Effect of photobleaching on calibration model development in biological Raman spectroscopy

Citation: Barman, Ishan et al. "Effect of Photobleaching on Calibration Model Development in Biological Raman Spectroscopy." Journal of Biomedical Optics 16.1 (2011) : 011004. © 2011 Society of Photo-Optical Instrumentation Engineers (SPIE)

As Published: http://dx.doi.org/10.1117/1.3520131

Publisher: Society of Photo-optical Instrumentation Engineers (SPIE)

Persistent URL: http://hdl.handle.net/1721.1/64970

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Effect of photobleaching on calibration model development in biological Raman spectroscopy

Ishan Barman,[a] Chae-Ryon Kong,[a] Gajendra P. Singh,[b] and Ramachandra R. Dasari[a]
[a]Massachusetts Institute of Technology, G. R. Harrison Spectroscopy Laboratory, Laser Biomedical Research Center, Cambridge, Massachusetts 02139
[b]University of St. Andrews, School of Physics and Astronomy, St. Andrew KY16 9SS, Scotland

Abstract. A major challenge in performing quantitative biological studies using Raman spectroscopy lies in overcoming the influence of the dominant sample fluorescence background. Moreover, the prediction accuracy of a calibration model can be severely compromised by the quenching of the endogenous fluorophores due to the introduction of spurious correlations between analyte concentrations and fluorescence levels. Apparently, functional models can be obtained from such correlated samples, which cannot be used successfully for prospective prediction. This work investigates the deleterious effects of photobleaching on prediction accuracy of implicit calibration algorithms, particularly for transcutaneous glucose detection using Raman spectroscopy. Using numerical simulations and experiments on physical tissue models, we show that the prospective prediction error can be substantially larger when the calibration model is developed on a photobleaching correlated dataset compared to an uncorrelated one. Furthermore, we demonstrate that the application of shifted subtracted Raman spectroscopy (SSRS) reduces the prediction errors obtained with photobleaching correlated calibration datasets compared to those obtained with uncorrelated ones. [DOI: 10.1117/1.3520131]

Keywords: Blood glucose monitoring; Raman spectroscopy; tissue fluorescence; photobleaching; shifted subtracted Raman spectroscopy; chemometrics.

Paper 10134SSR received Mar. 17, 2010; revised manuscript received May 14, 2010; accepted for publication Jun. 23, 2010; published online Jan. 7, 2011.

1 Introduction

Raman spectroscopy provides a powerful tool for noninvasive and real-time diagnostics of biological samples due to its exquisite molecular specificity and lack of sample preparation requirements. Specifically, the wealth of information available in Raman spectra has led to accurate quantification of molecular and morphological components embedded in complex biological samples, such as the determination of glucose, creatinine, and urea in whole blood and human eye aqueous humor.[1,2] Noninvasive blood glucose detection, in particular, is of considerable importance due to its implications for diabetes monitoring.[3] The potential of Raman spectroscopy is further substantiated by the promising results obtained in preliminary transcutaneous studies in human volunteers.[4,5] Nevertheless, it has proven to be more challenging to develop a robust calibration model that can make clinically accurate prospective predictions, irrespective of the nonanalyte specific properties of the sample (such as absorption and scattering in biological tissue[6]).

One of the primary hurdles in attaining the aforementioned calibration transfer is the presence of a varying luminescence (fluorescence) background. Additionally, the significantly greater intensity of tissue fluorescence frequently limits analysis to strong Raman bands only.[7] The associated photon shot noise and detector noise, which may have intensities similar to that of the Raman signal depending on the signal acquisition times, further compromise the diagnostic capability of the spectroscopic technique.

More importantly, in biological studies, quenching (photobleaching) of the endogenous fluorophores can change the acquired signal appreciably over time. In the context of the following discussion, photobleaching is defined as the reduction in sample autofluorescence intensity over a period of time under sustained laser exposure. A number of different mechanisms, including deactivation of the excited-state fluorophores on contact with other molecules (commonly known as quenchers) in the mixture sample, can contribute to the overall drop in fluorescence intensity.[8] In clinical Raman studies, the biological sample might be subject to a laser beam over a time period varying from a few seconds to several hours. For example, glucose tolerance tests[9] and glucose/insulin clamping studies[10] in humans and animal models can take anywhere between two to six hours. Naturally, fluorescence quenching manifests itself in the Raman spectra acquired over the duration of such studies.

To extract quantitative glucose information from the acquired spectra, it is necessary to employ implicit calibration methods such as partial least squares (PLS) and principal component regression (PCR).[11,12] These methods require only the concentrations of the analyte of interest and the spectra acquired from the calibration samples to construct a regression model. However, these calibration methods are often misled by chance correlations between nonanalyte specific spectral information and the concentrations of the analyte of interest, as originally demonstrated by Arnold, Burmeister, and Small.[13] Spurious

Address all correspondence to: Ishan Barman, Massachusetts Institute of Technology, G. R. Harrison Spectroscopy Laboratory, Laser Biomedical Research Center, Cambridge, Massachusetts 02139. Tel: 617-258-8694; Fax: 617-253-4513; E-mail: ishan@mit.edu.

1083-3668/2011/16(1)/011004/10/$25.00 © 2011 SPIE

effects such as system drift and covariations among sample constituents can lead to apparently functional models, which cannot be successfully employed for prospective prediction.[14]

We hypothesize that sample photobleaching introduces similar spurious correlations in the calibration models for biological Raman spectroscopy, particularly in cases where temporal correlations exist within the concentration dataset. Specifically for the characterization of metabolic activity of the blood glucose regulation system, functional tests such as glucose loading tests or glucose tolerance tests are frequently performed where the blood glucose levels are changed by ingestion of glucose or insulin. To prove the clinical feasibility of the spectroscopic techniques (e.g., infrared and Raman), spectra are typically acquired over the duration of these studies to form a calibration model (in conjunction with the measured analyte concentrations) for prospective prediction of blood glucose levels. In this work, we first assess the robustness (and validity) of models developed on samples undergoing photobleaching during glucose tolerance tests. Using numerical simulations and experimental studies on physical tissue models, we demonstrate that a significant deterioration in prospective prediction accuracy is observed when the model is developed on "correlated samples," i.e., calibration samples that demonstrate statistically significant correlation between (photobleached) fluorescence levels and glucose concentrations, compared to when it is developed on "uncorrelated samples." This is attributed to the construction of a spurious calibration model, where the regression vector is partly based on the fluorescence signal rather than only on the analyte of interest (glucose). This result has major implications for diabetics, as any tolerance-test-based protocol would lead to an approximately monotonic rise in glucose levels over the measurement period due to the inadequate insulin response. Evidently, the resultant covariation between the glucose concentrations and the photobleaching-induced decay in the tissue autofluorescence levels can lead to systematic errors in the calibration models.

The second goal of this work is to examine fluorescence removal methodologies as potential tools for avoiding the pitfalls associated with photobleaching correlated samples. Removal of fluorescence to isolate the Raman bands is a well-studied problem in the vibrational spectroscopy community, and several methodologies including time gating, derivative processing, least squares polynomial subtraction, shifted excitation Raman difference spectroscopy (SERDS), and shifted subtracted Raman spectroscopy (SSRS) have been proposed over the years.[15–20] SERDS, which is based on the differential shift response of the Raman and fluorescence signals to the shift in excitation wavelength, offers a promising solution, as it does not suffer from insufficient photon collection problems (as opposed to the time-gating schemes) and does not, in principle, introduce spectral artifacts like the numerical postprocessing techniques. Typically, a tunable laser source is employed to obtain the sample spectra at two shifted excitation frequencies, which are then subtracted to eliminate the fluorescence background, leaving only the Raman component as first derivative signals. SSRS is conceptually similar to the SERDS method, except that it involves shifting the spectrograph grating instead of the laser frequency. Although SERDS and SSRS have been used successfully to reconstruct "intrinsic" Raman spectra from simple mixture models, their efficacy in quantitative problems of concentration prediction, particularly when used in conjunction with an implicit calibration method, remains largely untested. Here we compare the effectiveness of SSRS to numerical postprocessing techniques, namely least squares polynomial subtraction and spectral derivative application, for both photobleaching correlated and uncorrelated samples. We observe that exploiting the shifted Raman spectra in SSRS to reject the fluorescence background reduces the prospective prediction errors for the calibration models developed on the correlated samples to the levels of the models calibrated on uncorrelated samples.

2 Experiments

2.1 *Experimental Studies on Human Subjects*

The concentration datasets as well as the tissue fluorescence and photobleaching profile used in our numerical simulations (Sec. 3) was originally described in one of our laboratory's previous publications.[4] Briefly, NIR Raman spectra were acquired from the forearms of 20 human volunteers undergoing oral glucose tolerance tests (OGTT). An 830-nm diode laser (Process Instruments) was used for Raman excitation, and the back-scattered light from the tissue was collected on a liquid-nitrogen cooled charge-coupled device (CCD) (Roper Scientific, Ottobrunn, Germany) coupled to a f/1.8i spectrograph (Kaiser Optical Systems, Ann Arbor, Michigan) For each volunteer, spectral measurements were initiated following the ingestion of a 220-ml SUN-DEX solution (Fisher Hamilton LLC, Two Rivers, Wisconsin) containing 75 g of glucose. Spectra were acquired for a total of 3 min per data point at 5-min intervals over approximately a two-hour time period, forming a measurement time series for each volunteer. A representative spectral time series acquired from a volunteer is shown in Fig. 1(a). The second panel, Fig. 1(b), gives the photobleaching profile as measured on this volunteer. It is a curve of I_λ/I_{exc} as a function of time, where I_λ is the area under the spectral curve over a 20-cm^{-1} band surrounding any arbitrarily selected central value (600 cm^{-1} used for data shown here), and I_{exc} is the measured intensity of the excitation source. The measured values of I_λ/I_{exc} were fitted with a double exponential curve,[21] as shown in the figure.

Concomitant with the spectral measurements, blood samples were also collected using finger sticks at regular 10-min intervals for estimation of glucose levels by a clinical glucose analyzer (HemoCue, Lake Forest, California). To correlate with the spectral measurements, the reference values were approximated at intermediate time points using spline interpolation. For comparison, two such profiles taken from two different volunteers are shown in Fig. 2. Profile 1 shows a rise in blood glucose concentration (to almost double its normal value), followed by a return to original levels, indicative of proper functionality of the blood glucose regulation system in the volunteer. Profile 2, in sharp contrast, exhibits a nearly monotonic rise in glucose levels over the test period, characteristic of a weak insulin-mediated response in this subject.

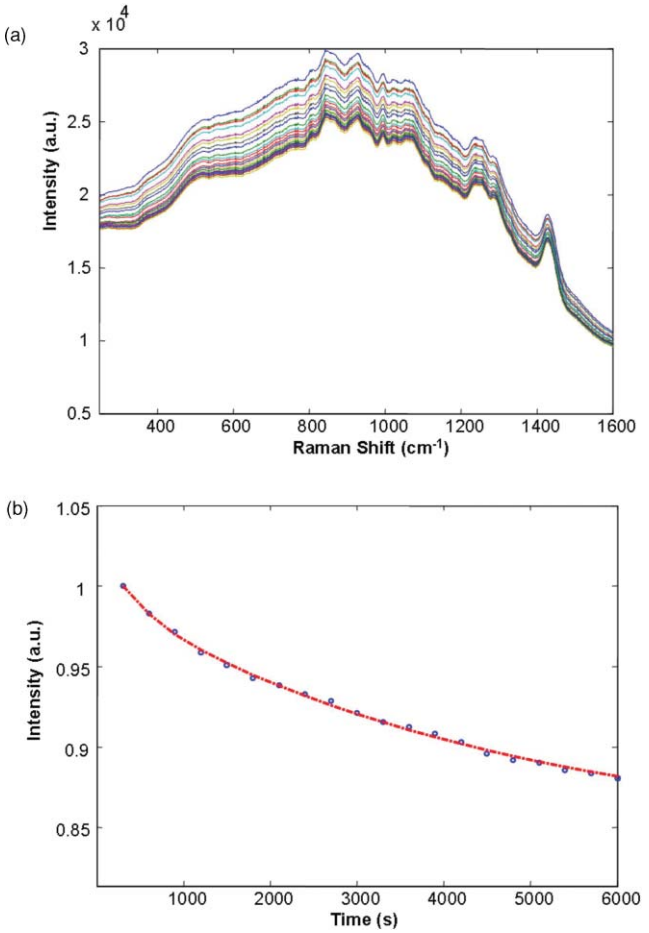

Fig. 1 (a) Raman spectra acquired from a human volunteer during an OGTT. (b) Normalized tissue autofluorescence decay (I_λ/I_{exc}) obtained as a function of time during the OGTT performance (the measured data points are in circles and the dotted line is the best-fit double exponential curve).

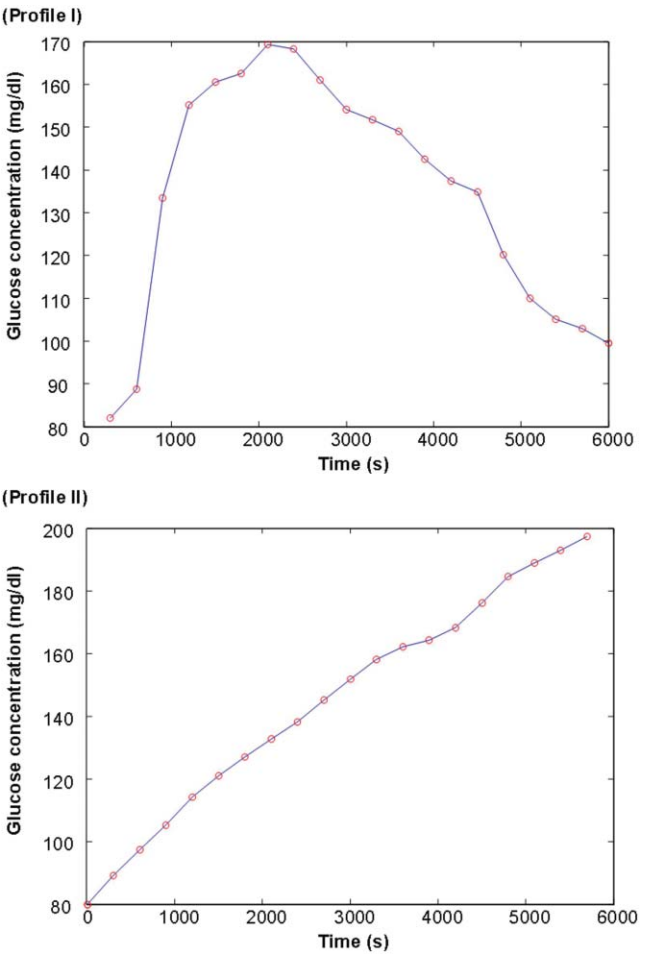

Fig. 2 Representative glucose concentration profiles taken from two human volunteers: profile 1 (top) and profile 2 (bottom).

2.2 *Experimental Studies on Tissue Phantoms*

Physical tissue models (tissue phantoms) in which the fluorescence and Raman scattering could be precisely varied were used to investigate the prediction performance of calibration models based on the correlated and uncorrelated spectral datasets, respectively. Additionally, in conjunction with the numerical simulations performed in Sec. 3, this study was used to understand the effect of fluorescence removal strategies on the prospective prediction accuracy of the developed calibration models.

A total of 50 tissue phantoms were prepared with aqueous solutions of glucose (analyte of interest), urea (spectral interferent), and indocyanine green (ICG). ICG was employed to produce a fluorescence background similar to the tissue autofluorescence observed with near-infrared (NIR) excitation. To mimic the (photobleached) fluorescence correlation observed in glucose tolerance testing of human subjects with impaired glucose tolerance (see Profile 2 in Fig. 2), 20 tissue phantoms were constructed, where the correlation between the glucose and ICG concentrations was high ($R^2 \approx 0.95$). The other 30 tissue phantoms had negligible correlation between the glucose and ICG levels ($R^2 \approx 0.1$). For both correlated and uncorrelated tissue phantoms, glucose and ICG concentrations varied from 0.9 to 1.2 M and 0.18 to 0.36 μM, respectively. The relatively high concentration of glucose in the tissue phantoms enabled higher fluorescence values to be studied while retaining a satisfactory signal-to-noise ratio. The urea concentration in all the tissue phantoms was randomized in the range of 250 to 400 mM.

For Raman excitation, an 830-nm diode laser was focused into the delivery fiber (200-μm core diameter) of an optical fiber and filter probe assembly[22] to deliver an average power of $\sim$100 mW on a spot size of $\sim$1 mm^2 at the sample. The sample solution was held in a cylindrical glass vial (one inch diameter) with a top opening, through which the probe tip was immersed directly into the solution without contacting the container (Fig. 3). The back-scattered Raman light was collected by ten collection fibers (200-μm core diameter) before being dispersed via a tunable f/2.0 spectrograph (LS-785, 5 cm^{-1} resolution, Princeton Instruments, Trenton, New Jersey) onto a TE-cooled CCD detector (PIXIS 256, Princeton Instruments). The acquisition time for each Raman spectrum was 20 sec. Spectra of the phantoms were acquired randomly with respect to the constituents' concentration to eliminate any potential temporal correlation. Since the optical fiber probe itself generated prominent Raman spectral features at wavelengths below 800 cm^{-1}, the spectral range of 950 to 1800 cm^{-1} was used for data analysis

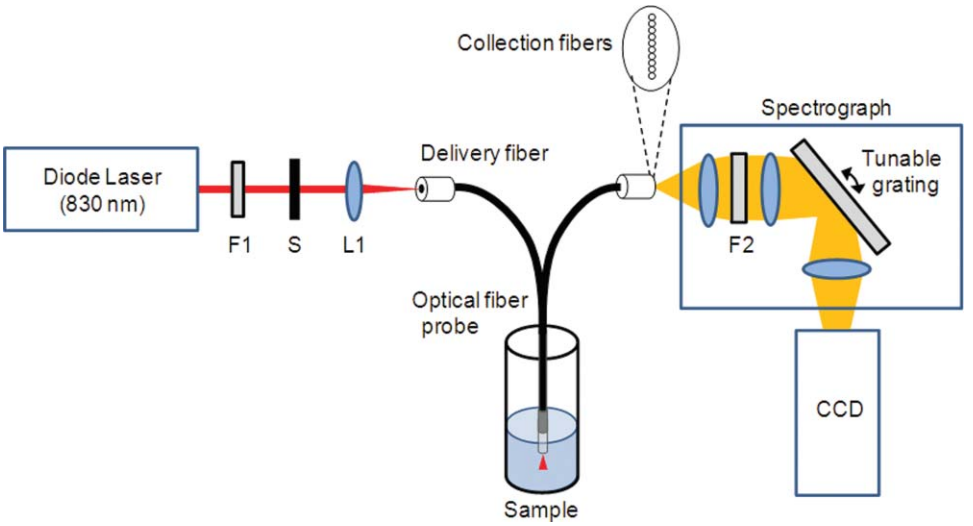

Fig. 3 A schematic diagram of the experimental setup. Raman spectra were obtained from tissue phantom solutions using an optical fiber probe, which included a laser light delivery fiber and ten collection fibers. The spectrograph was equipped with a micrometer, which was used to precisely tune the grating for implementing SSRS. F1: laser line filter. S: shutter. L1: focusing lens for optical fiber coupling. F2: Rayleigh rejection edge filter.

(PLS calibration and prediction) after vertical binning and cosmic ray removal. Additionally, to investigate strategies for removing the effects of the fluorescence background, SSRS was performed by acquiring the tissue phantom spectra after tuning the spectrograph grating position by ~25 cm^{-1} (approximately the full width half maximum of the characteristic glucose peak at 1100 cm^{-1}). The Raman shift axis was calibrated using the standard spectra measured from a 4-acetaminophenol powder sample. The tuning of the grating was monitored by the position of the laser line on the CCD detector such that an 18-pixel shift was maintained between the two positions. Further details of the calibration procedure followed for SSRS implementation are described elsewhere.[23] The difference between the pair of Raman spectra (acquired at two different grating positions) was used for PLS calibration and prediction, and the performance was compared with that of the PLS calibration model constructed using the raw Raman spectra. We have also employed first derivative spectra for comparison with the aforementioned calibration models. For the first derivative spectra, discrete differencing of the acquired spectra was performed. Additionally, we undertook a derivative application after smoothing of the acquired spectra (Savitzky and Golay,[24] nine-point window). All mathematical processing and numerical simulations were performed using in-house software codes written in Matlab (The Mathworks, Natick, Massachusetts).

3 Numerical Simulations

We performed two sets of numerical simulations using the experimental data from the human subject studies (Sec. 2.1), to investigate the questions outlined in Sec. 1. The first simulation study (Sec. 3.1) was aimed at determining the effect of photobleaching in introducing spurious correlations in calibration models. To characterize this relationship, we analyzed the robustness and prediction accuracy of calibration models developed on the photobleaching correlated and uncorrelated datasets, respectively. In the second study (Sec. 3.2), the same analysis was repeated, except that the spectra used in the study were processed for fluorescence removal. We investigated two fluorescence removal techniques in our simulations: least squares polynomial subtraction and SSRS. This study was used to compare the improvement in prediction performance obtained in both cases—correlated and uncorrelated—on employment of different fluorescence removal strategies.

3.1 *Effect of Photobleaching Study*

Simulation calibration spectra were first generated by forming weighted linear combinations of constituent analyte spectra of glucose, creatinine, and urea, as measured by our Raman system. For our application, glucose represents the analyte of interest, with the other two constituents playing the role of spectral interferents in the mixture model. The assigned weights in the linear combinations correspond to the concentrations of the analytes in the sample at that particular time instance. Two sets of simulation spectra were constructed: a correlated spectral dataset and an uncorrelated spectral dataset. Specifically, the correlation coefficient $\rho_{A,B}$ between two random variables A and B is given by:

$$\rho_{A,B} = \frac{\mathrm{cov}(A, B)}{\sigma_A \sigma_B}, \tag{1}$$

where cov(A, B) gives the covariance between A and B, and σ_A and σ_B are the standard deviations of A and B, respectively.[25] For the considered case, the variables A and B are constituted by the fluorescence intensity values and glucose concentrations, respectively. In the following, we state the strength of correlation by the R^2 value, which is the square of $\rho_{A,B}$.

To generate the correlated spectral dataset, the concentrations shown in profile 2 (Fig. 2) were assigned to glucose to formulate the mixture Raman spectra. The glucose concentration values of profile 2 show a strong negative correlation with the photobleached fluorescence levels plotted in Fig. 1(b) ($\rho_{A,B} \approx 0.95$, $R^2 \approx 0.9$). For the uncorrelated dataset, the concentration values from profile 1 (Fig. 2) were allocated for glucose, as the

correlation between this profile and the aforementioned photobleaching curve is insignificant ($R^2 \approx 0.1$). It should be noted that the concentrations of creatinine and urea in both sets of simulation spectra were kept nearly constant (i.e., within 5% variation) to replicate the negligible changes observed in the concentration levels of the other analytes during a typical OGTT.

To model the fluorescence, we performed hypersmoothing of a typical tissue spectrum to remove the sharper Raman features and retain only the autofluorescence background. For both correlated and uncorrelated dataset simulations, this hypersmoothed spectrum was scaled by a factor specified by the photobleaching profile [Fig. 1(b)] at that particular time before addition to the mixture of Raman spectra. Moreover, to fully evaluate the impact of fluorescence-to-Raman amplitude ratio on the performance of the calibration models for both the correlated and uncorrelated cases, the relative intensity of the fluorescence and Raman signals was also varied in the range of 5 to 20. Finally, noise was generated in accordance with the shot-noise-limited condition (which has previously been observed to hold true for spectra acquired using our Raman system[26]) for addition to the fluorescence-Raman combination spectra. We revisit the shot-noise limited assumption and its implications in Sec. 4.2. The signal-to-noise ratio (SNR), defined here as the ratio of the maximum intensity of the fluorescence-Raman combination signal to the mean noise magnitude, was varied from 50 to 200. The range of fluorescence-to-Raman amplitudes and SNR values mentioned before are consistent with that observed in Raman spectra acquired for these types of analytical measurements (typical values of observed SNR $\sim$ 100 and fluorescence-to-Raman ratio $\sim$ 10). At each setting of the tunable parameters, i.e., fluorescence-to-Raman ratio and SNR, 100 calibration spectra were generated for the correlated as well as the uncorrelated datasets.

The prediction spectra were generated following a scheme similar to that stated before, except that the concentrations of all the analytes were completely randomized in this dataset. The fluorescence levels in the prediction spectra were also completely randomized to simulate those observed in a set of unknown samples (such as a test set of human volunteers), where the fluorescence levels and the concentrations of the analytes would be expected to possess minimal correlation with one another. This situation typically occurs when a calibration model developed during a tolerance or loading test is used prospectively on different subjects to determine their glucose levels. For a given setting of the fluorescence-to-Raman ratio and SNR level, 50 such spectra were included in the prediction dataset.

Partial least squares (PLS) analysis was used to develop the calibration models based on leave-one-out cross-validation on the calibration (training) set. The reported final errors [root-mean-square error of prediction (RMSEP), Eq. (2)] were obtained by employing these calibration models on the independent prediction sets.

$$\mathrm{RMSEP} = \left[\sum_{i=1}^{n_P} \frac{\left(C_{\mathrm{act}} - C_{\mathrm{pred}}\right)^2}{n_p}\right]^{\frac{1}{2}}, \qquad (2)$$

where C_{act} and C_{pred} are the actual and predicted glucose concentrations in the prediction sample, and n_p is the number of samples used in the prediction set.[25] To ensure reproducibility of our results, the entire calibration-prediction procedure was repeated 20 times.

3.2 *Fluorescence Removal Study*

In this simulation study, we followed an algorithm similar to that outlined in Sec. 3.1, except that all the spectra (i.e., calibration and prediction, both for correlated and uncorrelated cases) were processed for fluorescence removal. Here, we employed SSRS to quantify its benefits, relative to that of numerical postprocessing schemes, for both the photobleaching correlated and uncorrelated cases.

For our simulations, we generated a second set of calibration and prediction spectra for both the correlated and uncorrelated datasets by shifting the composite mixture spectra (Raman plus fluorescence) by 25 cm^{-1}. These shifted spectra were then subtracted from the original set of spectra to obtain the Raman difference spectra (plus shot noise and residual fluorescence background). While most of the previous research has primarily been focused on reconstruction of the intrinsic Raman spectrum from the difference spectra,[27–29] we have employed the difference spectra directly in analysis to extract concentration information through PLS. The direct incorporation of the difference spectra into calibration analysis is advantageous, as it avoids the artificial spectral features and noise (ringing behavior) that may appear due to the application of a reconstruction method (such as Fourier deconvolution).

For fair comparison with numerical strategies, we implement three variations of least- squares polynomial subtraction as alternative fluorescence removal tools. First, a lower order polynomial is fit to the broad fluorescence background and subtracted. Specifically, we employ a second or third order polynomial for the PLS analysis, depending on which order provides the least error in cross-validation. This approach is widely pursued for signals with broad fluorescence backgrounds, because higher order polynomials tend to overfit the data, as also noted by Beier and Berger.[30] Second, an algorithm developed by Lieber and Mahadevan-Jansen,[17] which automatically fits a modified polynomial just below the original spectrum, is used. Third, a recently developed adaptive "minmax" scheme[31] is used to subtract the background fluorescence. PLS analysis is undertaken in conjunction with each of the three numerical processing schemes.

4 Results and Discussion

4.1 *Numerical Simulations*

4.1.1 *Effect of photobleaching study*

As detailed in Sec. 3.1, PLS was first used in a leave-one-out cross-validation routine on the calibration spectral set. The number of loading factors used to build the final model for prediction was in accordance with the number of constituents that provided the minimum error in cross-validation. The results of the simulations at varying levels of Raman-to-noise ratio for the photobleaching correlated and uncorrelated cases are plotted in Fig. 4, with error bars indicating the standard deviation of the RMSEP values at each setting. It is evident that the calibration model developed on the correlated calibration spectra always performs worse than the model developed on the uncorrelated

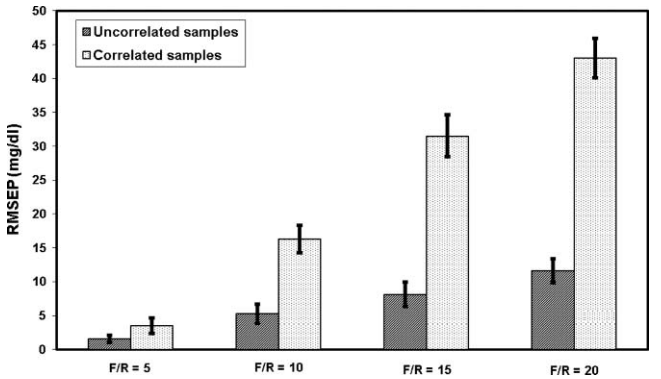

Fig. 4 Bar plot of RMSEP obtained for calibration models developed on photobleaching correlated and uncorrelated datasets, respectively, as a function of increasing fluorescence-to-Raman ratio (i.e., decreasing Raman-to-noise ratio). Here, the SNR is held constant (100). Identical results are obtained by changing the SNR while holding the fluorescence-to-Raman ratio fixed.

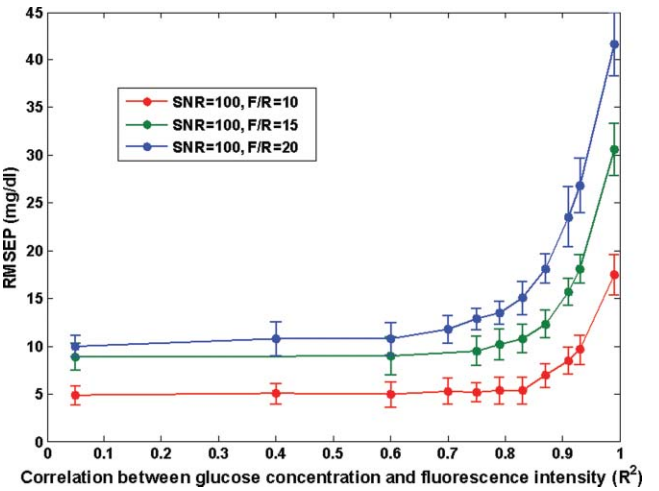

Fig. 5 RMSEP of simulations as a function of the correlation (R^2) between glucose concentration and fluorescence intensity. The fluorescence-to-Raman ratio (F/R) was varied from 10, 15 to 20, and 20 simulations were performed for each case.

set, irrespective of the SNR value or the fluorescence-to-Raman amplitude ratio. For example, at a setting of fluorescence-to-Raman of 10 and SNR of 100 (values observed typically for tissue Raman spectra), the mean prediction errors were measured to be 16.3 and 5.3 mM for the correlated and uncorrelated cases, respectively, thereby demonstrating an approximately three-fold increase in the prediction error. Expectedly, for lower levels of Raman-to-noise ratio, the prediction performance of both models drops considerably. (It is noted that changing the fluorescence-to-Raman ratio or the SNR has an identical effect, as they both modulate the Raman-to-noise ratio.) Importantly, the performance of the calibration model built on the correlated dataset drops off much more significantly than that developed on the uncorrelated dataset—as evidenced by the widening gap between the bars at higher fluorescence-to-Raman (i.e., lower Raman-to-noise) settings. This can be attributed to the fact that it becomes more difficult to distinguish the Raman spectrum of glucose from the confounding fluorescence background in a noisier spectral set. The simplest experimental solution to reducing the disparity between the correlated and uncorrelated datasets is to increase the overall SNR, which would then decrease the RMSEP; however, this would come at the cost of either increased exposure time or increased laser power.

The comparison of PLS prediction performance between correlated and uncorrelated datasets (Fig. 4) can be further generalized to show that the amount of deterioration in the prediction performance is dependent on the magnitude of correlation (R^2) between glucose concentration and tissue fluorescence. In Fig. 5, RMSEP increases with increasing R^2 for all values of fluorescence-to-Raman ratios. At low values of R^2, the RMSEP appears to be independent of R^2, but displays a steep rise after a threshold value of R^2. This suggests that the prediction error at low values of R^2 is determined primarily by the spectroscopic SNR, while at higher values of R^2, the prediction error depends on both SNR and R^2. This can also be visualized by observing the widening gap between any pair of RMSEP curves with an increase in R^2 in Fig. 5.

The results of Figs. 4 and 5 also have extensive clinical implications, particularly for the correlated case. Such a situation is representative of experimental observations from OGTTs performed on diabetic and insulin-resistant patients. Since it is difficult to perform glucose/insulin clamping studies in normal, and particularly, diabetic populations, tolerance tests have remained the most viable protocol for the development of a calibration model using spectroscopic techniques.[32,33] Moreover, while test-strip-based measurements and glycosylated hemoglobin (HbA1c) examination are more extensively used at the present time, tolerance tests remain the gold standard of clinical tests of insulin/glucose control systems, and have been used on a regular basis to screen for gestational diabetes[34,35] and to determine the level of insulin resistance.[36] However, our results indicate that photobleaching introduces a systematic error in Raman spectroscopic calibration model development using tolerance tests, particularly for people with impaired glucose regulation. Put another way, any model developed on a diabetic population is susceptible to substantial errors in prospective prediction, even when used for diabetic patients only. This is clearly undesirable. The potential remedies to break the unwanted correlation include a change in experimental study design (e.g., point-of-care measurements across a large population) and removal of the fluorescence background. The inclusion of a large number of human volunteers in the calibration set may, however, introduce additional complexities such as large variations in tissue optical properties and skin heterogeneity.

4.1.2 *Fluorescence removal study*

For this study, the fluorescence removed spectra—SSRS-processed and least squares polynomial subtracted—were used in conjunction with the concentration datasets to build the calibration models (as detailed in Sec. 3.2). The results of prospective prediction of the five models (i.e., based on raw, lower order polynomial subtracted, modified polynomial subtracted, minmax fit subtracted, and SSRS processed spectra) are summarized in Fig. 6. For this set of simulations, experimentally observed values of SNR (of 100) and fluorescence-to-Raman ratio (of 10) were used.

For the correlated case, SSRS provides substantial reduction of prediction error (from 16.7 to 3.8 mg/dl), even lower

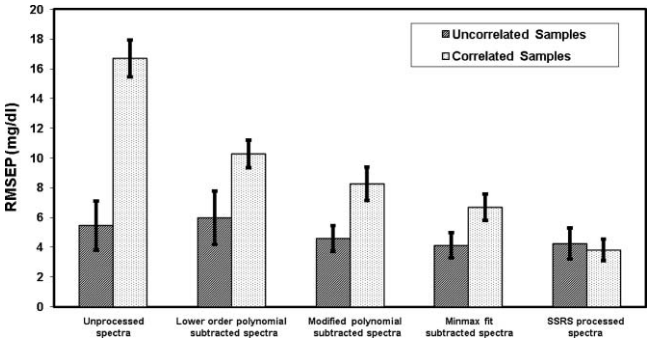

Fig. 6 Bar plot of RMSEP values obtained for glucose concentrations from simulations on photobleaching correlated (red) and uncorrelated (blue) datasets. The groups represent calibration and prediction performed using the following types of spectra: (from left to right) unprocessed, lower order polynomial subtracted, modified polynomial subtracted, minmax fit subtracted, and SSRS processed spectra, respectively.

than the levels obtained using the raw uncorrelated calibration model (5.5 mg/dl). Conventional lower-order polynomial subtraction, on the other hand, provides only about half the amount of improvement (from 16.7 to 10.3 mg/dl) as SSRS processing for the correlated case. This disparity in results could be attributed primarily to the introduction of nonphysical negative regions in the spectra and associated artifacts due to the process of fluorescence removal using the polynomial subtraction technique. Newer variants of the polynomial subtraction technique, namely the modified polynomial (8.3 mg/dl) and adaptive minmax fitting (6.7 mg/dl), provide lower average RMSEP values than conventional polynomial subtraction. Clearly, the inclusion of meaningful constraints in these techniques enhances the prospective prediction accuracy, though not to the levels obtained using SSRS.

Our results for the uncorrelated case suggest that application of SSRS or numerical processing do not cause similar changes in prediction performance from the raw spectra results. This is not unexpected, because implicit calibration schemes (PLS) do a reasonable job of separating the analyte of interest (e.g., glucose) from the confounding analytes (including tissue fluorescence), especially when reasonable levels of SNR are achieved.

Finally, it is important to understand that our numerical simulations have been based on the assumption that the main factor limiting SNR is the shot noise in the system. Although this assumption holds good for experiments on tissue samples where exposure times are limited by other constraints, this may not always be the case. As has been pointed out by Bell, Bourguignon, and Dennis,[23,37] and O'Grady et al.,[38] there are many situations where it is observed that the random but fixed variations in pixel-to-pixel response on the CCD might actually be more dominant than shot noise. This is especially true when high signal levels can be achieved and exposure times can be increased without any significant downside (such as when chemical mixtures and powder samples are tested). SSRS will clearly provide a bigger improvement in such applications, because the major cause of noise arising from the irregularity in detector response can be canceled when the shifted spectra are subtracted.

4.2 *Experimental Studies on Tissue Phantoms*

To validate our simulation results, we analyzed the spectral datasets obtained in our tissue phantom studies. Figure 7(a) shows a raw Raman spectrum obtained from a representative tissue phantom, Fig. 7(b) shows SSRS data obtained by subtracting two raw spectra obtained at slightly different spectrograph grating positions 25 cm^{-1} apart, and Fig. 7(c) shows the numerical first derivative of the raw spectrum shown in Fig. 7(a). It is clear

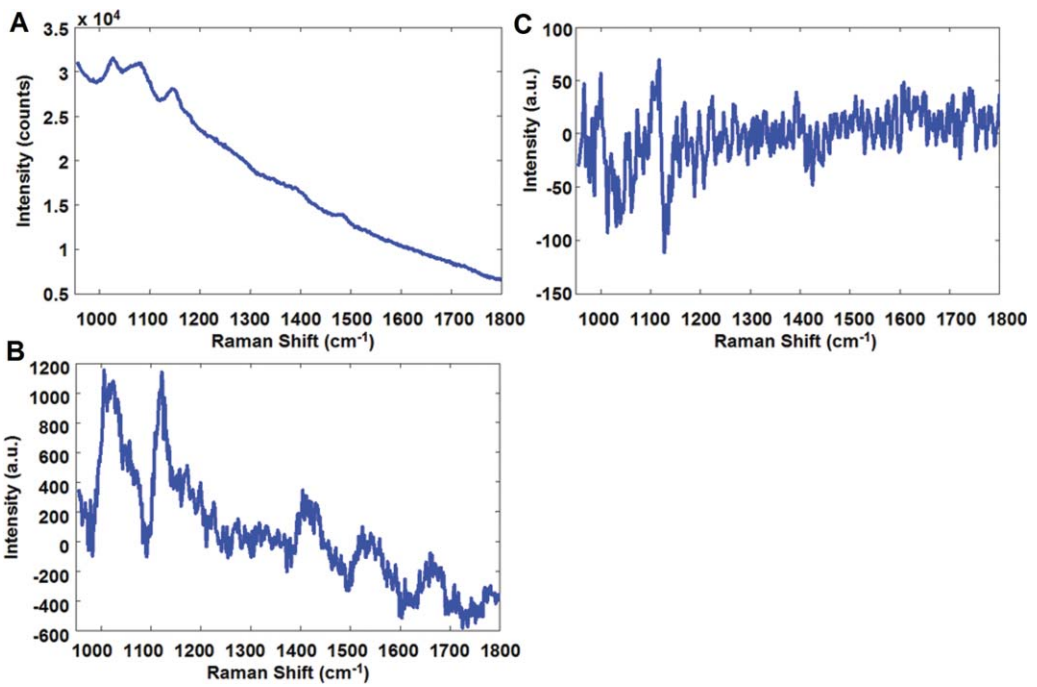

Fig. 7 (a) Representative Raman spectrum acquired from a tissue phantom. (b) SSRS spectrum obtained by subtracting two spectra, obtained at spectrograph grating positions 25 cm^{-1} apart. (c) First derivative spectrum.

Table 1 Summary of mean and standard deviation (in parentheses) of the RMSEP values obtained from the tissue phantom experiments for both the correlated and uncorrelated datasets.

	Raw	SSRS	First derivative	First derivative after smoothing
Correlated	53.8 (±7.8) mM	42.7 (±7.3) mM	66.3 (±9.2) mM	56.0 (±8.1) mM
Uncorrelated	39.1 (±6.7) mM	42.4 (±6.2) mM	65.0 (±8.1) mM	55.5 (±8.9) mM

that the fluorescence background can be suppressed using either shifted subtraction at two grating positions (SSRS) or taking the first derivative. However, the SNR of the SSRS data appears to be significantly better than that of the first derivative spectra. This can be attributed to the cancellation of the fixed pattern noise on the CCD detector, as discussed before. It should be noted that having the correct amount of shift between the two acquired spectra is critical to the application of SSRS, as too small a shift would result in poor SNR and too large a shift would cause improper background suppression. For the SSRS spectrum, we also observe the presence of a small residual background because of the corresponding shift in fluorescence in the two sets of spectra.

We first created the prediction dataset by randomly extracting ten uncorrelated tissue phantom data. PLS calibration models were then constructed based on the raw spectra, SSRS spectra, and first derivative spectra for both the fluorescence-correlated and uncorrelated datasets (which consisted of 20 tissue phantoms each). Model performance during optimization was judged by comparing the standard error for leave-one-out cross-validation in the calibration dataset. The prediction set was withheld from all optimizations and was solely employed to evaluate the performance of the final optimized models. The whole procedure was iterated 50 times to obtain an average and standard deviation of the RMSEP values. Table 1 summarizes the results of the prospective prediction.

We observe that the prediction performance of the calibration model built on the fluorescence-correlated dataset is substantially worse than that built on uncorrelated phantoms. The mean prospective prediction errors are 53.8 and 39.1 mM for the correlated and uncorrelated calibration models, respectively. Importantly, we find that the application of SSRS reduces the error in the correlated calibration model from 53.8 to 42.7 mM to the levels achieved with the uncorrelated dataset. We also observe that the application of SSRS marginally increases the prediction error of the uncorrelated dataset from 39.1 to 42.4 mM. This may be attributed to the imperfect rejection of the fluorescence background by SSRS application. Nevertheless, the two primary results observed in our previous numerical simulations are reinforced by this tissue phantom study: 1. calibrations developed from the fluorescence-correlated datasets have a greater prediction error than calibrations developed with fluorescence-uncorrelated datasets, and 2. the application of frequency-shifted spectra provides more accurate concentration predictions than those obtained from the unprocessed spectra, particularly for the fluorescence-correlated dataset.

In sharp contrast to SSRS, the first derivative application appears to adversely impact the prediction performance for both the fluorescence-correlated and uncorrelated datasets. One would expect that this is an outcome of the reduction in SNR of the first derivative spectra shown in Fig. 7(c). This is consistent with previous observations, especially when an additional filtering procedure is not employed.[39] It has also been previously reported that the derivative method makes the lower intensity Raman bands difficult to identify in the processed spectrum.[17] Smoothing of the acquired spectra before the differentiation procedure, however, reduces the average RMSEP value, probably due to the resultant increase in SNR[16] compared to simple derivative application.

Taken in conjunction with the least squares polynomial subtraction results from our simulations, these results indicate that while the numerical fluorescence removal strategies are able to suppress the fluorescence background, the instrumentation-based SSRS technique provides better quantitative predictions. Furthermore, the tissue phantom studies verify the robustness of the SSRS methodology in a realistic setting.

5 Conclusion

A central challenge in the development of a robust calibration model for transcutaneous blood analyte detection using Raman spectroscopy lies in overcoming the effects of photobleaching-induced nonanalyte specific variations. In this work we show that significant deterioration in model performance can be observed due to the photobleaching-induced spurious correlations. At typically observed levels of SNR and fluorescence-to-Raman ratio, this can result in a factor of 3 increase in prediction error. This has serious implications in developing a calibration model, particularly for diabetic patients using glucose tolerance test-like protocols. To remedy the situation, one needs to incorporate some combination of a change in experimental study design and fluorescence removal strategy. Evidently, a similar photobleaching-induced problem can be expected in other biological and industrial applications, especially when the measurement period is of similar time scales as the sample fluorescence lifetime. For example, this issue may interfere with Raman-spectroscopy-based investigation of single cells over the entire cell cycle.[40]

Furthermore, we demonstrate that SSRS reduces the prospective prediction error of the developed calibration models by rendering the model less susceptible to spurious effects associated with photobleaching-correlated samples. In this regard, SSRS appears to substantially outperform conventional numerical postprocessing schemes such as least squares polynomial subtraction and derivative application. It is worth noting that newer numerical processing schemes, such as the adaptive min-max approach and the automated background removal algorithm proposed by Beier and Berger[30] may provide better results

than the conventional numerical schemes, and the advantages of SSRS application over these methods needs to be carefully studied for the specific application. Nevertheless, our results imply that the problem of photobleaching-induced spurious correlation can be better solved by incorporating a tunable spectrograph, as compared to numerical processing schemes. Such implementation ensures that the principal advantages of glucose tolerance tests (i.e., the convenience of testing and the lack of stringent glucose and insulin control requirements as are necessary for clamping studies) are not sacrificed. For future work, detailed investigations need to be carried out to ascertain other sources of random errors that might be inherent due to tolerance test-like protocols. Moreover, to further improve accuracy of the obtained calibration models, we are currently investigating nonlinear calibration schemes such as support vector machines.[41]

Acknowledgments

This work was performed at the MIT Laser Biomedical Research Center and supported by the NIH National Center for Research Resources, grant P41-RR02594, and a grant from Bayer Health Care, LLC. Barman acknowledges support of the Lester Wolfe Fellowship from the Laser Biomedical Research Center. We would like to thank the Professor Michael Feld (late) for his interest and active support during the initial stages of this investigation. We are also grateful to the referees for their insightful comments.

References

1. A. M. K. Enejder, T. W. Koo, J. Oh, M. Hunter, S. Sasic, M. S. Feld, and G. L. Horowitz, "Blood analysis by Raman spectroscopy," *Opt. Lett.* **27**(22), 2004–2006 (2002).
2. C. C. Pelletier, J. L. Lambert, and M. Borchert, "Determination of glucose in human aqueous humor using Raman spectroscopy and designed-solution calibration," *Appl. Spectrosc.* **59**, 1024–1031 (2005).
3. J. N. Roe and B. R. Smoller, "Bloodless glucose measurements," *Crit. Rev. Ther. Drug.* **15**(3), 199–241 (1998).
4. A. M. K. Enejder, T. G. Scecina, J. Oh, M. Hunter, W. C. Shih, S. Sasic, G. Horowitz, and M. S. Feld, "Raman spectroscopy for non-invasive glucose measurements," *J. Biomed. Opt.* **10**(3), 031114 (2005).
5. J. Chaiken, W. Finney, P. E. Knudson, R. S. Weinstock, M. Khan, R. J. Bussjager, D. Hagrman, P. Hagrman, Y. W. Zhao, C. M. Peterson, and K. Peterson, "Effect of hemoglobin concentration variation on the accuracy and precision of glucose analysis using tissue modulated, noninvasive, in vivo Raman spectroscopy of human blood: a small clinical study," *J. Biomed. Opt.* **10**(3), 031111 (2005).
6. I. Barman, G. P. Singh, R. R. Dasari, and M. S. Feld, "Turbidity-corrected Raman spectroscopy for blood analyte detection," *Anal. Chem.* **81**(11), 4233–4240 (2009).
7. P. Matousek, M. Towrie, and A. W. Parker, "Fluorescence background suppression in Raman spectroscopy using combined Kerr gated and shifted excitation Raman difference techniques," *J. Raman Spectrosc.* **33**(4), 238–242 (2002).
8. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer, Berlin (1999).
9. American Diabetes Association, "Standards of medical care in diabetes," *Diabetes Care* **32**, S13–S61 (2009).
10. R. A. DeFronzo, J. D. Tobin, and R. Andres, "Glucose clamp technique: a method for quantifying insulin secretion and resistance," *Am. J. Physiol. Gastrointest. Liver Physiol.* **237**(3), G214–G233 (1979).
11. S. Wold, H. Martin, and H. Wold, *Lecture Notes in Mathematics*, Springer-Verlag, Heidelberg (1983).
12. R. F. Gunst and R. L. Mason, *Regression Analysis and Its Applications*, Marcel Dekker Inc., New York (1980).
13. M. A. Arnold, J. J. Burmeister, and G. W. Small, "Phantom glucose calibration models from simulated noninvasive human near-infrared spectra," *Anal. Chem.* **70**(9), 1773–1781 (1998).
14. R. Liu, W. Chen, X. Gu, R. K. Wang, and K. Xu, "Chance correlation in non-invasive glucose measurement using near-infrared spectroscopy," *J. Phys. D: Appl. Phys.* **38**, 2675–2681 (2005).
15. P. P. Yaney, "Reduction of fluorescence background in Raman spectra by the pulsed Raman technique," *J. Opt. Soc. Am.* **62**(11), 1297–1303 (1972).
16. P. A. Mosier-Boss, S. H. Lieberman, and R. Newbery, "Fluorescence rejection in Raman spectroscopy by shifted-spectra, edge detection, and FFT filtering techniques," *Appl. Spectrosc.* **49**(5), 9 (1995).
17. C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Appl. Spectrosc.* **57**(11), 1363–1367 (2003).
18. A. P. Shreve, N. J. Cherepy, and R. A. Mathies, "Effective rejection of fluorescence interference in Raman spectroscopy using a shifted excitation difference technique," *Appl. Spectrosc.* **46**(4), 5 (1992).
19. J. F. B. Brennan III, Y. Wang, R. R. Dasari, and M. S. Feld, "Near-infrared Raman spectrometer systems for human tissue studies," *Appl. Spectrosc.* **51**(2), 201–208 (1997).
20. M. D. Morris, P. Matousek, M. Towrie, A. W. Parker, A. E. Goodship, and E. R. C. Draper, "Kerr-gated time-resolved Raman spectroscopy of equine cortical bone tissue," *J. Biomed. Opt.* **10**(1), 014014 (2005).
21. H. Zeng, C. MacAulay, B. Palcic, and D. I. McLean, "Laser-induced changes in autofluorescence of in vivo skin," *Proc. SPIE* **1882**, 278–290 (1993).
22. O. Ščepanovic, Z. Volynskaya, C. R. Kong, L. Galindo, R. R. Dasari, and M. S. Feld, "A multimodal spectroscopy system for real-time disease diagnosis," *Rev. Sci. Instrum.* **80**(4), 043103 (2009).
23. S. E. J. Bell, E. Bourguignon, and A. Dennis, "Analysis of luminescent samples using subtracted shifted Raman spectroscopy," *Analyst* **123**, 1729–1734 (1998).
24. A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.* **36**(8), 1627–1639 (1964).
25. R. G. Brereton, *Applied Chemometrics for Scientists*, John Wiley and Sons Ltd., Chichester, West Sussex, England (2007).
26. W. C. Shih, "Quantitative biological Raman spectroscopy for non-invasive blood analysis," PhD Thesis, Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA (2007).
27. J. Zhao, M. M. Carrabba, and F. S. Allen, "Automated fluorescence rejection using shifted excitation Raman difference spectroscopy," *Appl. Spectrosc.* **56**(7), 12 (2002).
28. I. Osticioli, A. Zoppi, and E. M. Castellucci, "Shift-excitation Raman difference spectroscopy difference deconvolution method for the luminescence background rejection from Raman spectra of solid samples," *Appl. Spectrosc.* **61**(8), 839–844 (2007).
29. S. T. McCain, R. M. Willet, and D. J. Brady, "Multi-excitation Raman spectroscopy technique for fluorescence rejection," *Opt. Expr.* **16**(15), 10975–10991 (2008).
30. B. D. Beier and A. J. Berger, "Method for automated background subtraction from Raman spectra containing known contaminants," *Analyst* **134**, 1198–1202 (2009).
31. A. Cao, A. K. Pandya, G. K. Serhatkulu, R. E. Weber, H. Dai, J. S. Thakur, V. M. Naik, R. Naik, G. W. Auner, R. Rabah, and D. C. Freeman, "A robust method for automated background subtraction of tissue fluorescence," *J. Raman Spectrosc.* **38**(9), 1199–1205 (2007).
32. H. M. Heise, R. Marbach, T. H. Koschinsky, and F. A. Gries, "Non-invasive blood glucose sensors based on near-infrared spectroscopy," *Artif. Organs* **18**(6), 439–447 (1994).
33. C. Fischbacher, K. U. Jagemann, K. Danzer, U. A. Müller, L. Papenkordt, and J. Schüler, "Enhancing calibration models for non-invasive near-infrared spectroscopical blood glucose determination," *Fresenius J. Anal. Chem.* **359**(1), 78–82 (1997).
34. J. B. O'Sullivan and C. M. Mahan, "Criteria for the oral glucose tolerance test in pregnancy," *Diabetes* **13**, 278–285 (1964).
35. D. A. Sacks, J. S. Greenspoon, S. Abu-Fadil, H. M. Henry, G. Wolde-Tsadik, and J. F. F. Yao, "Obstetrics: toward universal criteria for gestational diabetes: the 75-gram glucose tolerance test in pregnancy," *Am. J. Obstet. Gynecol.* **172**(2), 607–614 (1995).

36. D. I. Phillips, P. M. Clark, C. N. Hales, and C. Osmond, "Understanding oral glucose tolerance: comparison of glucose or insulin measurements during the oral glucose tolerance test with specific measurements of insulin resistance and insulin secretion," *Diabetic Med.* **11**(3), 286–292 (1994).
37. S. E. J. Bell, E. S. O. Bourguignon, A. C. Dennis, J. A. Fields, J. J. McGarvey, and K. R. Seddon, "Identification of dyes on ancient Chinese paper samples using the subtracted shifted Raman spectroscopy method," *Anal. Chem.* **72**(1), 234–239 (2000).
38. A. O'Grady, A. C. Dennis, D. Denvir, J. J. McGarvey, and S. E. Bell, "Quantitative Raman spectroscopy of highly fluorescent samples using pseudosecond derivatives and multivariate analysis," *Anal. Chem.* **73**(9), 2058–2065 (2001).
39. I. Lewis and H. G. M. Edwards, *Handbook of Raman Spectroscopy* (*Practical Spectroscopy*), Marcel Dekker, New York (2001).
40. G. P. Singh, G. Volpe, C. M. Creely, H. Groetsch, I. M. Geli, and D. Petrov, "The lag phase and G1 phase of a single yeast cell monitored by Raman microspectroscopy," *J. Raman Spectrosc.* **37**, 858–864 (2006).
41. U. Thissen, B. Ustun, W. J. Melssen, and L. M. C. Buydens, "Multivariate calibration with least-squares support vector machines," *Anal. Chem.* **76**(11), 3099–3105 (2004).