive effect” [14].

While the dispersive effect can be seen explicitly in the analytical description of Mie scattering, and in the analytical description of the interaction of electromagnetic radiation with a thin film, it is in principle relevant for any scatterer of arbitrary shape and composition.

For the interested reader, we refer to references [11,14,37,38] to see examples of how dispersion affects Mie distorted spectra. Examples of the dispersive effect in films will be briefly discussed later in this article.

While the dispersive effect is especially strong in Mie distorted spectra, it is not as prominent in spectra from thin film biological samples. In the article at hand it is shown that the multiple internal reflection of thin biological films can be treated in a good approximation as additive to absorption. This permits the use of the Fringe EMSC model for fringe correction, as opposed to using an iterative approach. EMSC models are typically very stable and both easy and fast to use for correction of many spectra.

3.2 | The Fringe EMSC

In the following, we present the Fringe EMSC, in which we model the sample as a non-absorbing film, and treat the fringe signal as additive to the pure absorbance. When approximating the fringe signal emerging from a non-absorbing film, the apparent absorbance of the film can be expressed as [7]

$$A = \frac{\ln \left(1 + \frac{1}{4} \left(\frac{1}{n_0} - n_0 \right)^2 \sin^2(2\pi n_0 \tilde{\nu} d) \right)}{\ln 10}. \quad (7)$$

This equation is later in the article referred to as the full analytical model. By means of a Taylor expansion of the natural logarithm and the half-angle formula, the expression can be written

$$A = \frac{1}{4 \ln 10} \left(\left(\frac{1}{n_0} - n_0 \right)^2 \frac{1 - \cos(4\pi n_0 \tilde{\nu} d)}{2} \right) - \frac{1}{32 \ln 10} \left(\left(\frac{1}{n_0} - n_0 \right)^4 \frac{(1 - \cos(4\pi n_0 \tilde{\nu} d))^2}{4} \right) + \dots \quad (8)$$

provided that $\frac{1}{4} \left(\frac{1}{n_0} - n_0 \right)^2$ is between -1 and 1 , which holds when $n_0 \leq 2.4$ [7]. From this equation, we can see

that the fringe signal can be approximated by a single cosine wave (or sine wave, depending on the phase) when the first term dominates the expression. One should note that the frequency of the fringe signal is depending on the ratio of the optical thickness of the film to the wavelength of the employed radiation, and is therefore independent of a potential substrate. A substrate would however affect the amplitude of the fringes. In most cases, fringes are unwanted in infrared spectroscopy, since they alter peak ratios and hampers the analysis. However, fringes can be used to estimate the thickness of a sample of known refractive index, using the equation above.

In the Fringe EMSC, the apparent absorbance is modeled as

$$Z_{app}(\tilde{\nu}) = a + b \cdot Z_{ref}(\tilde{\nu}) + c \cdot \tilde{\nu} + Z_{fringes}(\tilde{\nu}) + \varepsilon(\tilde{\nu}) \quad (9)$$

where the terms a and $c \cdot \tilde{\nu}$ represent constant and linear baseline shifts, respectively. The model can be further extended by adding higher order polynomials. $Z_{ref}(\tilde{\nu})$ is a pure, scatter-free absorbance spectrum, called the reference spectrum, and $Z_{fringes}(\tilde{\nu})$ is the fringe signal. The term $\varepsilon(\tilde{\nu})$ accounts for the unmodelled part of $Z_{app}(\tilde{\nu})$. The Fringe EMSC builds on a standard EMSC, that is, an EMSC model which does not contain the term $Z_{fringes}(\tilde{\nu})$ [8]. In a standard EMSC, polynomials are added as model spectra to account for moderate baseline variations.

The fringe signal $Z_{fringes}(\tilde{\nu})$ we suggest to present by a number of sine and cosine waves with different frequencies

$$\begin{aligned} Z_{fringes}(\tilde{\nu}) = & d_1 \sin(f_1 \cdot \tilde{\nu}) + d_2 \cos(f_1 \cdot \tilde{\nu}) \\ & + e_1 \sin(f_2 \cdot \tilde{\nu}) + e_2 \cos(f_2 \cdot \tilde{\nu}) \\ & + \dots \end{aligned} \quad (10)$$

The EMSC parameters a , b , c , d_1 etc., are estimated by least squares regression, as will be explained shortly. With Equation (9), the measured spectrum is modeled as a combination of physical and chemical model functions. When the contribution from each of the model functions is estimated, the physical features of the absorbance spectrum can be removed, and we are left with the corrected spectrum $Z_{corr}(\tilde{\nu})$

$$Z_{corr}(\tilde{\nu}) = \frac{Z_{app}(\tilde{\nu}) - a - c \cdot \tilde{\nu} - Z_{fringe}(\tilde{\nu})}{b}. \quad (11)$$

Dividing by the scaling parameter b normalizes all spectra to match the optical thickness of the reference spectrum.

As mentioned, the EMSC parameters need to be estimated, which is done by least squares regression, as described in the following. Let the apparent

absorbance spectrum be represented by the column vector α , where the length i is the number of wavenumber channels,

$$\alpha = \begin{pmatrix} \alpha(\tilde{\nu}_1) \\ \alpha(\tilde{\nu}_2) \\ \alpha(\tilde{\nu}_3) \\ \vdots \end{pmatrix}. \quad (12)$$

The matrix containing the j model spectra are represented as the $[i \times j]$ matrix \mathbf{M} ,

$$\mathbf{M} = \begin{bmatrix} 1 & Z_{ref}(\tilde{\nu}_1) & \tilde{\nu}_1 & \sin(f_1 \cdot \tilde{\nu}_1) & \cos(f_1 \cdot \tilde{\nu}_1) & \dots \\ 1 & Z_{ref}(\tilde{\nu}_2) & \tilde{\nu}_2 & \sin(f_1 \cdot \tilde{\nu}_2) & \cos(f_1 \cdot \tilde{\nu}_2) & \\ 1 & Z_{ref}(\tilde{\nu}_3) & \tilde{\nu}_3 & \sin(f_1 \cdot \tilde{\nu}_3) & \cos(f_1 \cdot \tilde{\nu}_3) & \\ \vdots & & & & & \end{bmatrix} \quad (13)$$

Finally, the vector \mathbf{p} of length j which contains the EMSC parameters,

$$\mathbf{p} = \begin{pmatrix} a \\ b \\ c \\ \vdots \end{pmatrix} \quad (14)$$

can be found by regressing each apparent absorbance spectrum α onto the matrix of model spectra \mathbf{M} . The vector of EMSC parameters \mathbf{p} contains the regression coefficients obtained from the regression model for one apparent absorbance spectrum.

It is important to note that the EMSC needs as an input a very rough estimate of the pure absorbance spectrum as a reference spectrum. This estimate is preferably an absorbance spectrum from the same dataset that shows little or no scatter contributions. The aim of the Fringe EMSC is not to restore iteratively the pure absorbance spectrum. It aims at estimating the fringes, scaling and baseline effects and to subtract these from the measure absorbance spectrum which then results in a very good estimate of the pure absorbance spectrum. There is no requirement that the reference spectrum is chemically nearly identical with the underlying pure absorbance spectrum, chemical difference between the reference spectrum and the spectrum to be corrected can be handled by the residual term of the EMSC model. It has been shown previously that this approach has a stabilizing effect for the parameter estimation, as the reference spectrum has overall a very similar shape as the underlying pure absorbance spectrum except the tiny and interesting chemical variations which are taken care of by the

residual of the EMSC model [39,40]. Therefore, the Fringe EMSC approach does not require the knowledge of the refractive index function. The knowledge of the refractive index function is necessary if the dispersive effect becomes large and effects are no longer additive or multiplicative, as for Mie scattering, and for thin films of non-biological materials. In this case iterative algorithms need to be used [12–14,26,37].

3.3 | Frequency estimation

In the following we explain how the fringe frequencies f_i can be estimated automatically. We select a part of the spectrum $Z_{crop}(\tilde{\nu})$, preferably from the “silent” region where the fringe signal is most undisturbed by absorbance peaks, and perform a discrete Fourier transformation (DFT) on this region:

$$\hat{Z}_{crop}(k\Delta x) = \sum_{n=0}^{N-1} Z_{crop}(\tilde{\nu}_n) e^{-i\frac{2\pi}{N}kn}, \quad (15)$$

where $\hat{Z}_{crop}(k\Delta x)$ is the complex number associated with the k 'th frequency bin, N is the number of sampling points in the signal, and Δx denotes the distance between the sampling points in the Fourier domain. The amplitude of the k 'th frequency bin is found as $|\hat{Z}_{crop}(k\Delta x)|$, and the frequency steps in the Fourier domain are found through Δx as

$$\Delta x = \frac{2\pi}{N\Delta\tilde{\nu}}. \quad (16)$$

It is evident that the frequency steps in the Fourier domain can be decreased, and the number of frequencies increased, by increasing N . Zero-padding is therefore applied to $Z_{crop}(\tilde{\nu})$, extending the signal with zeros at the end. Padding with $n_{pad} = 5 \cdot N$ at the end of the signal is chosen as a standard. When the cropped signal is transformed with the discrete Fourier transform, we find which frequencies are most prominent (with the highest amplitude) in the fringe signal. These frequencies are then used as input for the sine and cosine wave model functions. The number of frequencies, and thus the number of terms to be used in Equation (10), will be evaluated later in the article.

3.4 | Improvements of the Fringe EMSC algorithm

In this study, several improvements to the Fringe EMSC algorithm are proposed. The algorithm is shown in Figure 2, where the improvements with respect to the algorithm proposed by Konevskikh et al. [7] are marked in red. The new algorithm includes a robust and precise frequency estimation and suggests to select and use several frequencies in the Fourier domain. In the following, a step-wise explanation of the new algorithm is presented. Subsequently, the algorithm is demonstrated in

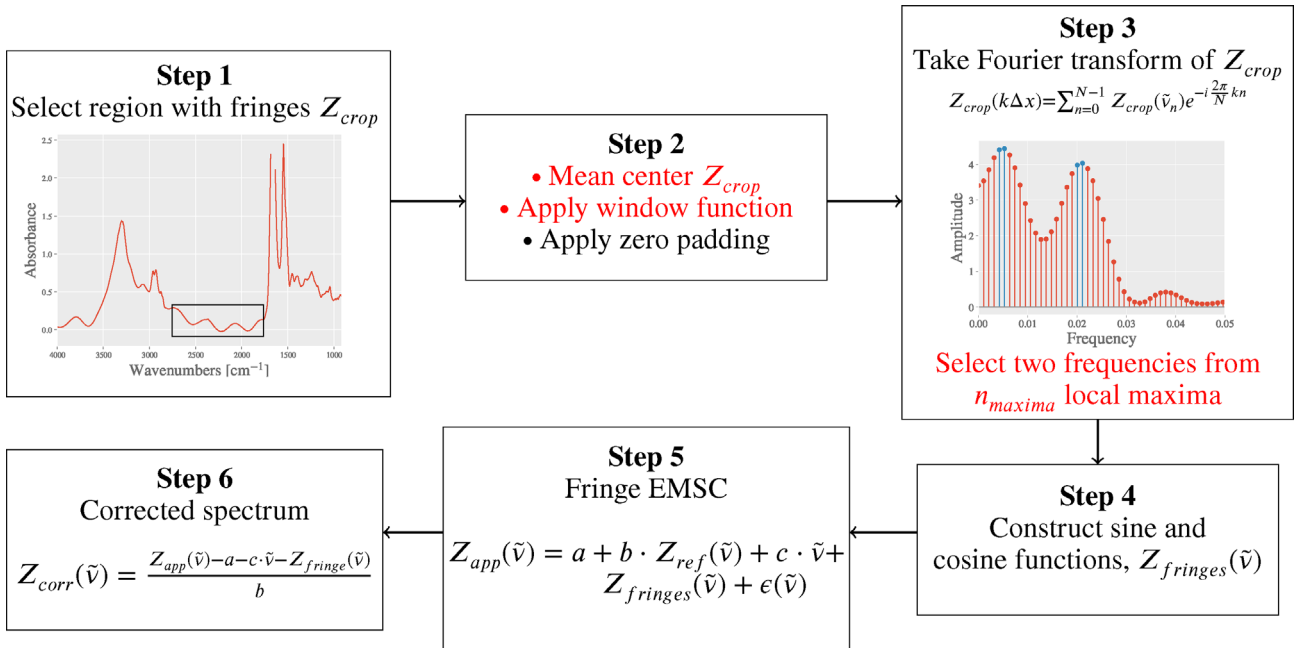


FIGURE 2 The Fringe EMSC algorithm with improvements marked in red (wrt. the algorithm of Konevskikh et al. [7]). Each step of the algorithm is explained in more detail in the text

use on two datasets of infrared spectroscopic measurements. Two filtering methods which are used when preprocessing infrared hyperspectral images are also presented. The first method is used to identify spectra which are affected by interference fringes, while the second method is used to discard empty background spectra in an infrared image. The improved Fringe EMSC algorithm consists of six steps (the algorithm is available open source at BioSpec Norway's GitHub page; <https://github.com/BioSpecNorway/biospectools>):

Step 1: The user specifies a region containing undisturbed fringes $Z_{crop}(\tilde{\nu})$, that is, within the “silent” region. The data used in this study did not show signals from CO₂ in this region. In cases where variations in CO₂ is a problem, we suggest to linearly interpolate the CO₂ region prior to the following steps.

Step 2: The selected region $Z_{crop}(\tilde{\nu})$ is then mean centered. This is because an offset in $Z_{crop}(\tilde{\nu})$ leads to high amplitudes for low frequencies in the Fourier domain, which are unrelated to the fringes in the measured spectra. These low frequencies are suppressed by mean centering.

Subsequently, a window function is applied to minimize spectral leakage caused by non-integer number of periods in the selected spectral window. This makes the algorithm more robust with respect to the choice of $Z_{crop}(\tilde{\nu})$. In the codes provided on GitHub, the triangular Bartlett window function is used. Finally, zero padding is performed in order to increase the resolution in the Fourier domain. However, since padding makes the Fourier transform more computationally expensive, we suggest as a trade-off to pad with $n_{pad} = 5 \cdot N$, that is, five times the signal length N at the end of the signal.

Step 3: A fast Fourier transform (FFT) is applied, and the two most pronounced frequencies are selected from a number of local maxima n_{max} . To find out how many local maxima one should include, the recommended strategy is to inspect the Fourier transformed signal to see how many prominent frequencies the signal consists of. Usually n_{max} should be set to 1 or 2. This way of selecting the most prominent frequencies deviates from previous practise. In the following we first present the method for a more precise fringe signal estimation by performing a simulation study, and thereafter the method for selecting frequencies from multiple maxima is presented for experimentally obtained data.

Method for a more precise fringe signal estimation (simulation study)

The Fringe EMSC model is based on two approximations: the first relates to the fact that fringes are described by treating the sample as a non absorbing film and the fringe signal is then considered as additive to the pure

absorbance. The second approximation is made by using only the first term of Equation (8), that is, the fringe signal is described by a sine wave.

To study the effect of approximations included in the Fringe EMSC, we compare the following three models: Using the (1) *Full analytical model*, where the fringe features are simulated by use of Equation (7), using a priori knowledge about n_0 and d . It is important to note that the model is the full analytical model for a non-absorbing film. To account for absorbance, a simulated pure Lorentzian absorbance spectrum was added to the model. With a pure Lorentzian absorbance we refer to an absorbance spectrum calculated from the dielectric function of a Lorentz oscillator (with the dispersive imaginary and real parts), as described in eg [41,42] Adding this spectrum to the model would be ideal if fringes could in fact be treated as additive to the chemical absorbance, and exact values for n_0 and d could be obtained. (2) *First term of the Taylor expansion*, where the fringe features are simulated using the first term on the right hand side of Equation (8). This model approximates model 1 if the first term dominates the expression, neglecting the higher frequency oscillations. As for model 1, both n_0 and d are known a priori. (3) *Sum of sine and cosine waves (The Fringe EMSC)*. In this model, the frequency is estimated by the approach described above. If the frequency of the fringe signal is estimated correctly, the model is expected to be very similar to model 2. The region 1880–6000 cm⁻¹ is selected as $Z_{crop}(\tilde{\nu})$ for frequency estimation.

The three models are compared to an exact apparent absorbance spectrum with fringes which is calculated from the transmittance of Equation (4), where the exact Fresnel equations were used with no approximations. The exact apparent absorbance spectrum has one absorbance band, simulated as a pure absorbance spectrum consisting of one Lorentzian peak at 1600 cm⁻¹. This pure absorbance spectrum served then as input for Equation (6), and the imaginary part of m , n' , was established. From n' , n_{kk} was found through the last term in Equation (5). A range of values for the constant part of the real part of the refractive index and the thickness of the film were also given, with $n_0 \in [1.2, 2.9]$ and $d \in [3 \mu\text{m}, 15 \mu\text{m}]$, respectively. An example of a simulated apparent absorbance spectrum is shown in red in Figure 3 together with the pure Lorentzian absorbance in blue. The apparent absorbance spectrum was obtained by setting $n_0 = 1.82$ and $d = 5.45 \mu\text{m}$.

Comparing the model performances: The simulated exact apparent absorbance spectra were compared to the three models. For each comparison, the root mean square error (RMSE) was found. We plot the RMSE as a 2D heatmap to visualize the fit for the three models for

different n_0 and d . The heatmaps are shown in Figure 4a–c. As is evident from the full analytical model in Figure 4a, treating the fringes as additive to the absorption leads to an error, especially when n_0

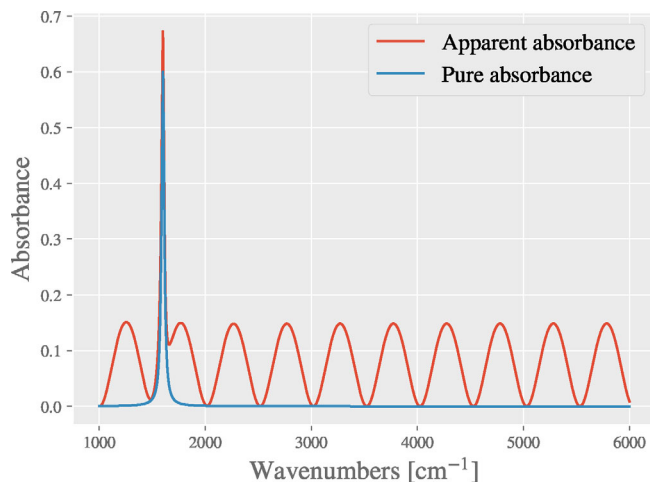


FIGURE 3 A simulated pure absorbance spectrum in blue, and a simulated apparent absorbance spectrum is calculated with Equation (4), using $n_0 = 1.82$ and $d = 5.45 \mu\text{m}$. The pure absorbance spectrum was used for calculating n' and n_{kk}

increases. By plotting the apparent absorbance (red), the estimated fringe signal (blue) and the residuals (purple) from a set of n_0 and d which correspond to high residuals ($n_0 = 2.41$ and $d = 6.92 \mu\text{m}$), we show that the error has a Fano line shape which is present in the residuals (see Figure 4d). This is because scattering affects absorption efficiency, and constructive and destructive interference tend to increase and decrease the absorption, respectively. Interference leads to electromagnetic field intensity changes within the layer, leading to non-linear absorption effects [36]. For biological samples, with a relatively low n_0 (with an n_0 between 1 and 1.5), the effect of dispersion is low. In order to evaluate the effect of the substrate, the relative refractive index between the sample and the substrate needs to be taken into account. For commonly used materials such as CaF_2 this ratio is within range that is valid for our approximation. Figure 6 shows that a ratio of around two still gives a valid approximation.

For the model 2 using the first term of the Taylor expansion we see that the RMSE increases as $n_0 > 2.4$, as expected. When $n_0 > 2.4$, the Taylor expansion is no longer valid. An increasing n_0 is also associated with more significant higher order terms. The first term of the

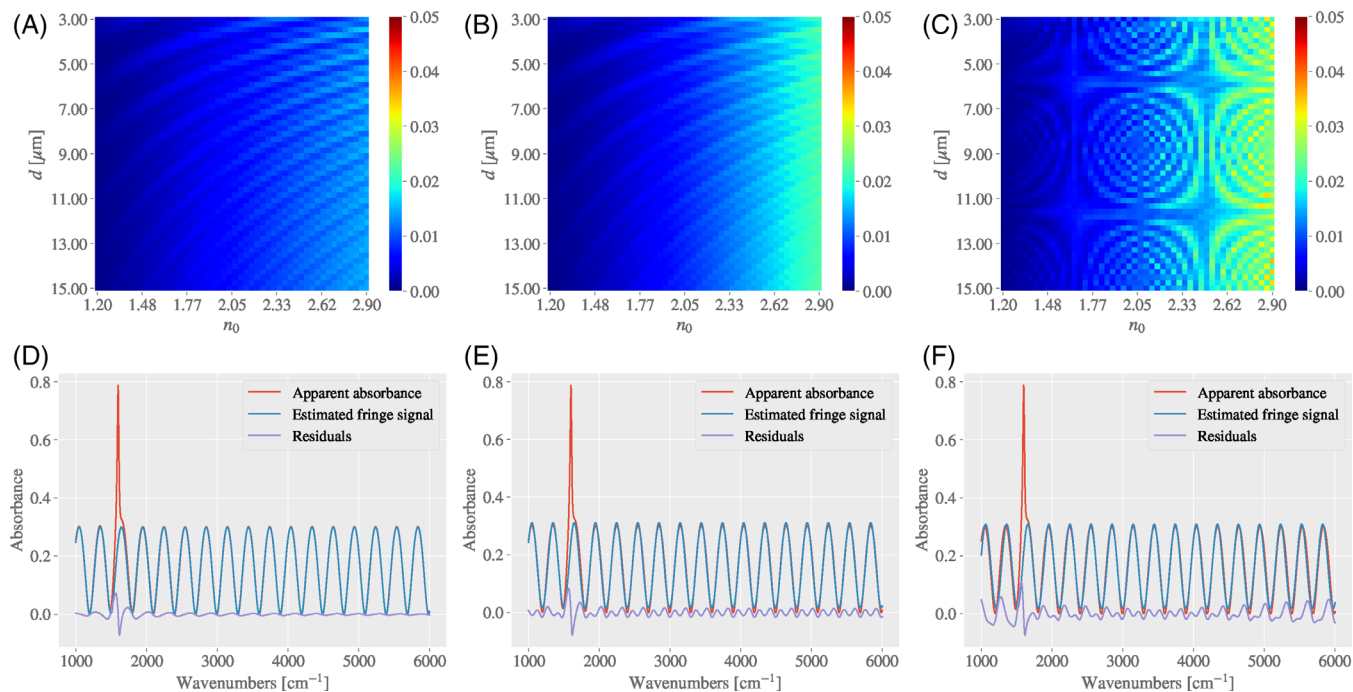


FIGURE 4 Upper panel: Heatmap showing the RMSE when correcting different apparent absorbance spectra, using different combinations of n_0 and d in the simulation. The models used for correction was (a) model 1 (full analytical model), (b) model 2 (first term of the Taylor expansion) and (c) model 3 (the Fringe EMSC model). Lower panel: Examples of correction using the different models are shown for $n_0 = 2.41$ and $d = 6.92 \mu\text{m}$. The apparent absorbance spectrum is plotted in red, the estimated fringe signal which is removed from the apparent absorbance spectrum is shown in blue, and the residuals in purple. (d) Using model 1, the residuals are low outside the chemical absorption band. (e) Using model 2, we get additional residuals from the lack of higher order cosine terms (Equation 8). (e) Using model 3, it is evident that the estimation of the frequency is not satisfactory

expansion is no longer sufficient to express the fringe signals. From Figure 4e we see that this is consistent with an increase in the higher frequency terms in Equation (8), which are present in the residuals. For the Fringe EMSC model, however, the RMSE is higher for some combinations of n_0 and d compared to the model using the first term of the Taylor expansion. This results in the circular patterns in the RMSE heatmap (Figure 4c). The reason for the circular patterns is that the estimated frequency does not correspond to the correct frequency of the fringe signal, as can be seen from Figure 4f. This can be confirmed by plotting the difference in the known main frequency of the fringe signal ($4\pi n_0 d$, from Equation 8) and the

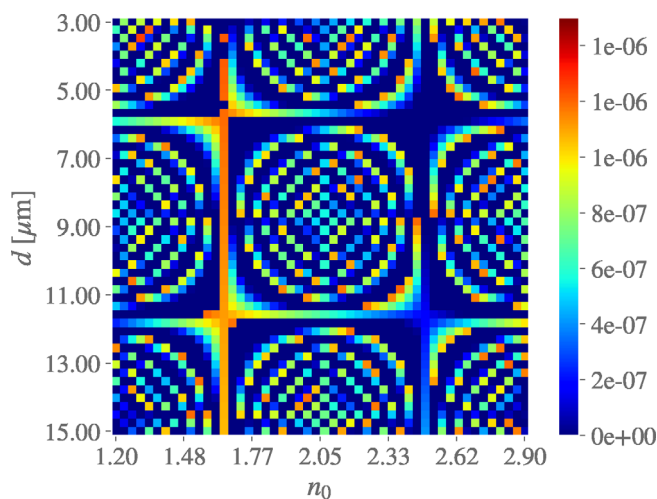


FIGURE 5 Heatmap showing the difference in the desired estimated frequency ($4\pi n_0 \tilde{d}$ from Equation 8) and the frequency estimated from the Fourier transform in the Fringe EMSC model. The same circular patterns as in Figure 4c is present, confirming that the difference between the performance of model 2 (the first term of the Taylor expansion) and model 3 (the Fringe EMSC) is due to the incorrect estimation of the frequency

estimated frequency, which results in the plot in Figure 5. We observe that the same circular patterns appear.

The erroneous estimation of the frequency is due to the resolution of the frequency axis in the Fourier domain. The error for the estimation of the frequency could in principle be avoided by using a higher n_{pad} for the zero padding. In practice, when increasing n_{pad} for zero padding, we did not observe overfitting. However, increasing n_{pad} increases the computation time considerably. For this example, setting $n_{pad} = 12.4 \cdot N$ reduced the error caused by the erroneous frequency estimation by 53%, but resulted in an increase of the computation time by a factor of 10.2 (averaged over 10 runs, for all combinations of n_0 and d). On a Lenovo P53 laptop, with 128 GB RAM and an intel i7 (9th gen) processor, it takes 6 ms to correct one spectrum consisting of 5000 data points, with $n_{pad} = 5 \cdot N$. To avoid increasing computation time, we propose another strategy for handling this issue. We suggest selecting two neighboring frequencies from each local maximum which is selected in the Fourier domain. The resulting RMSE heatmap for model 3 with this adjustment is shown in Figure 6. As we can see from this figure, the residuals caused by the wrongly estimated frequency are now reduced, and the heatmap in Figure 6a resembles the one in Figure 4b. This strategy is therefore proposed as a step in the Fringe EMSC algorithm. In the Fringe EMSC algorithm proposed by Azarfar et al. [16], they suggest to use multiple iterations. However, by selecting frequencies from multiple maxima in the Fourier domain, we experience that an iterative approach is not needed.

Method for selecting multiple frequency maxima in the Fourier domain (experimental data):

In order to illustrate the approach for selecting multiple frequencies, we used experimental data obtained from thin film hair cross sections. From the experimentally

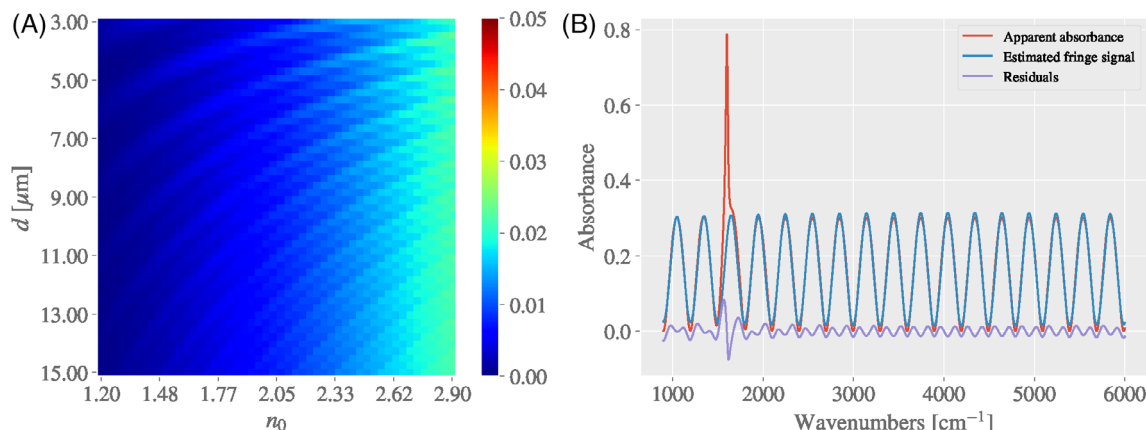
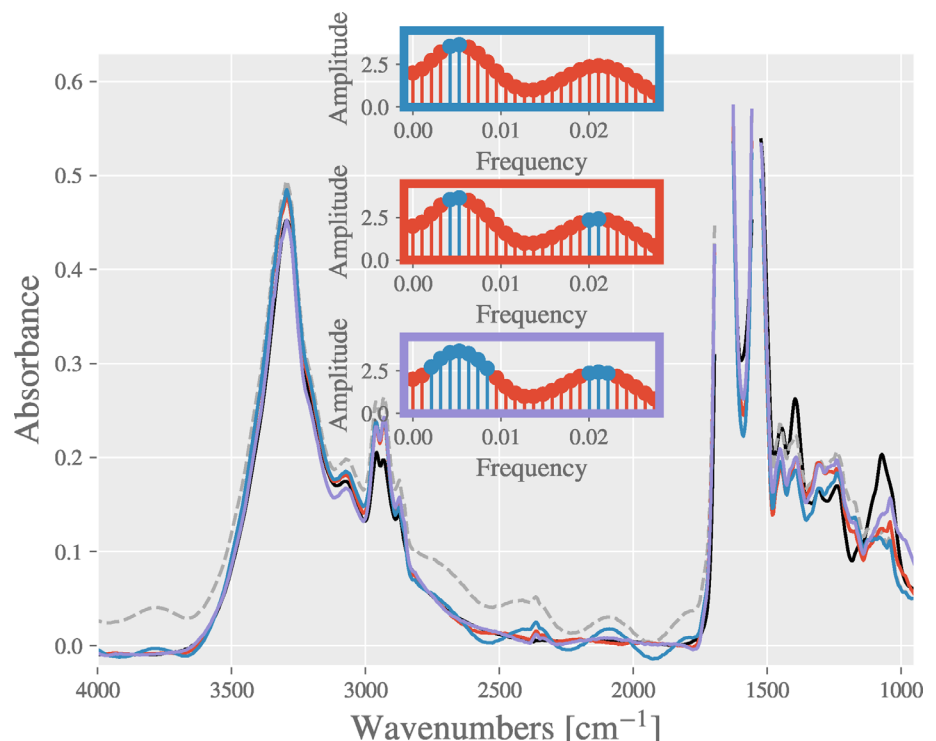


FIGURE 6 When two frequencies are included in model 3 (the Fringe EMSC), (a) the heatmap of the RMSE and (b) the residuals resemble the results from the model 2 in Figure 4b and e, respectively

obtained spectra, it is evident that there can be multiple frequencies present in the fringe signal. This can be seen from the raw spectrum plotted as a dashed gray line in Figure 7, where a relatively high-frequency signal is superimposed on a more low-frequency background signal in the region between 1750 and 2750 cm^{-1} . From the Fourier transform of this region, shown as inserted panels in Figure 7 (mean centering, windowing and zero-padding is applied), the two frequencies are present as two broad local maxima. In order to capture both frequency components of the signal, we propose to include two neighboring frequencies around a number n_{maxima} local maxima, instead of including a portion of the frequencies with the highest amplitude, as suggested by Azarfar et al. [16] The effects of including different sets of frequencies are shown in Figure 7, where the sets of selected frequencies are highlighted in the inserted panels. The corresponding corrected spectra are shown in Figure 7, where the frame color matches the color of the corrected spectrum. It is evident that using only the low frequencies results in a corrected spectrum where the main fringe signal is still present (blue line). The result of the proposed method (including frequencies from the middle panel) which is plotted in red is clearly favorable. Including 10 of the most prominent frequencies leads to a smooth silent region (purple line), but the method is prone to overfitting, which can be seen in the fingerprint region where the corrected spectrum assimilates the reference spectrum. Including 10 frequencies results in 20 model function in $Z_{\text{fringe}}(\tilde{\nu})$ and in a model complexity which is not needed.

FIGURE 7 The frequency selection is shown in the inserted panels, where the color of the frame corresponds to the colors of the corrected spectra in the main figure. The raw spectrum is shown in dashed gray and the reference spectrum in black. The raw spectrum is selected from the hair cross section dataset



We now return to the six steps of the Fringe EMSC algorithm:

Step 4: For each frequency selected in the Fourier domain, a pair of sine and cosine functions is created in the EMSC fringe model of Equation (9). The complete Fringe EMSC model includes the following: the reference spectrum, a constant and linear baseline, the sine and cosine functions with the selected frequencies.

Step 5: Estimate the EMSC parameters for all model functions, using least squares regression.

Step 6: The measured spectrum is finally corrected according to Equation (11).

The code published on GitHub performs all the steps one to six fully automatically. As input parameter, the operator chooses the number of frequency maxima n_{maxima} , and the spectral window used for the fringe frequency estimation.

In the following, the suggested algorithm is validated on two example data sets of spectra obtained from biological material.

4 | EXPERIMENTAL VALIDATION

4.1 | Case study: Correction of single spectra from pollen grains embedded in paraffin

In the following example, we show how the Fringe EMSC algorithm can be used on infrared spectra of pollen grains embedded in paraffin. Since pollen grains normally are of

spherical shape, with a size in the same order as the wavelength of the employed radiation, the samples are highly effective Mie scatterers [33]. In order to avoid Mie scattering in these measurements, the pollen grains were embedded in paraffin, and sandwiched between two sheets of a thin polyethylene film. While the paraffin embedding successfully suppressed the Mie scattering, fringes were observed in measured spectra due to the nearly perfect film caused by the paraffin. Since the refractive indices of paraffin and polyethylene are similar, 1.47 and 1.53, respectively [43,44], we can treat them optically as the same material. Raw spectra are shown in gray in Figure 8, where the mean spectrum is shown in black. The mean spectrum does not contain fringe signals, and is therefore well suited as a reference spectrum in the Fringe EMSC. In this dataset not all spectra are affected by fringes, and a filtering technique was therefore developed to detect spectra which need fringe correction.

4.1.1 | Selecting spectra requiring fringe correction

In order to select spectra that require fringe correction, we employ a standard EMSC model in a narrow range in the silent region. We select a narrow range such that it contains 4–8 fringe maxima and correct this range with a standard EMSC with polynomials up to the third order. As a reference, we use the same reference spectrum which was used for the Fringe EMSC, that is, the mean spectrum. It turned out that spectra without fringes could be modeled with the standard EMSC model with low

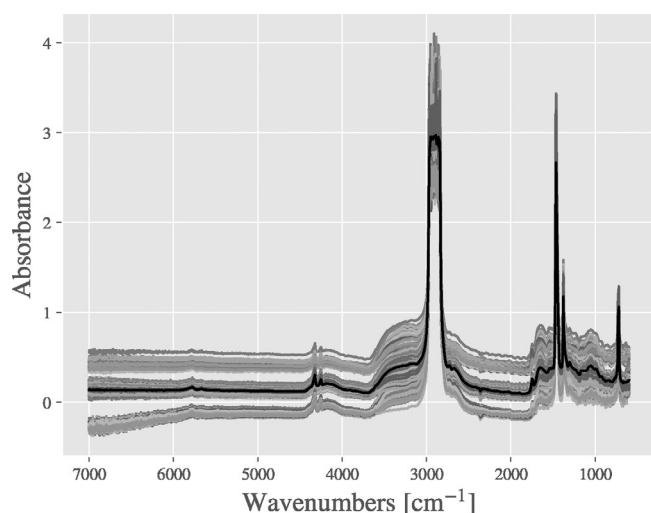


FIGURE 8 A selection of apparent absorbance spectra from transmission measurements of single pollen grains embedded in paraffin. The mean spectrum is shown in black

residuals, while spectra which contain fringes show relatively high residuals. This offers an opportunity to select spectra that require fringe correction based on the residuals of this standard EMSC. In Figure 9a,b, two examples for the correction of the narrow ranges with a standard EMSC are shown. The corrected spectra are shown in blue and the reference spectrum in purple. In Figure 9, fringes are not present in the raw spectrum (shown in red), and the residuals (gray) are small. In Figure 9, fringes are present in the raw spectrum, and since they cannot be modeled they are also present in the residuals. A cutoff value based on the RMSE, here set to 0.0027, allows to identify spectra which need fringe correction and separate them from spectra for which a standard EMSC is sufficient for the correction. The separation is shown in Figure 9, where Figure 9c shows the spectra without fringe signals, and Figure 9d shows spectra with fringes.

4.1.2 | Correcting pollen spectra

The thickness of the paraffin and polyethylene film of our pollen data set is in the order of 40–70 micrometer, which results in high-frequency fringes in the pollen spectra. According to our experience, the correction of these high-frequency fringes is difficult since high-frequency fringes in the EMSC model may model chemical signals in the chemically active regions. An example of our fringe correction is shown in Figure 10 where Figure 10b,c show two different magnified ranges of Figure 10a, to make the effect of the correction more visible. The mean spectrum was used as a reference spectrum, and two frequencies from two local maxima in the frequency domain were chosen. As we can see from Figure 10, there are still fringe signals left in the spectra, and in some regions the fringe signal is even increased. One should note that in this example, a relatively broad spectral region ($600\text{--}7000\text{ cm}^{-1}$) has been kept. However, it is common to cut the spectra above 4000 cm^{-1} , due to the lack of molecular vibrations in this region. We have chosen to keep the region above 4000 cm^{-1} for illustrating the challenges of correcting high-frequency fringes better, and how to overcome them. However, the suggested approach for fringe correction does not require to keep the region above 4000 cm^{-1} . In this correction we have employed weights, shown as a dashed gray line in Figure 10. Weights are used to increase or decrease the importance of different spectral regions and can for example be used to suppress the effect of variations in CO_2 content in the spectra [40]. In this example, weights are used to amplify the effect of residuals in regions which

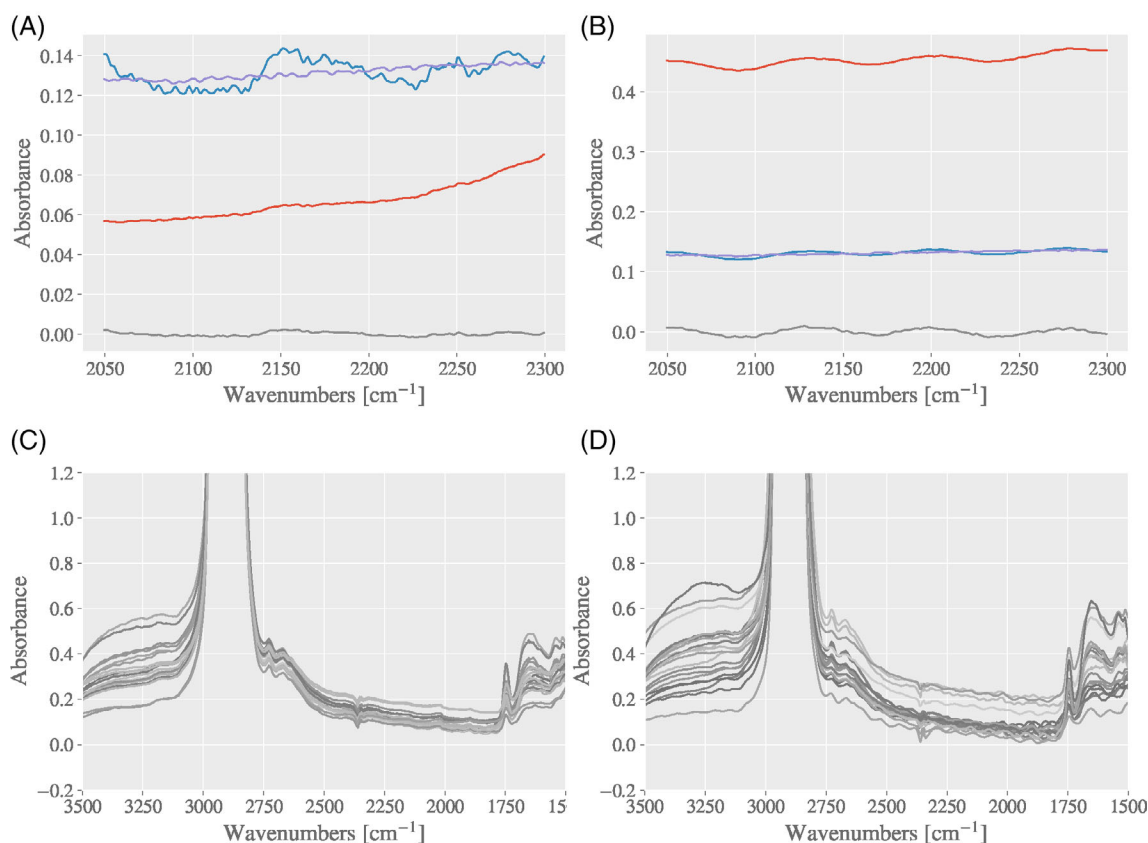


FIGURE 9 Filtering method for separating spectra of pollen embedded in paraffin which need fringe correction and spectra which do not need fringe correction. (a) An example of a spectrum which does not need fringe correction, showing the selected part of the silent region in red, the reference in purple, the corrected in blue and the residuals in gray. (b) The same is shown for a spectrum which needs fringe correction. Keep in mind that the y-axes in (a) and (b) are scaled differently, and that the residuals in (b) are significantly higher than in (a). The RMSE value of a standard EMSC model can be used to separate spectra into (c) spectra which does not need fringe correction, and (d) spectra which need fringe correction

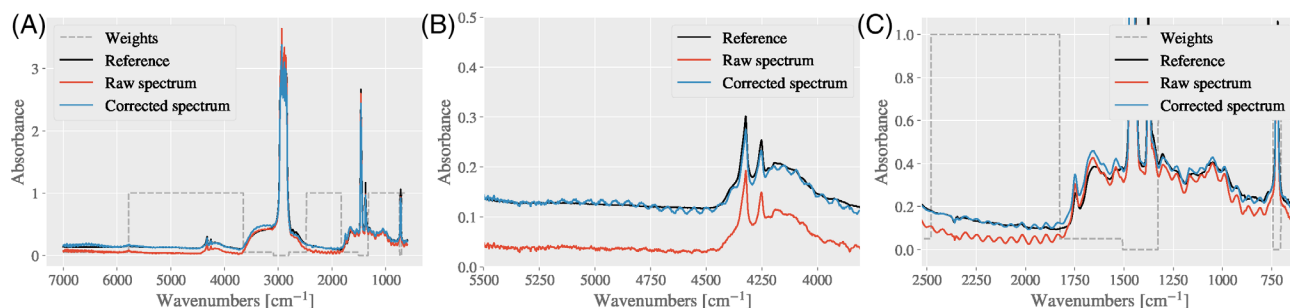


FIGURE 10 (a) Example of correcting a full spectrum with the Fringe EMSC model. The raw spectrum is show in red, the reference spectrum in black, weights in grey (dashed) and the corrected spectrum in blue. When the spectral range is broad, and the fringes are high-frequent with low amplitude, the fringe signal can in some cases be (b) artificially increased, or (c) simply not removed

are chemically inactive and where fringes are clearly present. Down-weighting of regions with strong paraffin signals is used as a measure to suppress the effect of chemical signals of paraffin on the estimation of the EMSC b parameter. Since the sample of interest is the pollen grain and not the paraffin, we aim at scaling the spectra according to the optical thickness of the pollen

grains. Using zero weights is analogous to removing the region completely.

An alternative to correcting the full spectral range (which in this case is 600–7000 cm⁻¹) is to perform piecewise correction, or selecting a smaller spectral region of interest. For the pollen samples, the region of interest is in the range 800–2000 cm⁻¹. In order to be able to

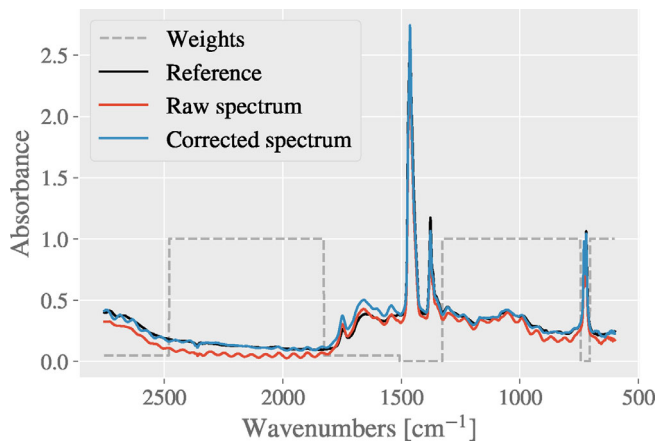


FIGURE 11 By selecting a shorter range, spectra which are affected by high-frequent and low-amplitude fringes can be successfully corrected. The raw spectrum is shown in red, the reference in black, weights in gray (dashed) and the corrected spectrum in blue

estimate the fringe frequency we included a chemically silent region as well, and chose the region 600–2750 cm^{-1} . As is shown in Figure 11, the fringe correction works well when a smaller spectral range is corrected. The fringe signal is removed completely. We evaluated the complete dataset of pollen spectra which were identified as needing fringe correction and found that it worked well when the fringe correction was applied to the region 600–2750 cm^{-1} .

4.2 | Case study: correction of infrared images of hair cross sections

The Fringe EMSC algorithm was further evaluated on a hyperspectral image of hair cross sections. In Figure 12a, a set of spectra that were randomly selected from inside the hair section is shown. The chemical image obtained

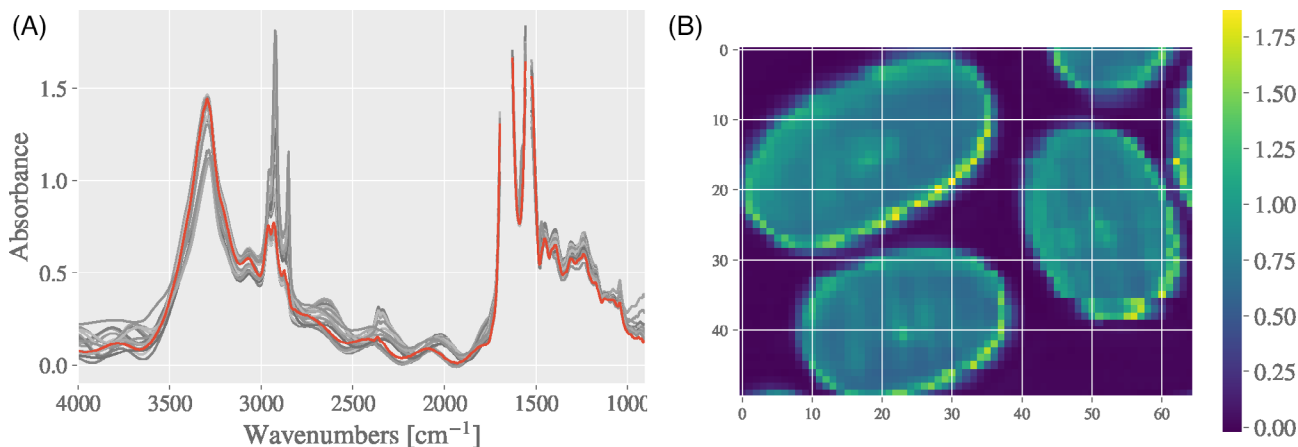


FIGURE 12 (a) A selection of infrared transmission spectra from a sample of thin film hair cross sections. One of the spectra is highlighted in red. The spectra are taken from a hyperspectral image, here plotted at 2952 cm^{-1} in (b)

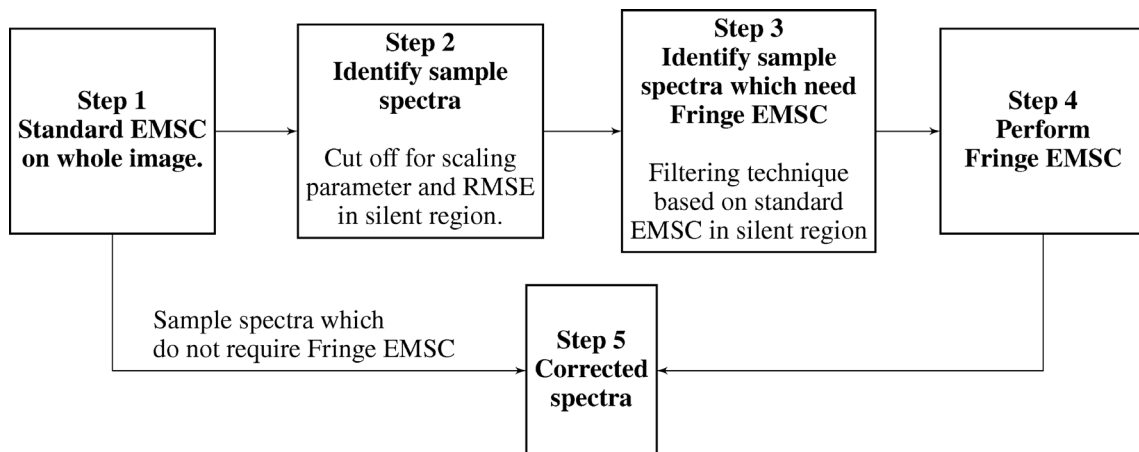


FIGURE 13 Shows the suggested pipeline for performing fringe correction on infrared hyperspectral images. The steps are explained in more detail in the text

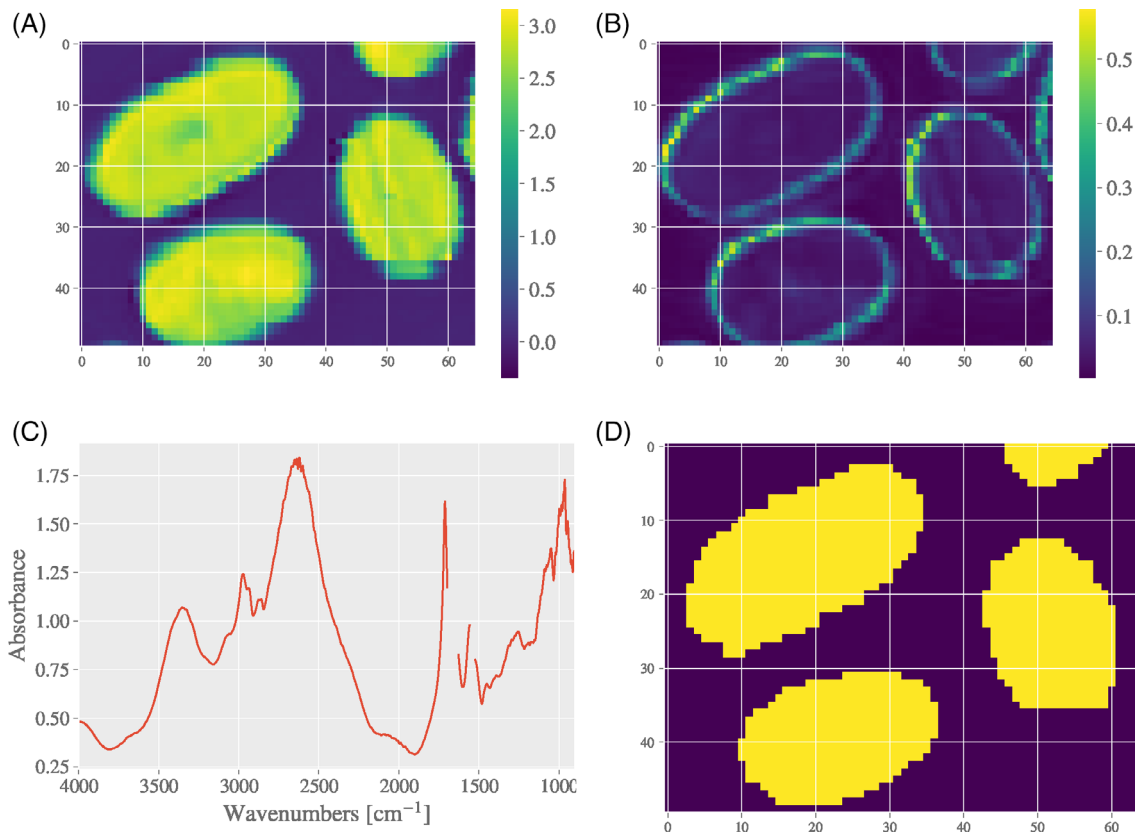


FIGURE 14 (a) Plot of the scaling parameter from a standard EMSC applied to the hair cross section image. The scaling parameter is interpreted as a measure of sample thickness in each pixel. (b) RMSE from the standard EMSC. It is evident that the RMSE is higher at the edges of the sample. (c) A spectrum from the edge is shown to illustrate the scattering effects which are present at the edges. (d) The mask which is used to separate the sample spectra from the empty slide spectra. The mask is based on a cutoff value for the scaling parameter and the RMSE value of the EMSC model

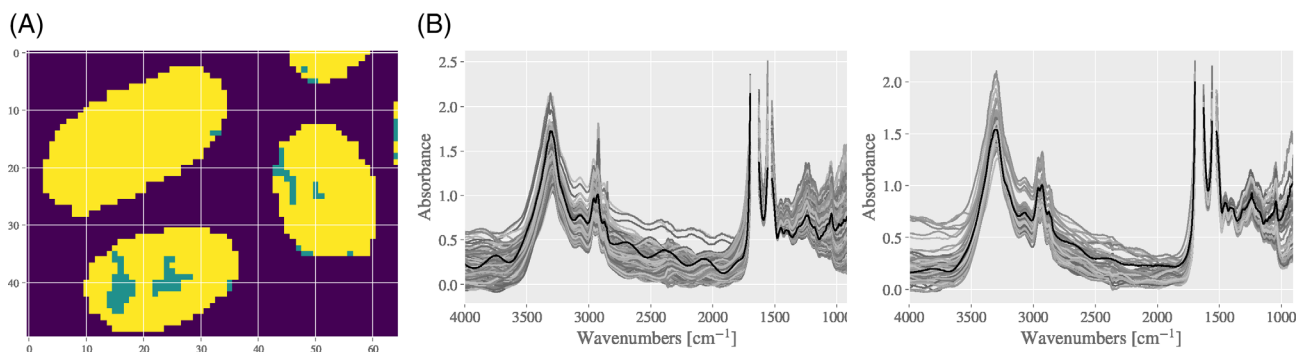


FIGURE 15 Separation of spectra which need fringe correction (yellow pixels in (a), shown in (b)) and spectra which do not need fringe correction (green pixels in (a), spectra shown in (c))

from the absorbance values at 2952 cm^{-1} from the same image is shown in Figure 12b. The amide I and II peaks are partly removed from the spectra due to saturation of the absorbance signal, that is, the sample absorbs the incoming radiation completely in these spectral regions. For preprocessing hyperspectral images, different filtering techniques are often needed prior to the

preprocessing itself, and in the following we suggest a robust method for preprocessing image data with fringe signals. The pipeline is shown in Figure 13, and each step is described in the following.

Step 1: For hyperspectral infrared images, the first step is often to separate the sample spectra from the image background spectra (pixels which refer to the empty slide).

This can be achieved easily with a basic EMSC by including polynomials up to the second degree [45]. We use the Matrigel spectrum as a reference spectrum [14], since the mean spectrum from the whole image contains fringe signals, which makes it unsuitable as a reference spectrum. From the standard EMSC, the scaling parameter referring to the effective optical path length is used to obtain a measure of the thickness of the sample in each pixel. The scaling parameter for each pixel is shown in Figure 14a. In our case a threshold of 2.3 was suitable to separate the sample spectra from the background. The exact threshold used depends on the scaling of the reference spectrum compared to the sample spectra, and needs to be chosen for each application. An additional threshold for the RMSE in the region 1761–2840 cm^{-1} was set to 0.1 in order to filter out pixels at the edge of the hair samples. At the edge of the sample, strong scattering signals are present in the spectra, which lead to high residuals in the silent region (Figure 14b). An example of such a spectrum obtained at the edge of the hair sample, in pixel (8,9), can be seen in Figure 14c. These spectra would need a different preprocessing technique than Fringe EMSC, a study which will be described elsewhere. Here, they are treated as image background spectra. The separation into image background pixels and sample pixels can be seen in Figure 14d.

Step 2: After separating the sample spectra from the image background, we identify spectra which need fringe correction. This is done with the method described in section 4.1.1. The range 1906–2681 cm^{-1} was used for the EMSC for filtering spectra that need fringe correction. The threshold for the RMSE was set to 0.015. Since this region contains also spectral bands for CO_2 , the region between 2340 and 2380 cm^{-1} was removed before calculating the RMSE value. The identification of spectra which need fringe correction can be seen in Figure 15.

Figure 15a shows the pixels which need Fringe EMSC in yellow, and pixels for which a standard EMSC is sufficient in green. The corresponding spectra are shown in Figure 15b and c, respectively.

Step 3: The spectra which were identified as needing fringe correction were then corrected with the Fringe EMSC. The region 1906–2681 cm^{-1} was used to estimate the frequency of the fringe signal, and two local frequency maxima were chosen from the Fourier domain. For each local maximum, two frequencies were selected. The weights shown in Figure 16a,b as a gray dashed line were both applied for the standard EMSC for selecting sample spectra and for the Fringe EMSC. Polynomials up to the second order were used in the Fringe EMSC. The fringe corrected spectra are shown as raw spectra in Figure 16a, and corrected spectra in Figure 16b. The reference spectrum is plotted in red, and one of the sample spectra is plotted in black.

Step 4: The corrected spectra can now be combined to a complete preprocessed dataset. The spectra which needed Fringe EMSC are taken from step 3, while the spectra which did not need the Fringe EMSC were already corrected with a standard EMSC in step 1. Since the same reference spectrum was used in both corrections, all spectra are scaled in the same way. An image of the corrected spectra at 2952 cm^{-1} is shown in Figure 16c. Compared to the image of the raw data at the same wavenumber (Figure 12b) it is evident that the contrast is better. For the largest hair sample, even the lipid rich medulla is visible. One should note that with the Fringe EMSC, the spectra are baseline corrected and scaled according to the reference spectrum, in addition to the fringe correction. The biggest difference between the raw spectra and the corrected spectra is due to baseline correction and scaling (see Figure 16a,b). This demonstrates one of the advantages of the Fringe EMSC. All physical effects are comprised in the model and therefore handled simultaneously.

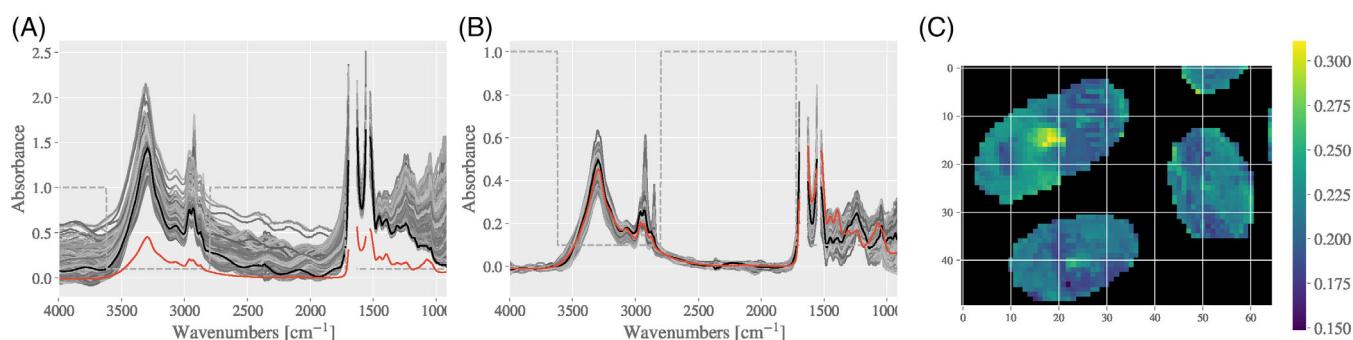


FIGURE 16 Correcting hair cross section spectra with the Fringe EMSC model. (a) Raw spectra and (b) corrected spectra, with the reference spectrum (Matrigel) in red, and one of the spectra is highlighted in black. The weights are shown in dashed grey. (c) An image from the corrected spectra at 2952 cm^{-1} , corresponding to Figure 12b. The improvement in the image contrast is clearly visible

5 | CONCLUSION

In this study, we developed the Fringe EMSC into a stable and automated algorithm for fringe removal in infrared microspectroscopy and infrared microspectroscopic imaging. We presented several improvements to the algorithm. A more precise estimation of the frequency of the fringe signal is achieved by mean centering the signal and applying a window function to it prior to the Fourier transform. Since the fringe frequencies are estimated from the silent region, it cannot be guaranteed that the fringes in the region with strong absorption are exactly modeled, since there might be slight frequency shifts. However, visually, the correction seems to work well. To better understand the effects of the approximations made in the Fringe EMSC, a simulation study was carried out. Consequently, the practice for including multiple frequencies in the Fringe EMSC was revised, and a best practice was found by including two frequencies per n_{maxima} local maxima in the Fourier domain. The presented algorithm is robust, e.g. the choice of the number of frequencies for the fringe corrections does not affect the correction result notably. The improved Fringe EMSC is demonstrated on two experimental datasets, accompanied by two filtering methods. The first filtering method allows for separation of sample spectra from empty background spectra in infrared imaging, while the second filtering method identifies spectra which are affected by fringe signals. Both filtering methods are based on standard EMSC, which makes them robust and easy to use. Python codes for the Fringe EMSC and filtering methods are available on GitHub (<https://github.com/BioSpecNorway/biospectools>).

ACKNOWLEDGMENTS

The research reported in this article was supported by the Research Council of Norway through the DeepHyperSpec project (Grant/Award Number: 289518) and the PollChem project (Grant/Award Number: 249844). The author would also like to acknowledge funding and the use of instrumentation at the SMIS beamline from the SOLEIL synchrotron.

CONFLICT OF INTEREST


The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Johanne Heitmann Solheim  <https://orcid.org/0000-0001-6726-1692>

Ferenc Borondics  <https://orcid.org/0000-0001-9975-4301>

Boris Zimmermann  <https://orcid.org/0000-0001-5046-520X>

Christophe Sandt  <https://orcid.org/0000-0002-6432-2004>

Florian Muthreich  <https://orcid.org/0000-0002-1242-7381>

REFERENCES

- [1] M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, et al., *Nat Protoc* **2014**, 9(8), 1771.
- [2] D. Naumann, D. Helm, C. Schultz, *Bacterial Diversity and Systematics*, Boston: Springer, **1994**, p. 67.
- [3] M. Kümmerle, S. Scherer, H. Seiler, *Appl. Environ. Microbiol.* **1998**, 64(6), 2207.
- [4] J. Schmitt, H.-C. Flemming, *Int. Biodeter. Biodegr.* **1998**, 41(1), 1.
- [5] J. Brynmor, P. S. C. Davis, R. Bhargava, *Anal. Chem.* **2010**, 82(9), 3474.
- [6] J. Brynmor, P. S. C. Davis, R. Bhargava, *Anal. Chem.* **2011**, 83(2), 525.
- [7] T. Konevskikh, A. Ponossov, R. Blümel, R. Lukacs, A. Kohler, *Analyst* **2015**, 140(12), 3969.
- [8] H. Martens, E. Stark, *J. Pharm. Biomed. Anal.* **1991**, 9(8), 625.
- [9] B. Mohlenhoff, M. Romeo, M. Diem, B. R. Wood, *Biophys. J.* **2005**, 88(5), 3635.
- [10] A. Kohler, J. Sulé-Suso, G. Sockalingum, M. Tobin, F. Bahrami, Y. Yang, J. Pijanka, P. Dumas, M. Cotte, D. G. van Pittius, G. Parkes, H. Martens, *Appl. Spectrosc.* **2008**, 62, 259.
- [11] P. Bassan, H. J. Byrne, F. Bonnier, J. Lee, P. Dumas, P. Gardner, *Analyst* **2009**, 134(8), 1586.
- [12] Thomas Van Dijk, David Mayerich, P Scott Carney, Rohit Bhargava, *Appl. Spectrosc.* **2013**, 67 (5), 546–552.
- [13] T. Konevskikh, R. Lukacs, R. Blumel, A. Ponossov, A. Kohler, *Faraday Discuss.* **2016**, 187, 235.
- [14] J. H. Solheim, E. Gunko, D. Petersen, F. Großerüschkamp, K. Gerwert, A. Kohler, *J. Biophotonics* **2019**, 12(8), e201800415.
- [15] E. Ly, O. Piot, R. Wolthuis, A. Durlach, P. Bernard, M. Manfait, *Analyst* **2008**, 133(2), 197.
- [16] G. Azarfar, E. Aboualizadeh, N. M. Walter, S. Ratti, C. Olivieri, A. Norici, M. Nasse, A. Kohler, M. Giordano, C. J. Hirschmugl, *Analyst* **2018**, 143(19), 4674.
- [17] F. Capitani, C. Marini, S. Caramazza, P. Dore, A. Pisanu, L. Malavasi, L. Nataf, F. Baudelet, J.-B. Brubach, P. Roy, P. Postorion, *J. Phys. Chem. C* **2017**, 121(50), 28125.
- [18] T. Hirschfeld, A. W. Mantz, *Appl. Spectrosc.* **1976**, 30(5), 552.
- [19] A. Baghdadi, *Applied Spectroscopy* **1983**, 37(6), 520.
- [20] F. R. S. Clark, D. J. Moffatt, *Appl. Spectrosc.* **1978**, 32(6), 547.
- [21] T. Iwata, J. Koshoubu, *Appl. Spectrosc.* **1994**, 48(12), 1453.
- [22] A. M. A. Pistorius, W. J. DeGrip, *Vibrat Spectrosc* **2004**, 36(1), 89.
- [23] S. W. King, M. Milosevic, *J. Appl. Phys.* **2012**, 111(7), 073109.
- [24] M. Milosevic, S. W. King, *J. Appl. Phys.* **2012**, 112(9), 093514.
- [25] M. Milosevic, S. W. King, *ECS J Solid State Sci Technol* **2014**, 4(1), N3146.
- [26] G. Thomas, S. P. Mayerhöfer, U. Hübner, J. Popp, *Analyst* **2020**, 145(9), 3385.
- [27] B. Zimmerman, V. Tafintseva, M. Bagcioglu, M. H. Berdahl, A. Kohler, *Anal. Chem.* **2016**, 88(1), 803.
- [28] D. F. Swinehart, *J. Chem. Educ.* **1962**, 39(7), 333.

- [29] A. Kohler, C. Kirschner, A. Oust, H. Martens, *Appl. Spectrosc.* **2005**, 59(6), 707.
- [30] M. Harald, S. A. Jensen, P. Geladi, *Proceedings of the Nordic Symposium on Applied Statistics*, Stokkand Forlag Publishers Stavanger, Norway **1983**, p. 205.
- [31] P. Geladi, D. MacDougall, H. Martens, *Appl. Spectrosc.* **1985**, 39(3), 491.
- [32] J. L. Ilari, H. Martens, T. Isaksson, *Appl. Spectrosc.* **1988**, 42(5), 722.
- [33] R. Lukacs, R. Blümel, B. Zimmerman, M. Bağcıoğlu, A. Kohler, *Analyst* **2015**, 140(9), 3273.
- [34] C. F. Bohren, D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, Weinheim, Germany: John Wiley & Sons, **2008**.
- [35] M. Romeo, M. Diem, *Vibrat Spectrosc* **2005**, 38(1–2), 129.
- [36] M. A. Brandsrud, R. Blümel, R. Lukacs, E. Seim, E. S. Marstein, E. Olsen, A. Kohler, *Journal of Photonics for Energy* **2021**; 11(2), 024501.
- [37] P. Bassan, A. Kohler, H. Martens, J. Lee, H. J. Byrne, P. Dumas, E. Gazi, M. Brown, N. Clarke, P. Gardner, *Analyst* **2010**, 135(2), 268.
- [38] P. Bassan, H. J. Byrne, J. Lee, F. Bonnier, C. Clarke, P. Dumas, E. Gazi, M. D. Brown, N. W. Clarke, P. Gardner, *Analyst* **2009**, 134(6), 1171.
- [39] N. K. Afseth, A. Kohler, *Chemom. Intel. Lab. Syst.* **2012**, 117, 92.
- [40] S. Brown, R. Tauler, B. Walczak, *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Amsterdam, Netherlands: Elsevier, **2020**.
- [41] G. R. Fowles, *Introduction to Modern Optics*, New York: Courier Corporation, **1989**.
- [42] E. Hecht, A. Zajac, *Optics*, Vol. 4, Addison Wesley, San Francisco **2002**.
- [43] J. W. Horwitz, *Optic Eng* **2011**, 50(9), 093603.
- [44] P. R. Cooper, *Appl. Optics* **1982**, 21(19), 3413.
- [45] V. Tafintseva, V. Shapaval, M. Smirnova, A. Kohler, *J. Biophotonics* **2020**, 13(3), e201960112.

How to cite this article: J. H. Solheim, F. Borondics, B. Zimmermann, C. Sandt, F. Muthreich, A. Kohler, *J. Biophotonics* **2021**, 14 (12), e202100148. <https://doi.org/10.1002/jbio.202100148>