

A robust method for automated background subtraction of tissue fluorescence

Alex Cao,^{1,2*} Abhilash K. Pandya,² Gulay K. Serhatkulu,² Rachel E. Weber,³ Houbei Dai,³ Jagdish S. Thakur,² Vaman M. Naik,⁴ Ratna Naik,³ Gregory W. Auner,² Raja Rabah⁵ and D. Carl Freeman⁶

¹ Department of Pediatric Surgery, Children's Hospital of Michigan, Detroit, MI 48201, USA

² Department of Electrical & Computer Engineering, Wayne State University, Detroit, MI 48202, USA

³ Department of Physics, Wayne State University, Detroit, MI, 48202, USA

⁴ Department of Natural Sciences, University of Michigan-Dearborn, Dearborn, MI 48128, USA

⁵ Department of Pathology, Children's Hospital of Michigan, Detroit, MI 48201, USA

⁶ Department of Biological Science, Wayne State University, Detroit, MI 48202, USA

Received 17 May 2006; Accepted 5 March 2007

This paper introduces a new robust method for the removal of background tissue fluorescence from Raman spectra. Raman spectra consist of noise, fluorescence and Raman scattering. In order to extract the Raman scattering, both noise and background fluorescence must be removed, ideally without human intervention and preserving the original data. We describe the rationale behind our robust background subtraction method, determine the parameters of the method and validate it using a Raman phantom against other methods currently used. We also statistically compare the methods using the residual mean square (RMS) with a fluorescence-to-signal (F/S) ratio ranging from 0.1 to 1000. The method, 'adaptive minmax', chooses the subtraction method based on the F/S ratio. It uses multiple fits of different orders to maximize each polynomial fit. The results show that the adaptive minmax method was significantly better than any single polynomial fit across all F/S ratios. This method can be implemented as part of a modular automated real-time diagnostic *in vivo* Raman system. Copyright © 2007 John Wiley & Sons, Ltd.

KEYWORDS: background subtraction; fluorescence; polynomial fit; adaptive; minmax

INTRODUCTION

Raman spectra of tissues contain a combination of Raman scattering, intrinsic tissue fluorescence and noise. Over the past decade, Raman spectroscopy has emerged as a promising diagnostic tool for cancer detection.¹ The Raman scattering effect (typically very small) is unique to the chemical composition of the tissue and therefore can be used for diagnosis. One of the end goals of this field of research is to provide surgeons with a real-time *in vivo* tool in the operating room, with the ability to differentiate pathological tissue from normal tissue, to help guide the surgical excision. Such a device may help reduce operative time and cost involved in frozen section diagnosis during surgery. One of the key issues that need to be addressed before Raman spectroscopy reaches its ultimate potential is the automated background subtraction of fluorescence from the spectra. Here, we

present a robust method that takes into consideration the fluorescence-to-signal (F/S) ratio, to minimize the residual mean square (RMS) error. A comparison and evaluation with current techniques is provided.

Organs and tumors are not homogeneous, and a number of common constituents produce fluorescence when excited with light in the visible to near-infrared region. Therefore, it is important that the preprocessing of Raman signals be as accurate and as automated as possible. Several methods using hardware techniques have been used to reduce fluorescence^{2–5}; however, it is more practical and easier to deal with fluorescence using software. Ideally, the instrumentation should be optimized to minimize the amount of fluorescence. Nevertheless, fluorescence is typically present in most Raman spectra, and most background subtraction techniques require some user intervention. This leads to subjective results that vary from person to person, and therefore requires training individuals in selecting the best method to remove the background fluorescence. Subjective background subtraction on a large number of samples would

*Correspondence to: Alex Cao, Rm 3157, Department of Electrical and Computer Engineering, 5050 Anthony Wayne Drive, Detroit, MI 48202, USA. E-mail: caoa@wayne.edu

be time consuming, and inherent subjectivity may preclude comparisons (and diagnosis) among the resulting spectra. Accordingly, for real-time operation, automated fluorescence background subtraction is required.

The most promising type of background subtraction algorithms use polynomial fits because they can approximate the fluorescence profile while excluding the Raman peaks. However, there is no consensus on the best polynomial fit order for fluorescence background subtraction.

Vickers *et al.*⁶ utilize a two-step process for background subtraction – a third-order fit for background generation and a fifth-order fit for spectral smoothing. They use a piecewise procedure for the background estimation and then reassemble the segments to form the spectra. The user decides on the boundaries of the segment to meet the algorithm's collinearity requirement. Short *et al.*⁷ use a cubic spline algorithm for background fluorescence to eliminate most regions of negative Raman intensity that would be present in an n th order least-squares polynomial fit. Nodal regions were chosen in places where there appeared to be no Raman scattering. The cubic spline parameters were chosen by trial and error. The end nodes were not used, as they were estimated by extrapolation, causing distortions in the fits of noisy spectra. Different numbers of nodes and nodal regions were used depending on the sample type. Gao *et al.*⁸ used a cubic polynomial fit, while Brennan *et al.*⁹ used a fourth-order polynomial version for background subtraction. Huang *et al.*¹⁰ used a fifth-order polynomial excluding all significant Raman peaks. Lieber and Mahadevan-Jansen¹¹ propose an automated fifth-order polynomial using a least-squares fit excluding the Raman peaks. They were able to extract the Raman scattering from spectra where the ratio of fluorescence-to-Raman intensity ranged from 10:1 to 10⁷:1. The variation in methods is not only limited to between groups but also within groups. For example, in Raman studies of breast tissue, Feld's group did not consistently use a single method but rather several methods (fourth-, fifth- and sixth-order polynomial) for background subtraction.^{12–14} In addition, each tissue diagnostic group had different signal-to-noise ratios.

How researchers decide which order of fit to be used is often not clearly stated, nor is the rationale defended. It seems reasonable that they are choosing the order that best approximates the fluorescence for the spectra they have already collected. For real-time analysis, if the amount of fluorescence is known before the Raman measurement, then it makes it easier to choose which fit is to be used for the correction. But, if the magnitude of the tissue's intrinsic fluorescence is not known, then choosing the correct fit becomes problematic. Hence the need for a robust automated fluorescence background subtraction method.

This paper deals only with the background subtraction portion of the data processing using polynomial fits. See Lieber and Mahadevan-Jansen¹¹ and Schulze *et al.*¹⁵ for a more thorough list of background subtraction techniques.

We will not deal with the subtraction of noise in this paper and assume that the reference spectra are already de-noised. The method is independent of any de-noising technique (wavelets, Savitzky–Golay filter, median filters) and therefore can be coupled with the researcher's preference.

DEVELOPMENT OF THE ADAPTIVE MINMAX METHOD

Modified polynomial fit

The modified polynomial fit outlined in Lieber and Mahadevan-Jansen¹¹ seems most promising for three reasons: (1) it uses a least-squares fit that excludes the Raman peaks; (2) it works for a range of F/S ratios; and (3) it requires no user intervention. The method involves a series of curve fits and iterative re-assignment of the control points. The method begins with a polynomial fit of the original spectrum. In the next iteration, for each x -value, the minimum y -value of either the polynomial fit or the original spectrum is selected as a control point for the subsequent curve fit. The process is repeated until a specified convergence criterion is met or until the maximum number of iterations is reached. Our criterion for convergence was that the number of re-assigned points would remain constant for 10 consecutive iterations or 200 iterations, whichever came first. From here on, we will drop the term 'modified' and assume that all polynomial fits mentioned hereafter are modified unless stated otherwise.

Adaptive fit

In practice, one encounters a wide range of F/S ratios, depending on instrumentation and the sample, and it is difficult to predict the amount of correction needed beforehand for online analyses. We define the F/S ratio as the maximum fluorescence divided by the maximum Raman scattering (i.e. subtracted spectrum) signal when measured with respect to a minimum intensity value of zero. By definition, the F/S ratio remains constant if the spectrum is offset along the y -axis (i.e. intensity), as the profile (fluorescence + Raman scattering) remains identical.

Polynomial fits do not perform consistently at all F/S ratios. To illustrate, we added a Gaussian background to a normalized tissue spectrum at two different F/S ratios: 10 and 100 (Fig. 1). The outline of the spectrum can be seen at the lower F/S. On the other hand, the spectrum is saturated with fluorescence at the higher F/S. Fourth- and fifth-order polynomial fits (most commonly mentioned in literature) were applied to the spectra + fluorescence. Figure 2 shows the normalized subtracted spectra (top row) and the residuals (bottom row) of the polynomial fits. For F/S = 10 (left column), the fourth-order fit has a smaller absolute residual than the fifth-order fit. However, for F/S = 100 (right column), the fifth-order fit was superior. Accordingly, we hypothesize that an adaptive algorithm, based on an estimated F/S ratio of the spectrum, will perform better than a single-order fit across all F/S ratios.

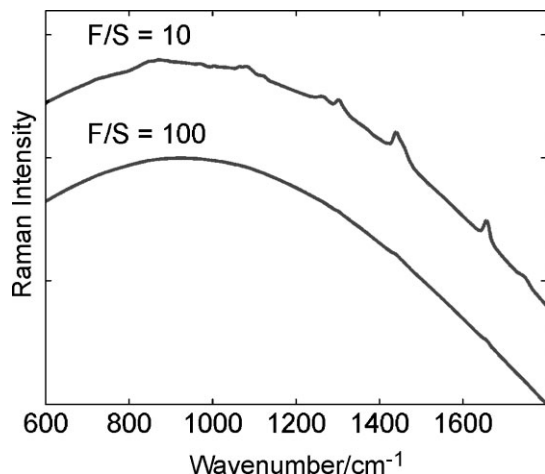


Figure 1. Sample spectrum with Gaussian background added at two different F/S ratios. The outline of the original spectrum is visible at a ratio of 10 while the original spectrum is saturated by the fluorescence at a ratio of 100.

Constrained polynomial fit

Another problem with polynomial fits that can be illustrated from the previous example is that the endpoints of the fit can suffer from instability. The fifth-order fit (Fig. 2) diverges

from the reference spectrum near 750 cm^{-1} in the lower F/S scenario and both orders diverge around 750 cm^{-1} at the higher F/S ratio. To minimize this problem, we propose to modify the polynomial fit by including the spectra endpoints in the control points to prevent the polynomial from diverging significantly from the spectrum at the two extremes (Fig. 3). The constraints help keep the fit close to the spectrum at the endpoints.

Minmax polynomial fit

There is a trade-off with the constrained polynomial fit (Fig. 3), as it introduces distortion (i.e. area under the curve) in the mid-sections of the fit as a result of the constraints. Accordingly, we propose a two-step technique that takes advantage of the benefits of each polynomial fit.

The first step involves performing both the unconstrained and constrained polynomial fit for two different orders. The fit orders will be determined by the adaptive part of the algorithm based on the F/S ratio. The second step takes the maximum value among the initial fits as the points for the final fit (Fig. 4). We name this technique the ‘minmax’ method. The ‘min’ part of the technique is the first part of the algorithm, which takes the minimum point between the spectrum and the polynomial fit. This is done to prevent overfitting the data. The ‘max’ part of the technique is the

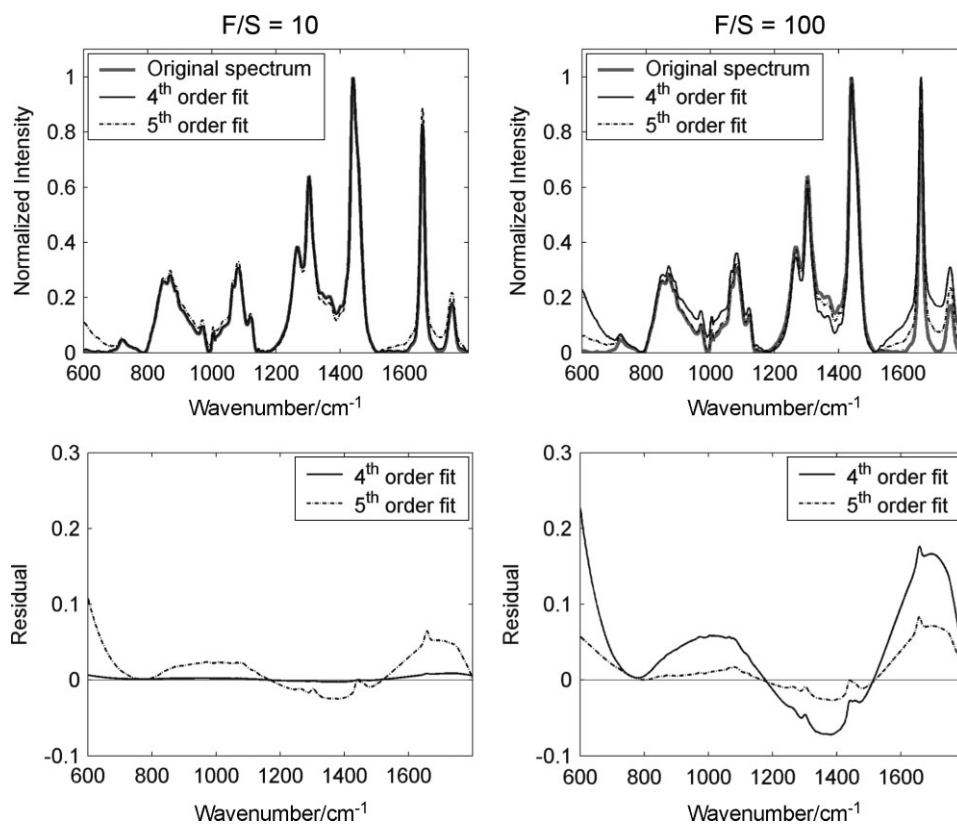


Figure 2. Polynomial fit background subtraction on the two spectra shown in Fig. 1. The top row shows the subtracted spectra compared with the original spectrum, while the bottom row shows the residuals. The left and right columns are for F/S = 10 and F/S = 100, respectively. The fourth-order fit performs better at F/S = 10 while the fifth-order fit is superior at F/S = 100.

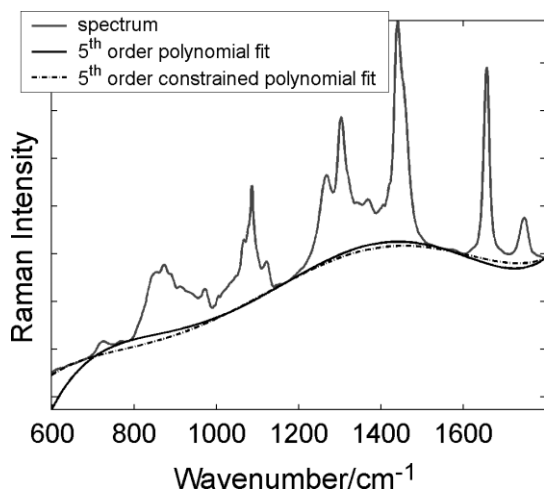


Figure 3. Unconstrained and constrained polynomial fit on a tissue spectrum. Note that the unconstrained fit diverges from the spectra at the left endpoint.

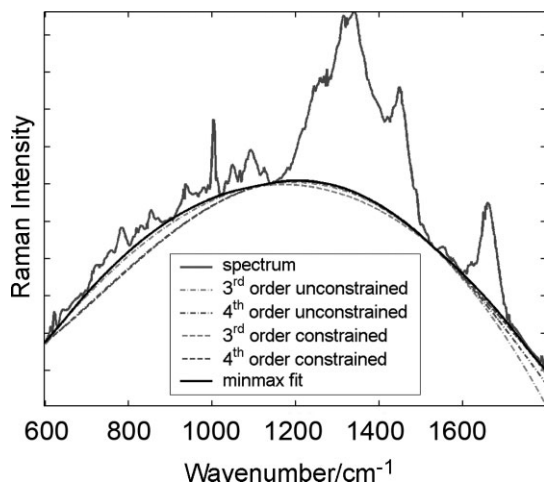


Figure 4. Example of the minmax fit on a spectrum with a Gaussian background. The fit takes the maximum value of the four polynomial fits performed as the final fit.

second part of the algorithm that takes the maximum value among all the polynomial fits performed, which prevents underfitting the data – assuming that the fit from the first part of the technique has not overfitted the data. We hypothesize that the adaptive minmax method performs better than a single-order polynomial fit across all F/S ratios.

TEST OF THE ADAPTIVE MINMAX METHOD

We used a data set of 650 Raman spectra (each with 1300 data points) collected and processed from tissues from 11 mice to test the method. Briefly, the Raman spectra were recorded with a Renishaw RM1000 Raman microscope (Renishaw, Wotton-under-Edge, UK) using a 50× objective. The 785 nm laser line was used to excite the Raman spectra, and the power

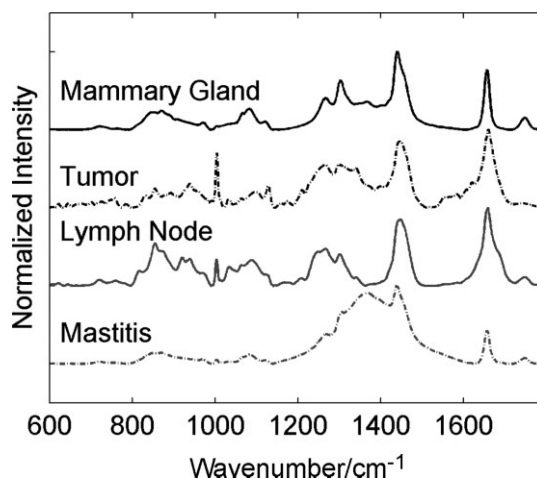


Figure 5. Representative spectra for the four types of tissues found in the mouse data set used in this study. The spectra shown have been processed and normalized. F/S ratios for this data set were found to range between 0.2 and 3.3.

used was 100 mW at the sample. Spectra were measured with a 10-s exposure time, and three exposures were averaged to obtain a spectrum. The spectral resolution was about 4 cm^{-1} .

The mice were injected on one side with a highly malignant mammary tumor cell line and sacrificed 10–14 days later. Pathologically, the data consisted of four tissue types: (# of spectra): normal mammary gland (234), tumor (220), lymph node (147) and mastitis (49). The tissues have different Raman profiles, making it impractical to use an algorithm that requires *a priori* knowledge of the spectrum when the tissue is unknown (Fig. 5). Half the data for each tissue type was used to construct the algorithm, while the other half was used for validation.

Four types of synthesized backgrounds were used in this analysis. They were taken from portions of the following four functions: Gaussian, Lorentzian, Fourier series and arctangent (Fig. 6). The function parameters were randomized to change the profile of the synthesized background fluorescence added to each of the samples to simulate sample variability. These curves were chosen because they are slowly varying functions similar in nature to background fluorescence. Here, we consider backgrounds with an interval ranging from 0.1 to 1000 to cover a wide range of fluorescence as in human and animal tissues.

After the background subtraction is performed, the subtracted spectrum is normalized to range between 0 and 1. The subtracted spectrum is then compared to the original spectrum using a goodness-of-fit measure. We calculated the square root of the RMS value (aka standard error of estimate or standard error of the regression), where the observed value was the predicted spectrum after background subtraction and the predicted value was the spectrum deemed as true at the onset. The mouse data set was assumed to represent the ‘truth’ i.e. what the spectra should look like after the

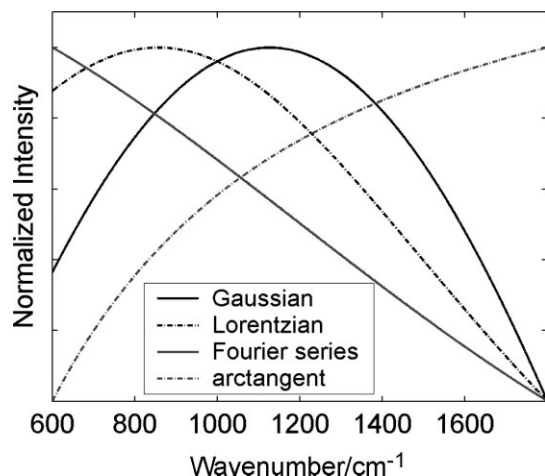


Figure 6. Sample spectra of the four synthesized backgrounds used in this study.

synthesized fluorescence is added and then subtracted using one of the background subtraction methods. Matlab v6.5 was used to preprocess the data, and SPSS v13 was used to carry out all statistical analyses.

Construction of the adaptive minmax fit parameters

The method requires that the adaptive fit parameters be determined first: specifically, the orders for the polynomial fit that will be used based on the F/S ratio. We sampled, with replacement, ten spectra from the mice construction data set. For each spectrum, we added the four synthesized backgrounds mentioned above with nine different F/S ratios (0.1, 0.5, 1, 5, 10, 50, 100, 500 and 1000) for a total of 360 spectra. Polynomial fit orders ranging from 1 to 8 were used for the polynomial fit. The literature suggests that the optimal order lies in this range. An unconstrained polynomial fit was then applied to each spectrum for a total of 2880 observations.

The two orders with the lowest RMS for a specific F/S ratio were chosen for the adaptive minmax fit (Table 1),

resulting in a staircase progression indicating that as the F/S ratio increases, a higher-order polynomial fit is needed to minimize the RMS. The next step in the process determined when to effect the transition from one polynomial order to another. Regression fits ($R^2 > 0.99$) using the means of the orders was employed to find the transition points between the fit orders (Table 2).

Once the parameters of the fit are determined, implementation of the adaptive minmax fit is straightforward. First, the method needs to estimate the F/S ratio of the spectrum by initially performing a fourth-order unconstrained polynomial fit. On the basis of the estimated F/S ratio, four polynomial fits are performed to determine the final fit. For example, if the algorithm estimates an F/S ratio of 5, then a third- and fourth-order, unconstrained and constrained polynomial fits will be performed (according to Table 2). The method then takes the maximum value of these four fits as the final fit for subtraction.

Validation of the adaptive minmax fit

We compare the adaptive minmax fit constructed above to three other polynomial fits (fourth-, fifth- and sixth-order unconstrained polynomial fit) currently used. We sampled, with replacement, ten spectra from the mice validation data set. For each of the spectra, we added the four synthesized backgrounds mentioned above with nine different F/S ratios

Table 2. Parameters to be used for the adaptive algorithm

F/S range	Order of fit
(0, 0.2)	1st, 2nd
(0.2, 0.75)	2nd, 3rd
(0.75, 8.5)	3rd, 4th
(8.5, 55)	4th, 5th
(55, 240)	5th, 6th
(240, 517)	6th, 7th
(517, 1000)	6th, 8th

Table 1. Mean RMS for experiment 1 with respect to the F/S ratio and polynomial fit order. Highlighted are the two orders with the lowest RMS values for a specific F/S ratio

F/S ratio	Polynomial fit order							
	1st	2nd	3rd	4th	5th	6th	7th	8th
0.1	0.010	0.012	0.021	0.023	0.052	0.058	0.084	0.086
0.5	0.050	0.019	0.022	0.023	0.052	0.058	0.084	0.086
1	0.078	0.029	0.024	0.023	0.052	0.058	0.084	0.086
5	0.196	0.106	0.038	0.026	0.052	0.058	0.084	0.086
10	0.232	0.147	0.058	0.027	0.052	0.058	0.084	0.086
50	0.274	0.218	0.162	0.056	0.058	0.059	0.084	0.086
100	0.284	0.220	0.188	0.080	0.068	0.060	0.084	0.086
500	0.290	0.218	0.229	0.153	0.106	0.072	0.086	0.086
1000	0.294	0.218	0.223	0.162	0.113	0.087	0.094	0.087

(0.1, 0.5, 1, 5, 10, 50, 100, 500, and 1000) for a total of 360 files. Each of the 4 subtraction methods was then applied to each of the spectrum for a total of 1440 observations. These data were subjected to ANOVA, with background, F/S ratio and subtraction method as fixed factors.

The results indicate that the adaptive minmax fit is better than a fifth- and sixth-order fit across all F/S ratios (Fig. 7). The adaptive minmax fit is better than a fourth-order fit for all F/S ratios except at F/S = 10. Examining only spectra pertaining to F/S = 10, our results indicate that the subtraction method is significant ($F_{3,144} = 90.9$, $p < 0.001$) but there is no significant difference between the fourth-order and the adaptive minmax fit according to the *post hoc* Student–Newman–Keuls test. If we consider all F/S ratios, the subtraction method is significant ($F_{3,1296} = 109.3$, $p < 0.001$). A Tamhane's T2 *post hoc* test (for nonhomogeneous variance) indicates that the adaptive minmax fit is indeed significantly better when compared to each of the other three methods ($p < 0.001$). This indicates that the adaptive minmax fit is working as intended. It is minimizing the RMS by attempting to take the best part of each polynomial fit.

A second comparative study was performed using a Raman phantom. A plant leaf was used as the fluorescence phantom as it did not exhibit any Raman peaks in the 600–1800 cm^{-1} range (Fig. 8). The spectrum was obtained with an integration time of 10 s at a power of 50 mW averaged over three accumulations. Since the raw spectrum contained noise, we used a median filter followed by wavelets to de-noise the spectrum before applying any background subtraction technique. The same methodology from the first validation comparative study is used, except that the Raman phantom is added at different F/S ratios to the spectra from the validation data set instead of the synthesized backgrounds. The same 4 subtraction methods are then applied for a total of 360 observations. The F/S ratio and subtraction method were the fixed factors for the ANOVA analysis.

The results show that the adaptive minmax fit is better than the three other orders across all F/S ratios (Fig. 9).

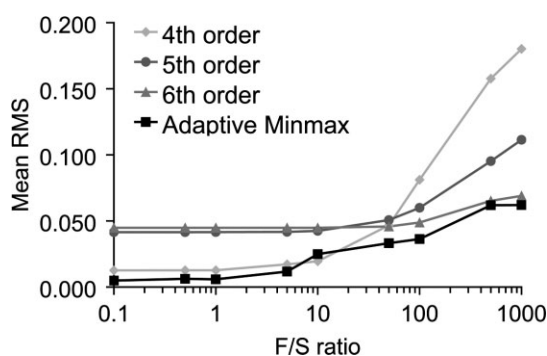


Figure 7. A comparison of the adaptive minmax polynomial fit vs the fourth-, fifth- and sixth-order polynomial fit using the residual mean square (RMS) across different F/S ratios (plotted on a log scale for clarity).

ANOVA results indicate that the subtraction method and the F/S ratio are significant factors ($F_{3,324} = 28.6$, $p < 0.001$ and $F_{8,324} = 190.9$, $p < 0.001$, respectively) in addition to the interaction between the two variables ($F_{24,324} = 4.7$, $p < 0.001$). Tamhane's T2 *post hoc* test indicates a significant difference ($p \leq 0.031$) between the adaptive minmax fit and the three other fits, while there were no significant differences ($p > 0.05$) among the three themselves.

We now apply the adaptive minmax fit to cases where the true Raman spectra are not known. We acquired the raw Raman spectrum of human skin, bone and brain. The spectrum were acquired *in vitro* with a power of 50 mW, integration time of 10 s and averaged over 3 accumulations. We applied a median filter and wavelets to de-noise and smooth the spectra. Then the adaptive minmax fit was applied followed by normalization to produce the subtracted spectra (Fig. 10). The algorithm retains the outline of the spectra, while the endpoints remain near zero.

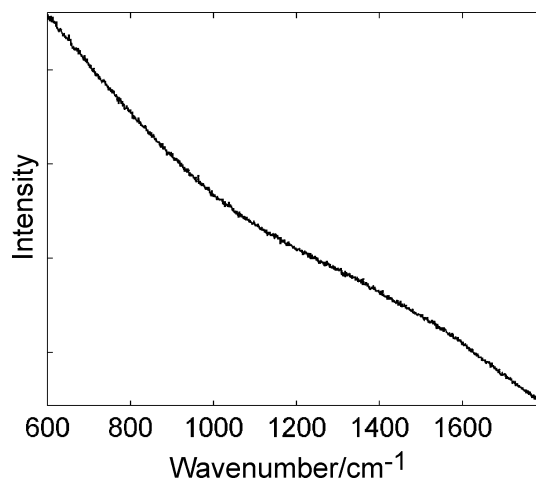


Figure 8. The raw spectrum of the fluorescence Raman phantom used in the study.

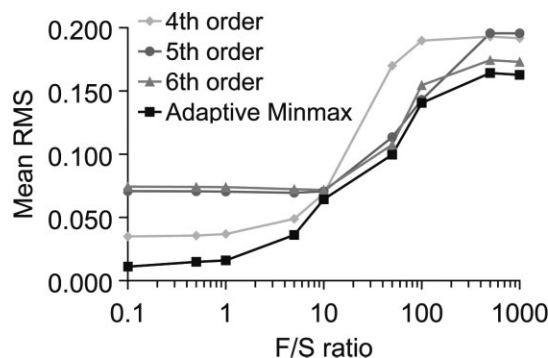


Figure 9. A comparison using a Raman phantom of the adaptive minmax polynomial fit vs the fourth-, fifth-, and sixth-order polynomial fit using the residual mean square (RMS).

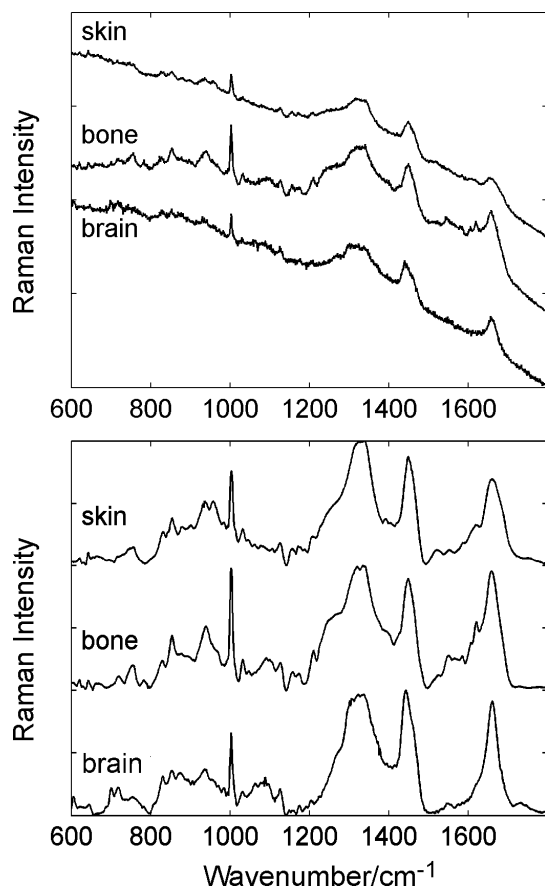


Figure 10. Example of the adaptive minmax fit on human tissue. The top figure shows the raw Raman spectra of human skin, bone and brain obtained *in vitro*. The bottom graph shows the spectra after de-noising and the adaptive minmax fit were applied. The spectra are normalized and offset for clarity.

As mentioned previously, this method is independent of other steps in the processing of the data and therefore can be coupled with other techniques for de-noising, calibration and statistical analysis. The method can be implemented into a real-time data analysis design similar to the one proposed by Bakker Schut *et al.*¹⁶ Their design is divided into three parts: construction of a data analysis model, initialization of the model and application of the model. The modular design allows the option for different models of statistical analysis and preprocessing methods to be implemented – thereby giving the user the choice of selecting which background subtraction method to employ.

We do not expect to have the same parameters for each data set (i.e. tissue type) although we expect to see the same trend. As the F/S increases, the optimal polynomial orders will also increase (staircase progression). We have presented and validated a methodology for implementing a robust method for background subtraction.

CONCLUSIONS

We have described an adaptive minmax method to subtract the fluorescence background in Raman spectra of biological tissues. The method is unbiased and requires no user intervention. Comparison shows that the adaptive minmax fit performs significantly better than the current polynomial fits in removing fluorescence background. We have shown that an adaptive minmax algorithm works best at minimizing RMS error especially when encountering a variety of F/S ratios.

Acknowledgements

This work was supported in part by ENSURE, the Endowment for Surgical Research, and the Festival of Trees to the Computer-Assisted Robot-Enhanced Surgery Program in the Children's Research Center of Michigan at the Children's Hospital of Michigan. This work was also supported by Wayne State University President Reid's Research Enhancement Program.

REFERENCES

1. Wolthuis R, Bakker Schut TC, Caspers PJ, Buschman HPJ, Romer TJ, Bruining HA, Puppels GJ. Raman spectroscopic methods for *in vitro* and *in vivo* tissue characterization. In *Fluorescent and Luminescent Probes for Biological Activity*, Chap. 32, (2nd edn). WT Mason (ed.). Academic Press: London, 1999; 433.
2. Santos LF, Wolthuis R, Koljenovic S, Almeida RM, Puppels GJ. *Anal. Chem.* 2005; **77**: 6747.
3. Shim MG, Wilson BC. *J. Raman Spectrosc.* 1997; **28**: 131.
4. Angel SM, DeArmond MK, Hanck KW, Wertz DW. *Anal. Chem.* 1984; **56**: 3000.
5. Shreve AP, Cherepy NJ, Mathies RA. *Appl. Spectrosc.* 1992; **46**: 707.
6. Vickers TJ, Wambles RE, Mann CK Jr. *Appl. Spectrosc.* 2001; **55**: 389.
7. Short KW, Carpenter S, Freyer JP, Mourant JR. *Biophys. J.* 2005; **88**: 4274.
8. Gao M, Lewis G, Turner GM, Soubret A, Ntziachristos V. *Appl. Opt.* 2005; **44**: 5468.
9. Brennan JF, Wang Y, Dasari RR, Feld MS III. *Appl. Spectrosc.* 1997; **51**: 201.
10. Huang Z, McWilliams A, Lui H, McLean DI, Lam S, Zeng H. *Int. J. Cancer* 2003; **107**: 1047.
11. Lieber CA, Mahadevan-Jansen A. *Appl. Spectrosc.* 2003; **57**: 1363.
12. Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. *Proc. Natl. Acad. Sci. U. S. A.* 2005; **102**: 12371.
13. Haka AS, Volynskaya Z, Gardecki JA, Nazemi J, Lyons J, Hicks D, Fitzmaurice M, Dasari RR, Crowe JP, Feld MS. *Cancer Res.* 2006; **66**: 3317.
14. Shafer-Peltier KE, Haka AS, Fitzmaurice M, Crowe J, Myles J, Dasari RR, Feld MS. *J. Raman Spectrosc.* 2002; **33**: 552.
15. Schulze G, Jirasek A, Yu MML, Lim A, Turner RFB, Blades MW. *Appl. Spectrosc.* 2005; **59**(5): 545.
16. Bakker Schut TC, Wolthuis R, Caspers PJ, Puppels GJ. *J. Raman Spectrosc.* 2002; **33**: 580.