



Characterization and subtraction of luminescence background signals in high-wavenumber Raman spectra of human tissue

E.M. Barroso¹  | T.C. Bakker Schut² | P.J. Caspers² | I.P. Santos²  | E.B. Wolvius¹ | S. Koljenović³ | G.J. Puppels²

¹Department of Oral & Maxillofacial Surgery, Special Dental Care, and Orthodontics, Erasmus MC Cancer Institute, Rotterdam 3015 CN, The Netherlands

²Center for Optical Diagnostics and Therapy, Department of Dermatology, Erasmus MC Cancer Institute, Rotterdam 3015 CN, The Netherlands

³Department of Pathology, Erasmus MC Cancer Institute, Rotterdam 3015 CN, The Netherlands

Correspondence

Tom C. Bakker Schut, Center for Optical Diagnostics and Therapy, Department of Dermatology, Erasmus MC Cancer Institute, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands.
Email: t.bakkerschut@erasmusmc.nl

JEL Classification: E10; 123

Abstract

Raman spectroscopy in the high-wavenumber spectral region (HWR) is particularly suited for fiber-optic in vivo medical applications. The most-used fiber-optic materials have negligible Raman signal in the HWR. This enables the use of simple and cheap single-fiber-optic probes that can be fitted in endoscopes and needles. The HWR generally shows less tissue luminescence than the fingerprint region. However, the luminescence can still be stronger than the Raman signal. Hardware- and software-based strategies have been developed to correct for these luminescence signals. Typically, hardware-based strategies are more complex and expensive than software-based solutions. Effective software strategies have almost exclusively been developed for the fingerprint region. First-order polynomial baseline fitting (PBF) is the most common background/luminescence estimation employed for the HWR. The goal of this study was to characterize the luminescence background signals of HW spectra of human oral tissue and compare the performance of two algorithms for correction of these background signals: PBF and multiple regression fitting (MRF). In the MRF method, we introduce here, prior knowledge of the range of Raman signals that can be obtained from the tissues of interest is explicitly used. MRF is more robust than PBF because it does not require an a priori choice of the polynomial order for fitting the background signal. This is important because, as we show, no single polynomial order can optimally characterize all backgrounds that are encountered in HW tissue spectra. We conclude that MRF should be the preferred method for background subtraction in the HWR.

KEYWORDS

background estimation, bleaching, multiple regression fitting

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.

© 2018 The Authors Journal of Raman Spectroscopy Published by John Wiley & Sons Ltd

1 | INTRODUCTION

Raman spectroscopy has been broadly explored for medical applications in oncology aiming at early diagnosis, biopsy guidance, and surgery guidance.^[1–3] The technique can provide qualitative and quantitative molecular information about tissues and enables classification of tissues with high sensitivity and specificity. Raman spectroscopy is directly applicable for *in vivo* and *ex vivo* analysis of tissue, without the need for labels and/or reagents.^[4] Elimination of interference from laser-induced tissue luminescence is still a challenge for many medical applications of Raman spectroscopy. This intrinsic luminescence signal can be several orders of magnitude more intense than the Raman signal, which limits the analysis.^[5] To overcome this, both hardware- and software-based strategies have been developed to reduce the interference from luminescence.^[3,5–29]

An example of a hardware-based strategy is the use of specific excitation wavelengths far from the visible range. For example, the intensity of a fluorophore is dependent on the excitation wavelength. With exception of porphyrins and melanin,^[17] the excitation and emission of most tissue fluorophores exhibit maxima at ultraviolet and visible wavelengths.^[18] Thus, longer wavelengths are preferred over ultraviolet/visible.^[5,6] The selection of appropriate near infrared source is an optimization between competing factors. For example, while tissue luminescence decreases with increasing wavelength, also the Raman scattering cross section and the quantum efficiency of charge-coupled device (CCD) detectors decrease. For CCD detectors, the quantum efficiency drops below 15% at 1,000 nm and cannot be used at all above 1,100 nm. Detectors such as indium gallium arsenide, germanium, and indium phosphide can be used for wavelengths above 950 nm. Although these detectors have lower quantum efficiency and exhibit much more detector noise in comparison to silicon detectors,^[5] new types of indium gallium arsenide detectors are emerging which have very good noise characteristics, approaching those of CCD detectors.^[6]

Another hardware-based strategy for luminescence rejection is multiexcitation wavelength Raman spectroscopy, also called as shifted excitation Raman difference spectroscopy (SERDS).^[29] Raman spectra shift with the excitation wavelength, and luminescence is invariant for small excitation-wavelength shifts. The use of multiple excitation frequencies can thus be employed to separate the Raman signal from the luminescence background. The efficiency of the rejection increases with the number of excitation frequencies, which adds significant complexity to this hardware solution.^[20] This technique has been mainly used on samples that have spectral bands with similar spectral widths (powders, crystalline proteins,

etc.). Recently, Cordero et al.^[29] demonstrated that this technique may also be applicable for biological tissue, demonstrating SERDS to correct background from chicken meat. They concluded that SERDS could be a good choice for background correction when backgrounds are too complex to be estimated, and when signal-to-noise of the spectra is not a limiting factor.^[29]

Software-based strategies aim at algorithms to accurately subtract luminescence background signals from tissue spectra. Such algorithms have been widely developed and tested for the fingerprint region of the Raman spectrum (400 to 2,000 cm^{-1}), including methods based on wavelet transformation,^[21,22] iterative weighted least squares,^[23] geometric approach,^[24] polynomial baseline fitting (PBF),^[7–9,25] iterative morphological and mollifier-based baseline correction,^[26] auto-adaptive background,^[27] and genetic algorithm-based cubic spline smoothing.^[28]

For the high-wavenumber region (2,500 to 4,000 cm^{-1}), these methods have shown limited value, mainly because of limited knowledge about the shape of the luminescence backgrounds of high-wavenumber Raman spectra of human tissues. The high-wavenumber region consists of a small number of relatively broad and overlapping bands, which makes the estimation of the background signals much more difficult than for the fingerprint region. Polynomial subtractions of first and fifth orders are the few and most common solutions mentioned for background correction in the high-wavenumber region.^[3,10–15] First-order polynomial subtraction has been used on high-wavenumber Raman spectra acquired from oral cavity (gingiva, buccal, dorsal tongue, and palate), adenomatous polyps, esophageal squamous cell carcinoma, nasopharyngeal tissues, and breast tissue.^[3,10,12–15] The spectra in these studies were acquired with fiber-optic Raman probes and a confocal Raman microscope. The excitation laser wavelengths used were 785 and 1,064 nm. A fifth-order polynomial subtraction method has been reported for high-wavenumber Raman spectra of urine samples, using a confocal Raman system with a 785 nm laser.^[11]

Raman spectroscopy in the high-wavenumber region demonstrates significant advantages for *ex vivo* and *in vivo* medical applications when compared to the commonly used fingerprint region. The intensity of the signal is higher, and therefore, measurements can be performed in shorter integration times. Additionally, much used fiber-optic materials such as fused silica do not have Raman signal at the high-wavenumber region of the Raman spectrum.^[30] This enables the use of simple and cheap single fiber optic probes that can be easily inserted in endoscopes and needles to perform *in vivo* measurements in hollow organs or surface assessment of solid organs such as oral cavity,^[31,32] lung,^[33,34] upper gastrointestinal tract,^[35–37] and colorectal.^[14,38]

A study by Koljenović et al. showed that, for discriminating cancer from healthy tissue, essentially the same diagnostic information is obtained in either of the two spectral regions (fingerprint or high-wavenumber).^[30] Although the high-wavenumber spectra show considerably less interference from tissue luminescence compared to spectra in the fingerprint region, the background luminescence signal in the high-wavenumber can still be some orders of magnitude higher than the Raman signal itself.

This study reports a more elaborate method for luminescence background subtraction from tissue spectra in the high-wavenumber region. The method is based on a detailed characterization of the shapes of the luminescence background signals in ex vivo high-wavenumber Raman spectra of freshly excised oral cavity tissue. The aim of this study was to characterize the luminescence background and to compare the performance of two algorithms for correction of these backgrounds: the most used algorithm (PBF) and a multiple regression fitting (MFR) based method.

2 | MATERIALS AND METHODS

2.1 | Instrumentation and spectral preprocessing

For all Raman experiments, a confocal Raman microscope was used. The setup comprised a multichannel Raman Module (HPRM 2500, RiverD International B.V., The Netherlands), a 671 nm laser (CL671-150-SO, CrystaLaser), and a CCD camera fitted with a back-illuminated deep depletion CDD-chip (Andor 316, Andor Technology Ltd., UK). All the different parts of the instrument are described in detail in earlier work.^[39,40] All spectra acquired with this setup were calibrated to a relative wavenumber axis and corrected for the wavelength dependent detection efficiency of the setup. The calibration was performed following the instructions of the spectrometer supplier (RiverD International B.V., The Netherlands). Spectral preprocessing was performed by removal of cosmic ray events and Raman signal generated in the instrument (caused by the passage of laser light through optical components), which was recorded by measuring a Raman spectrum without a sample being present.

2.2 | Luminescence background signal

Bleaching experiments were performed to evaluate the shapes of the luminescence background signals that are present in ex vivo high-wavenumber human tissue spectra. The bleaching was induced by continuous exposure of the samples to the focused Raman excitation laser light. This produced spectra with an almost constant Raman

contribution and decaying luminescence contributions. In this subsection, the bleaching experiments and the method used to extract the luminescence background signals from these experiments are described.

2.2.1 | Tissue samples

Ex vivo experiments were performed on fresh resection specimens from patients undergoing surgery for oral cavity squamous cell carcinoma. Informed consent was obtained from the patients prior to the surgical procedure. The study protocol has been approved by the Medical Ethics Committee of the Erasmus MC, University Medical Center Rotterdam, The Netherlands (MEC-2013-345). The surgeon brought the specimen right after resection to the pathology department. The pathologist performed assessment of resection margins by cutting the specimen, perpendicularly to the resection surface, in about 5-mm-thick cross sections. After this intraoperative assessment, one of the cross sections was chosen by the pathologist for measurements, and the remaining cross sections were immersed in formalin. Blood was rinsed with physiological salt solution (0.9% NaCl) and gently patted dry with gauze. The section was inserted in a cartridge with the tissue in contact with a fused silica window on one side of the cartridge. The window allowed scanning of a 3×3 cm tissue area. Experiment time was limited to 60 min, after which the section was immersed in formalin and together with the rest of the resection specimen, and followed the standard pathology workflow.

2.2.2 | Bleaching experiments to isolate luminescence background signals

For bleaching experiments, Raman spectra were collected from different tissue locations on the freshly excised oral cavity tissue sections. Approximately 80 mW of laser light was focused to a spot of 4 μ m in diameter. Up to 120 spectra with 1 s exposure time were obtained at each measurement location.

In the experiments, breakdown of the luminescence molecules was induced by continuous exposure to the focused Raman excitation laser light. Time decay of the luminescence component of the signal and time invariance of the Raman component were determined by calculating difference spectra between consecutive spectra in a bleaching experiment. If difference spectra showed a broad background signal without observable Raman features, the difference between the first spectrum (spectrum with the highest background) and last spectrum (spectrum with the lowest background) was used to estimate the luminescence background signal. If Raman features were present in the difference spectra between the first spectrum and consecutive spectra, the experiment was discarded.

2.2.3 | Optimization of polynomial approximation to the background

The background signals estimated from each bleaching measurement were normalized, and polynomials of increasing orders from 1st to 10th order were fitted to the normalized backgrounds by the classical least-squares method.^[41] The norm of the residuals was plotted as a function of polynomial order. The optimal polynomial order was defined as the order after which there was no more significant decrease in residual (<5%).

2.3 | Bovine serum albumin (BSA) reference spectra

BSA was purchased from Sigma-Aldrich (A9647, 66 kDa). Eleven solutions of BSA and water were prepared with mass percentages from 5% to 40%. Per solution, 60 spectra were acquired with an exposure time of 1 s and then averaged, resulting in a set of reference spectra characterized by a low background signal and varying fractions of protein and water (CH_3 stretching vibrations, 2,910–2,965 cm^{-1} , and OH stretching, 3,350–3,550 cm^{-1}). First-order polynomial subtraction was used to subtract the little existing background signal.

2.4 | Tissue reference spectra

A total of 10,147 tissue spectra, collected in Raman mapping experiments described earlier,^[40] were used to create a set of tissue reference spectra with low luminescence backgrounds and a high variance with respect to Raman content. Ratios between peak content and background content were calculated based on the 2,910–2,965 cm^{-1} CH_3 stretching band. A first-order polynomial baseline was fitted between the spectral points at 2910 and 2965 cm^{-1} . Peak content was calculated as the integrated area above this baseline. Background was calculated as the integrated area below the baseline. The spectra with the 25% highest CH_3 content to background ratio were selected for the reference set. These spectra were scaled on the mean of the data set by an extended multiplicative scatter correction algorithm.^[41,42] Hierarchical cluster analysis was used to find the largest clusters in the group of the spectra with the 25% highest CH_3 content to background ratio. Principal component analysis was first used for data reduction.^[43] The clustering method was Ward's agglomerative algorithm with $1 - R^2$ as the distance metric, where R^2 is the squared Pearson's correlation coefficient. The result of the hierarchical cluster analysis is a membership matrix $N \times N$ (N is the number of spectra) that represents the clustering at each level of agglomeration.^[44] The largest clusters, composing 90% of the spectra

(from the group of spectra with the 25% highest CH_3 content to background ratio), were selected as a representation of the most common tissue structures present in the oral cavity with the lowest luminescence background. Cluster averages were calculated, and the residual background subtracted by first-order PBF.

2.5 | Algorithms for luminescence background subtraction

Two algorithms were compared for characterization and subsequent subtraction of luminescence background signals: a PBF algorithm and an algorithm based on MRF.

2.5.1 | PBF algorithm

The PBF algorithm used is an iterative function that fits a polynomial, of a user-specified order, through a selected set of spectral points. In the first iteration, all points of the spectrum are used, and in each iteration, the number of spectral points is reduced by only including spectral points with lower intensity than the fitted polynomial at those points. In this way, the polynomial is iteratively adapted to fit the lowest points of the spectrum. The iteration is stopped when the number of points used for the baseline fit is below a threshold set by the user. Finally, the offset of fitted baseline polynomial is adapted to ensure that the value of the fitted baseline was never above the value of the spectrum over the whole spectral range.

2.5.2 | MRF algorithm

In the MRF method, the spectra were fitted with a set of library spectra (independent variables), that have low background, and a polynomial, of a user-specified order, using a nonnegative least squares method (constraining fit coefficients to greater or equal to zero).

2.6 | Algorithm performance evaluation

The performance of the PBF and MRF algorithms was evaluated using two artificial data test sets that were constructed from the extracted luminescence background spectra, BSA reference spectra, and tissue reference spectra described above. Different luminescence background spectra were digitally added to the reference sets, which resulted in test spectra for which the luminescence background component was exactly known. The result of the luminescence background correction methods was compared to the corresponding reference spectrum (without artificially added luminescence background). Performance of the two background correction

algorithms was evaluated for different values of the polynomial order to define optimal values for both methods. Evaluation criteria were the similarity between the background corrected spectrum and the original spectrum using Pearson's correlation coefficient, and second, the influence of the background correction on the calculated water percentage. The water concentration was calculated from the ratio of the Raman bands at 3,390 and 2,935 cm^{-1} , as described by Caspers et al.^[45]

3 | RESULTS

3.1 | Luminescence background signal data set

3.1.1 | Tissue samples & bleaching experiments

Bleaching experiments were performed to isolate the luminescence background signals that are present in ex vivo tissue spectra. Freshly excised oral cavity specimens from six patients were analyzed on 35 different

tissue locations. An example of a bleaching experiment is presented in the Figure 1a.

3.1.2 | Optimization of polynomial approximation to the background

Table S1 shows the residuals of the polynomial fits to the 35 different luminescence background spectra and the optimal polynomial order. The results for orders 1 to 7 are shown in the table. Three out of the 35 luminescence backgrounds were optimally approximated by a first-order polynomial (8%), 22 background signals were optimally approximated by a second-order polynomial (63%), eight by a third-order polynomial (23%), and two by fourth-order polynomial (6%). An example of the polynomial approximation to the background can be seen in the Figure 1b–d.

3.2 | BSA test data set

Figure 2 shows the reference set of 11 spectra of BSA solutions with different mass percentages (4.92%, 5.91%,

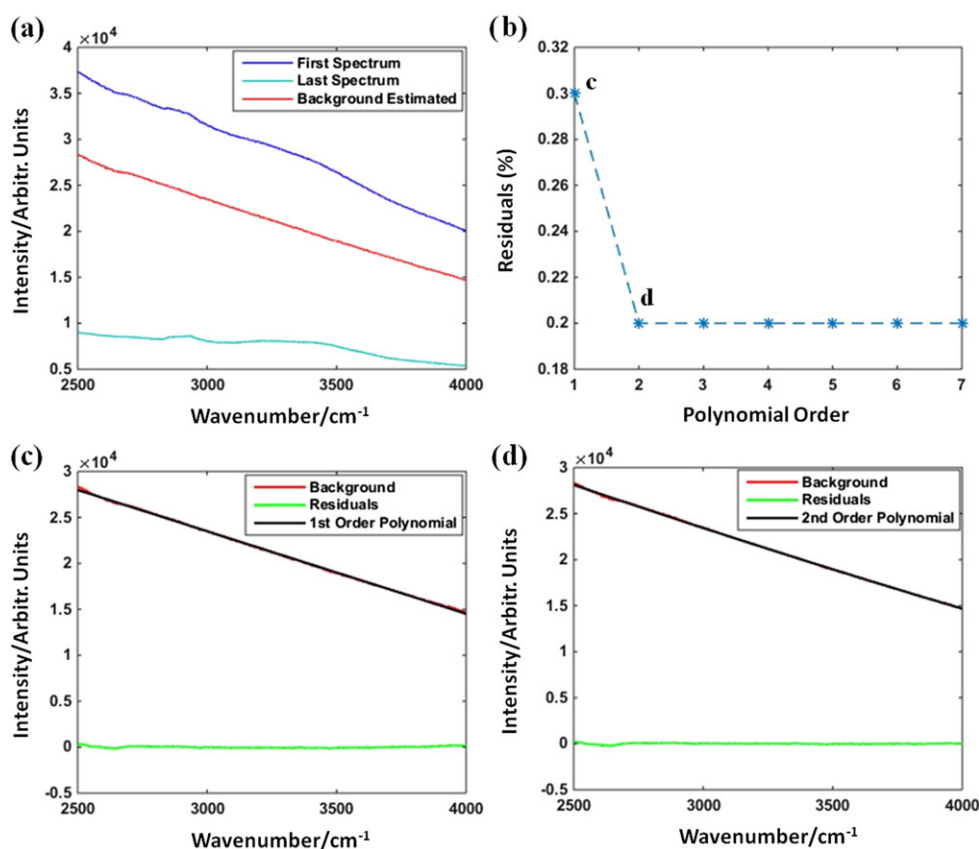


FIGURE 1 Example of background estimation from a bleaching experiment. (a) Bleaching time-series: first spectrum (dark blue), last spectrum (light blue), and the difference between the two as the estimated background (red). (b) Residuals for different polynomial order fits to the estimated background. (c) Estimated background (red), fitted first-order polynomial (black), and residual (green). (d) Estimated background (red), optimal (second order) polynomial (black), and residual (green). The optimal polynomial order is defined as the minimal order after which there was no more significant decrease in residual ($<5\%$) [Colour figure can be viewed at wileyonlinelibrary.com]

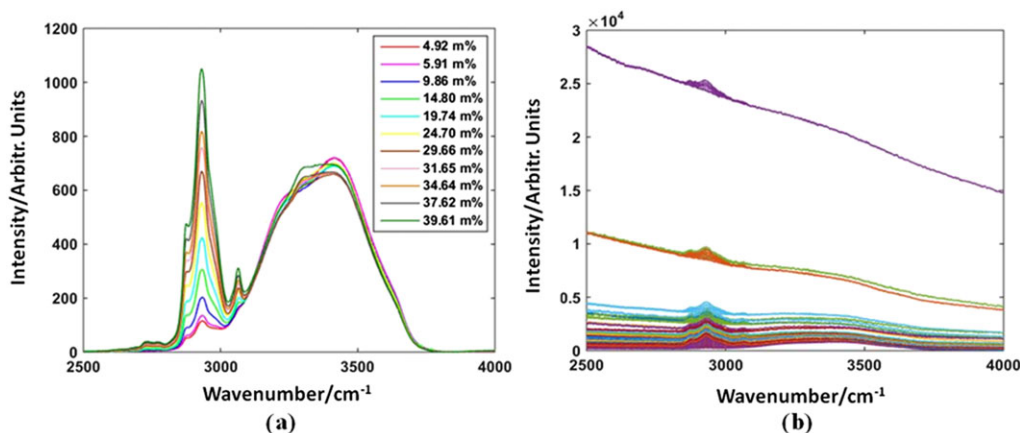


FIGURE 2 Bovine serum albumin (BSA) reference spectra data set and BSA spectra test set. (a) Averaged Raman high-wavenumber spectra from different albumin solutions. The mass percentage (m%) of the solutions varied between 5% and 40%. For better visualization of Raman signal, the low intensity background was subtracted as a first order polynomial. (b) BSA spectra test set: 280 Raman spectra with BSA signal and luminescence content [Colour figure can be viewed at wileyonlinelibrary.com]

9.86%, 14.80%, 19.74%, 24.70%, 29.66%, 31.65%, 34.64%, 37.62%, and 39.61%). Eight spectra were used for testing of the PBF and MRF algorithms. The remaining three spectra were used as independent data in the MRF algorithm. Each of the eight BSA reference spectra was combined with all 35 different background luminescence spectra, resulting in a test set of 280 spectra shown in Figure 2b.

3.3 | Tissue test data set

Nineteen tissue reference spectra with low luminescence background were obtained, which are shown in Figure 3a. Thirteen spectra were used for testing of the PBF and MRF algorithms. The remaining six spectra were used as independent data in the MRF algorithm, and each of the

13 tissue reference spectra was combined with all 35 different background luminescence spectra, resulting in a test set of 455 spectra shown in Figure 3b.

3.4 | Algorithm performance evaluation

3.4.1 | BSA test data set

For all 280 spectra of the BSA test data set, the background was estimated with the two algorithms using a first-, second-, third-, fourth-, and fifth-polynomial orders and subtracted. The similarity of the background-corrected data with the original data is listed in Table 1. The mean and the standard deviation of the Pearson's correlation coefficients are shown per test data set and per algorithm used for testing. The table also shows the

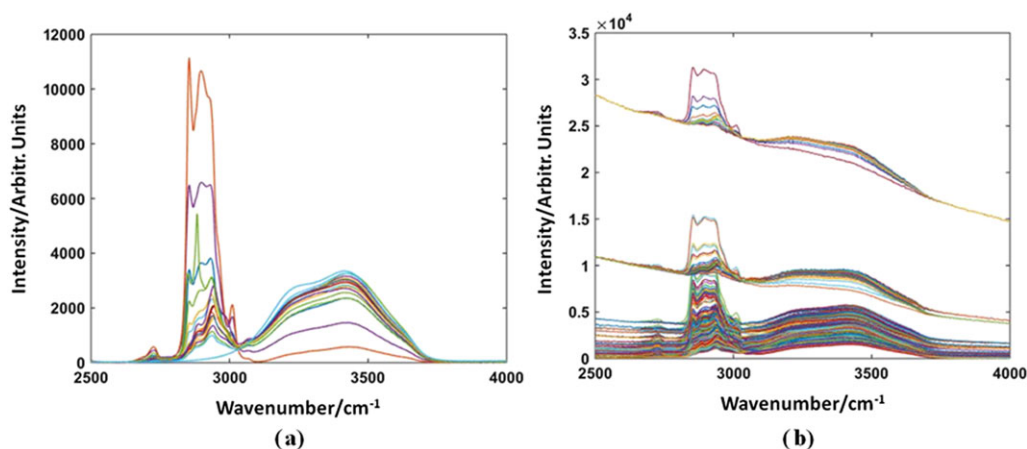


FIGURE 3 Ex vivo high-wavenumber tissue reference spectra data set and high-wavenumber tissue spectra test set. (a) Nineteen ex vivo high-wavenumber tissue reference spectra with low luminescence background were obtained after cluster analysis of the low luminescence spectra in the ex vivo tissue spectra data set. For better visualization of Raman signal, the low intensity background was subtracted as a first order polynomial. (b) Ex vivo high-wavenumber tissue spectra test set: 455 Raman spectra with known Raman and luminescence content [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Evaluation of the performance of the polynomial baseline fitting (PBF) and multiple regression fitting (MRF) algorithms on bovine serum albumin (BSA) spectra and on ex vivo high-wavenumber tissue spectra. Average \pm standard deviation of the Pearson's correlation coefficient between background corrected spectra and their reference spectra for the two different correction algorithms (PBF and MRF) and different values of the algorithm background polynomial. Average \pm standard deviation of the absolute error in water concentration (%) between background corrected spectra and their reference spectra for the two different background correction algorithms and different values of the algorithm background polynomial

Background polynomial	BSA spectra			
	Pearson's correlation coefficient		Absolute error in H ₂ O (%)	
	PBF	MRF	PBF	MRF
1st order	0.973 \pm 0.076	0.974 \pm 0.075	2.74 \pm 10.60	2.74 \pm 10.60
2nd order	0.995 \pm 0.009	0.996 \pm 0.006	0.64 \pm 0.57	0.61 \pm 0.52
3rd order	0.993 \pm 0.021	0.996 \pm 0.006	0.74 \pm 1.22	0.56 \pm 0.48
4th order	0.747 \pm 0.136	0.996 \pm 0.007	29.24 \pm 34.10	0.58 \pm 0.54
5th order	0.516 \pm 0.197	0.996 \pm 0.006	46.54 \pm 16.39	0.58 \pm 0.52
Background polynomial	Tissue spectra			
	Pearson's correlation coefficient		Absolute error in H ₂ O (%)	
	PBF	MRF	PBF	MRF
1st order	0.996 \pm 0.008	0.996 \pm 0.007	0.81 \pm 1.30	0.81 \pm 1.30
2nd order	0.996 \pm 0.010	0.997 \pm 0.007	0.62 \pm 1.00	0.52 \pm 0.72
3rd order	0.996 \pm 0.010	0.997 \pm 0.006	0.60 \pm 0.82	0.48 \pm 0.54
4th order	0.728 \pm 0.140	0.997 \pm 0.007	12.99 \pm 22.30	0.48 \pm 0.53
5th order	0.578 \pm 0.200	0.997 \pm 0.007	37.00 \pm 13.97	0.49 \pm 0.52

mean and the standard deviations of the differences in calculated water concentrations.

Examples for the PBF algorithm with first-, third-, and fourth-polynomial orders are shown in Figure 4a–c. The distributions of the corresponding Pearson's correlation coefficients are shown in Figure 4d. Examples for the MRF algorithm with first-, second-, and fourth-polynomial orders are shown in Figure 4e–g. The distributions of the corresponding Pearson's correlation coefficients are shown in Figure 4h.

3.4.2 | Tissue reference test data set

For all 455 spectra of the tissue test data set, the background was estimated with the two algorithms using a first-, second-, third, fourth-, and fifth-polynomial orders. The similarity of the background-corrected data with the original data and the errors in calculated water concentration after background correction are listed in Table 1. Two examples of the results obtained using PBF algorithm and MRF algorithm are shown in Figure 6a–d.

4 | DISCUSSION & CONCLUSIONS

When measuring Raman spectra from biological tissue, high luminescence backgrounds can (almost) completely

obscure the Raman signal. Both hardware- and software-based strategies have been developed to correct for these luminescence background signals.

Hardware-based strategies can be efficient, but their implementation is associated with higher instrument complexity and associated with higher costs. Software-based strategies are easier to implement and have lower costs. Most software-based strategies have been developed for the fingerprint region. In contrast to the fingerprint region, not many background subtraction algorithms have been developed for the high-wavenumber region of the Raman spectrum.^[7,9,21–28,46]

For high-wavenumber region spectra, PBF is the only solution mentioned for background correction, using either a first-order^[3,10,12–15] or a fifth-order polynomials.^[11] However, this method produced unsatisfactory results for our ex vivo high-wavenumber spectra of human oral tissues. First, we experienced that not all luminescence backgrounds could be characterized by a first-order polynomial, and the estimation of the background by higher order polynomials did not produce stable results. We observed that the polynomial fits have problems in distinguishing between luminescence background and the rather broad band Raman spectral features in the high-wavenumber region.

Therefore, in this article, we present a new and effective method for correction of luminescence backgrounds

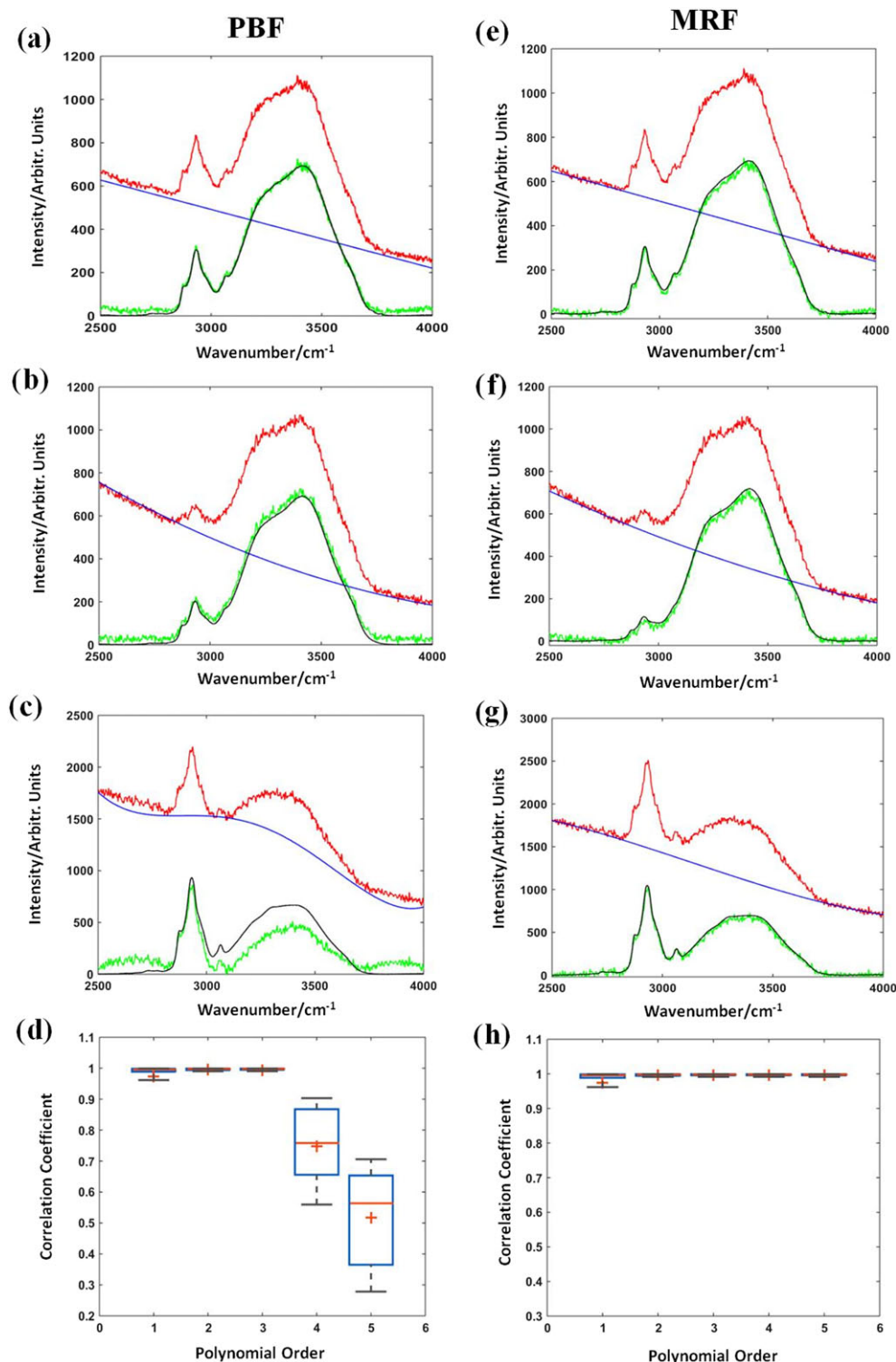


FIGURE 4 Luminescence background correction on bovine serum albumin test data using the polynomial baseline fitting (PBF) and multiple regression fitting (MRF) algorithms. PBF results are shown in the left column and MRF results in the right column. (a) Example of PBF first-order polynomial background subtraction. (b) PBF second-order polynomial. (c) PBF fourth-order polynomial. (d) Box plot of the Pearson's correlation coefficients for PBF with different polynomial orders. (e) Example of MRF first-order polynomial background subtraction. (f) MRF second-order polynomial. (g) MRF fourth-order polynomial. Spectra displayed: test spectrum (red), estimated background (blue), corrected spectrum (green), and corresponding reference spectrum (black) [Colour figure can be viewed at wileyonlinelibrary.com]

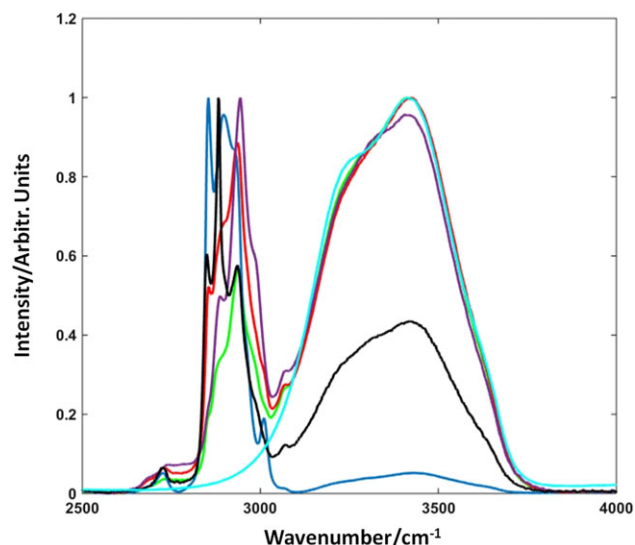


FIGURE 5 The six ex vivo high-wavenumber tissue reference spectra that were used as independent variables in the MRF algorithm. For better visualization of the differences in the Raman signal, the spectra were normalized [Colour figure can be viewed at wileyonlinelibrary.com]

in high-wavenumber region Raman spectra. The method is based on MRF and on high-wavenumber luminescence-free Raman signal. The goal of this study was to characterize the luminescence background signals of ex vivo human oral tissue high-wavenumber region spectra and to compare the performance of the MRF method with the more common PBF.

The luminescence background signals were measured in laser-induced bleaching experiments on freshly excised human oral tissue. The bleaching process was not complete for all the bleaching experiments performed, due to time constraints (fresh specimen(s) that needed to be sent to pathology department). Although the estimated backgrounds did not bleach completely, we do not believe that this affects the outcome of this study in any way.

Polynomials of different orders were required for optimal approximation of the luminescence backgrounds, ranging from first order (8% of the cases), second order (63% of the cases), third order (23%), to fourth order (6%). These results indicate that it is not possible to choose

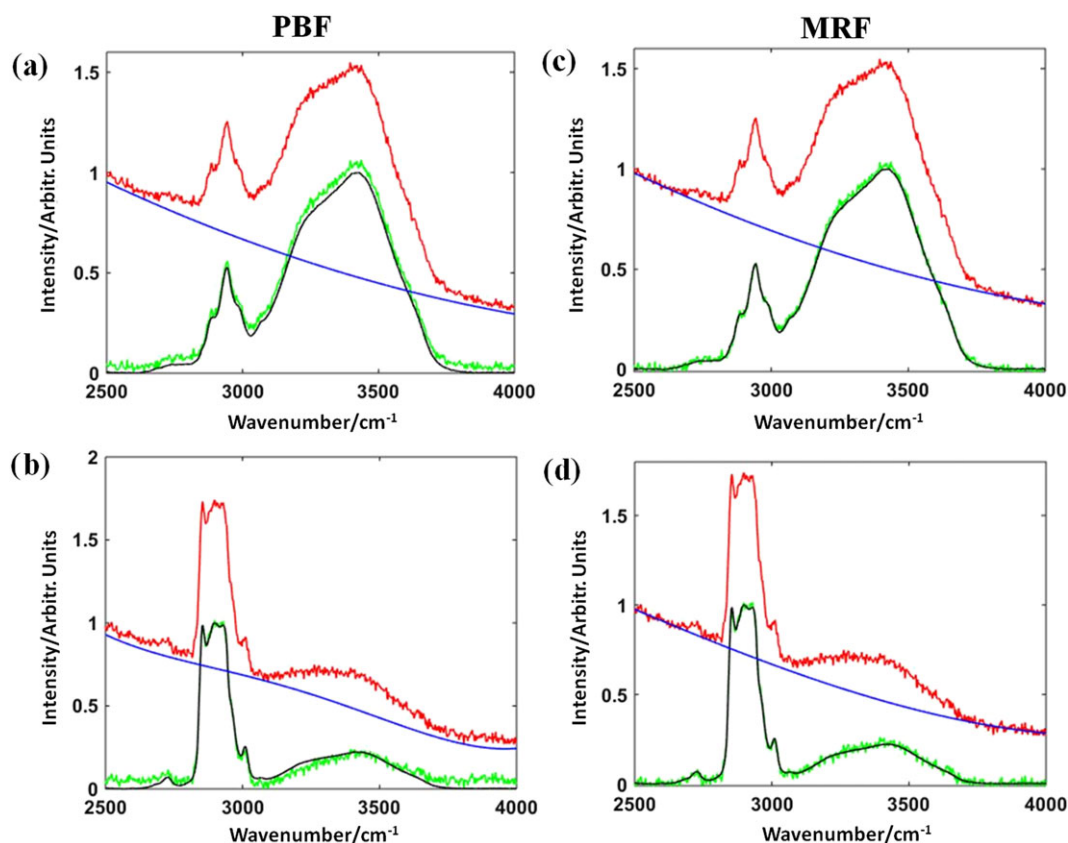


FIGURE 6 Luminescence background correction on tissue test data using the polynomial baseline fitting (PBF) and multiple regression fitting (MRF) algorithms. PBF results are shown in the left column and MRF results in the right column. (a) Example of PBF second-order polynomial background subtraction. (b) PBF fourth-order polynomial. (c) MRF fourth-order polynomial. (d) MRF fourth-order polynomial. Test spectrum (red), estimated background (blue), corrected spectrum (green), and corresponding reference spectrum (black) [Colour figure can be viewed at wileyonlinelibrary.com]

a single optimal polynomial order that will give the best background estimation for all cases. Hence, operator-supervision is required, and interoperator subjectivity may play a role. Moreover, the selection of a polynomial order that is not optimal can result in significant artifacts in the Raman spectrum from residual luminescence contributions or from removal of Raman signal by polynomial overfitting.

The luminescence background signals obtained from the bleaching experiments were used to generate test data sets of spectra with known luminescence and Raman contributions (Figures 2 and 3). For the Raman contributions, two different data sets were used: a set of spectra from pure BSA solutions with different concentrations and fresh ex vivo tissue spectra from oral cavity tissue with very low luminescence background. Because the luminescence backgrounds in the test data sets were digitally added, the test data enabled accurate evaluation of the performance of the background subtraction algorithms (Figures 4-6). Evaluation criteria were spectral correlation and error in the calculated water content after background correction, as compared to the corresponding reference spectra. The results in Table 1 show that for both data sets, MRF performs equally well or better than PBF for all the polynomial background orders evaluated. The results test show that the MRF is more robust and more effective than PBF in accurately estimating the luminescence background in high-wavenumber region tissue and BSA spectra. Also here, it is clear that the MRF algorithm gives significantly better results when higher order polynomials are used in the fitting process. For the tissue reference test data set, the MRF algorithms gives better results and again especially for the higher order polynomials.

For the classical PBF method, the overall best results were obtained by fitting a second-order polynomial. This is mainly determined by the observation that the luminescence backgrounds obtained from the bleaching experiments predominantly had a second-order polynomial shape. Higher polynomial orders lead to large errors due to overfitting artifacts. This is a common risk in polynomial background fitting of high-wavenumber region Raman spectra of tissues because these spectra are characterized by broad partially overlapping bands, and because no reliable baseline points are available to anchor the central part of the spectrum around 2,800–3,800 cm^{-1} . As a result, the broad high-wavenumber Raman bands are partly fitted as background by higher order polynomials (see Figure 4c).

For MRF, second-order polynomials or higher gave the best results. This confirms that the MRF-method is robust and can be reliably used with higher order polynomials, because the Raman signal is fitted by the

independent spectra included in the method. Thus, higher order polynomials do not result in fitting of the Raman signal as background. The MRF algorithm is not dependent of the shape of the Raman signal, provided that the independent data set of luminescence-free Raman spectra is representative of the Raman signal variance measured. This effectively eliminates subjectivity in choosing an optimal polynomial order, as required by the PBF method, and the associated problem that a single optimal polynomial order does not exist for a variety of backgrounds, as shown in the Section 3.

The results of this study were obtained on high-wavenumber region Raman spectra. However, we believe that the method may not be limited to this spectral region but may prove equally useful for luminescence background correction in the fingerprint region. Further studies will be conducted to prove the usefulness of MRF for luminescence background subtraction in the fingerprint region.

In conclusion, this study has demonstrated that MRF is a more accurate and robust solution than PBF for correction of luminescence backgrounds and can be used for unsupervised and automated background subtraction as needed in real time tissue diagnostic applications.

ORCID

E.M. Barroso  <http://orcid.org/0000-0001-7255-2881>

I.P. Santos  <http://orcid.org/0000-0003-2463-246X>

REFERENCES

- [1] C. Kallaway, L. M. Almond, H. Barr, *Photodiagnosis Photodyn. Ther.* **2013**, 3, 10.
- [2] B. Broadbent, J. Tseng, R. Kast, *J. Neurooncol* **2016**, 1, 130.
- [3] J. Wang, K. Lin, W. Zheng, *Sci. Rep.* **2015**, August, 5.
- [4] M. Çulha, *Bioanalysis* **2015**, 21, 7.
- [5] I. Pence, A. Mahadevan-Jansen, *Chem. Soc. Rev.* **2016**, 7, 45.
- [6] I. P. Santos, P. J. Caspers, T. C. Bakker Schut, *J Raman Spectrosc.* **2015**, 7, 46.
- [7] C. Gallo, V. Capozzi, M. Lasalvia, *Vib. Spectrosc.* **2016**, 83, 132.
- [8] B. D. Beier, A. J. Berger, *Analyst* **2009**, 6, 134.
- [9] C. A. Lieber, A. Mahadevan-Jansen, *Appl. Spectrosc.* **2003**, 11, 57.
- [10] W. Huang, S. Wu, M. Chen, *J Raman Spectrosc.* **2015**, 6, 46.
- [11] E. Brindha, R. Rajasekaran, P. Aruna, *Spectrochim Acta-Part A Mol Biomol Spectrosc.* **2017**, 171.
- [12] A. F. García-Flores, L. Raniero, R. A. Canevari, *Theor. Chem. Acc.* **2011**, 4-6, 130.
- [13] W. Liu, Z. Sun, J. Chen, *J. Spectrosc.* **2016**, 2016.
- [14] M. S. Bergholt, K. Lin, J. Wang, *J. Biophotonics* **2016**, 4, 9.
- [15] M. S. Bergholt, W. Zheng, Z. Huang, *J. Biomed. Opt.* **2013**, 3, 18.

- [16] I. P. Santos, P. J. Caspers, T. C. Bakker Schut, *Anal. Chem.* **2016**, *15*, 88.
- [17] Z. Huang, H. Zeng, I. Hamzavi, *J. Biomed. Opt.* **2006**, *3*, 11.
- [18] J. R. Lakowicz, *Principles of Spectroscopy*, MA, Springer US, Boston **2006**.
- [19] L. F. C. S. Carvalho, F. Bonnier, K. O'Callaghan, *Proc. SPIE 9531, Biophotonics South America* **2015**, 9531.
- [20] S. T. McCain, R. M. Willett, D. J. Brady, *Opt. Express* **2008**, *15*, 16.
- [21] C. M. Galloway, E. C. Le Ru, P. G. Etchegoin, *Appl. Spectrosc.* **2009**, *12*, 63.
- [22] Z. M. Zhang, S. Chen, Y. Z. Liang, *J Raman Spectrosc.* **2010**, *6*, 41.
- [23] H. Ruan, L. K. Dai, *Asian J. Chem.* **2011**, *12*, 23.
- [24] N. Kourkoumelis, A. Polymeros, M. Tzaphlidou, *Spectrosc An Int J.* **2012**, 5-6, 27.
- [25] T. Wang, L. Dai, *Appl. Spectrosc.* **2016**, *0*, 0.
- [26] M. Koch, C. Suhr, B. Roth, *J Raman Spectrosc* **2016**; July 2016.
- [27] Y. Xie, L. Yang, X. Sun, *Spectrochim Acta - Part A Mol Biomol Spectrosc.* **2016**, 161.
- [28] S. He, S. Fang, X. Liu, *Chemom. Intel. Lab. Syst.* **2016**, *152*, 1.
- [29] E. Cordero, F. Korinth, C. Stiebing, *Sensors* **2017**, 1724.
- [30] S. Koljenović, T. C. Bakker Schut, R. Wolthuis, *J. Biomed. Opt.* **2005**, *3*, 10.
- [31] H. Krishna, S. K. Majumder, P. Chaturvedi, *J. Biophotonics* **2014**, *9*, 7.
- [32] K. Guze, H. C. Pawluk, M. Short, *Head Neck* **2015**, *4*, 37.
- [33] M. A. Short, S. Lam, A. McWilliams, *Opt. Lett.* **2008**, *7*, 33.
- [34] H. C. McGregor, M. A. Short, A. McWilliams, *J. Biophotonics* **2017**, *1*, 10.
- [35] S. K. Teh, W. Zheng, K. Y. Ho, *Br. J. Cancer* **2008**, *2*, 98.
- [36] M. S. Bergholt, W. Zheng, K. Lin, *Analyst* **2010**, *12*, 135.
- [37] S. Duraipandian, M. S. Bergholt, W. Zheng, *J. Biomed. Opt.* **2012**, *10*, 17.
- [38] M. A. Short, I. T. Tai, D. Owen, *Opt. Express* **2013**, *4*, 21.
- [39] E. M. Barroso, R. W. H. Smits, T. C. Bakker Schut, *Anal. Chem.* **2015**, *4*, 87.
- [40] E. M. Barroso, R. W. H. Smits, C. G. F. van Lanschot, *Cancer Res.* **2016**, *20*, 76.
- [41] T. J. Vickers, R. E. J. Wambles, C. K. Mann, *Appl. Spectrosc.* **2001**, *4*, 55.
- [42] H. Martens, E. Stark, *J. Pharm. Biomed. Anal.* **1991**, *8*, 9.
- [43] I. T. Jolliffe, *Principal Component Analysis*, Second ed., Springer-Verlag, New York **2002**.
- [44] A. K. Jain, M. N. Murty, P. J. Flynn, *ACM Comput Surv.* **1999**, *3*, 31.
- [45] P. J. Caspers, G. W. Lucassen, E. A. Carter, *J. Invest. Dermatol.* **2001**, *3*, 116.
- [46] F. Bonnier, S. M. Ali, P. Knief, *Vib. Spectrosc.* **2012**, *61*, 124.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Barroso EM, Bakker Schut TC, Caspers PJ, et al. Characterization and subtraction of luminescence background signals in high-wavenumber Raman spectra of human tissue. *J Raman Spectrosc.* 2018;49:699–709. <https://doi.org/10.1002/jrs.5338>