Multimodal multiplex Raman spectroscopy optimized for in vivo chemometrics

S. T. McCain, M. E. Gehm, Y. Wang,
N. P. Pitsianis, and D. J. Brady

Duke University Fitzpatrick Center for Photonics and Communication Systems and
Department of Electrical and Computer Engineering, Box 90291, Durham, NC, 27708, USA

ABSTRACT

We have designed and constructed a multimodal multiplex Raman spectrometer which uses multi-wavelength excitation to better detect signals in the presence of fluorescence by taking advantage of the shift-variance of the Raman signal with respect to excitation frequency. Coupled with partial-least-squares (PLS) regression, the technique applied to ethanol estimation in a tissue phantom achieves root-mean-squared-cross-validation errors (RMSCVE) of 9.2 mmol/L with a model formed with 2 principal components, compared to a single wavelength data set with equivalent energy where 7 principal components were used to achieve an RMSCVE of 39.1 mmol/L.

Keywords: Raman Spectroscopy, Tissue Spectroscopy, Chemometrics, Multivariate Methods

1. INTRODUCTION

Raman spectroscopy is a powerful diagnostic tool due to its high specificity and possibility for *in vivo* applications.[1] At the same time, its very weak signal strength and incoherent scattering properties make it a challenging signal to detect. In biological systems where components scatter strongly and fluoresce, this detection problem is compounded even further. We have designed and constructed a multimodal multiplex Raman spectrometer which uses coded aperture spectroscopy to have a large collection area and multi-wavelength excitation to better detect signals in the presence of fluorescence.

Raman chemometrics has been shown to be very effective in determining analyte concentrations in biological media such as urine[2] and aqueous humor.[3] In whole blood and tissue, however, systems are not as successful because of the presence of interferences such as autofluorescence of non-target molecules, intense scattering, and strong absorption by other constituents. Recent work has shown good progress,[4] however high excitation powers and exposure times are necessary, and robust models with repeatable multi-day and multi-person performance have yet to be demonstrated. Through the use of multi-wavelength excitation it should possible to form robust models which improve the performance of the system.

2. MULTI-WAVELENGTH RAMAN SPECTROSCOPY

The wavelength of light scattered inelastically from molecular species can be shifted as a result of energy coupling to the internal vibrational modes of the scattering species. Such frequency shifts result in a discrete array of sharp lines corresponding to harmonic modes of the molecule. This spectral shift is generally referred to as the Raman effect. Raman spectroscopy may be used to identify the presence of a target molecule, to analyze the structure or densities of molecular species, or for other purposes. However, fluorescence produced by elastic scattering and stray light obscures the Raman portion of the spectrum.

Shifted Excitation Raman Difference Spectroscopy (SERDS) may be used to reduce interference from fluorescence and stray light by shifting the frequency of a laser light that is incident on a specimen. The Raman bands are generally shifted in response to a shift in excitation frequency, while the broad background fluorescence and stray light are generally much less affected by the excitation frequency shift. SERDS generally involves a

Further author information: (Send correspondence to David J. Brady: E-mail: dbrady@duke.edu, Telephone: 1 919 660 5598

Biomedical Vibrational Spectroscopy III: Advances in Research and Industry,
edited by Anita Mahadevan-Jansen, Wolfgang H. Petrich,
Proc. of SPIE Vol. 6093, 60930P, (2006) · 1605-7422/06/$15 · doi: 10.1117/12.646755

subtraction of two spectra obtained from two different excitation frequencies. The subtraction can result in a derivative spectrum that may reduce the background and fluorescence spectra.[5]

However, SERDS may not sufficiently reduce fluorescence and stray light in all environments. For example, Raman spectroscopy may be more difficult in diffuse media, such as biological tissue, because the Raman effect may be obscured by a high level of incoherent fluorescence spectra or by the diversity of molecules and the associated Raman signals that may be present in the media.

A spectral impulse response is used to make a mathematical distinction between a non-Raman signal (i.e., the component of a spectrum that can be attributed to non-Raman scattering) and the Raman signal (i.e., the component of a spectrum that can be attributed to Raman scattering). Both spectral components include sub-components that are linear and non-linear with respect to the exciting wavelengths. However, the linear components are generally stronger than the non-linear components. The linear terms are described by an impulse response. The non-Raman spectrum, S_{NR}, is assumed to be constant for small shifts in excitation wavelength, leading to an impulse response of the form $h_R(\nu, \nu') = h_R(\nu)$, where ν is the measured optical output frequency, and ν' is the excitation frequency. Thus the non-Raman signal can be expressed as

$$S_{NR} \simeq \int h_{NR}(\nu) S_e(\nu') d\nu', \tag{1}$$

where S_e is the exciting signal as a function of optical output frequency. The distinguishing feature of the Raman and the non-Raman signals is that the Raman impulse response is typically shift invariant in the exciting field, while the non-Raman impulse response is not shift invariant. The term "shift invariant" means that the Raman impulse response is a function only of the difference between the observed frequency and the exciting frequency, e.g., $h_R(\nu, \nu') = h_R(\nu - \nu')$. Therefore, the Raman spectrum is expressed as follows:

$$S_R = \int h_R(\nu - \nu') S_e(\nu') d\nu'. \tag{2}$$

The total inelastically scattered spectrum from a source is the sum of the non-Raman and the Raman terms. An array of individual laser sources that can be modulated can be used to implement a coded Raman spectrometer system. By designing an impulse response which uses multiple shifts, the spectrum can be separated into the Raman and non-Raman components by taking into account the differences in the impulse responses of the Raman and non-Raman features.

Instead of explicitly solving for the Raman portion of the signal, we propose to use the prior knowledge of the excitation spectrum of an array of lasers to shift the measured Raman spectra. These spectra can then be summed, and PLS-regression used to determine which portions of the signal correspond to the molecule of interest. By shifting the spectra, the Raman peaks will overlap and add, whereas any non-Raman signals will be distributed among many spectral channels thus making the estimation more robust against unknown interferents.

2.1. Multivariate Chemometrics

Multivariate techniques are often used in spectroscopy in order to provide a dimensionality reduction, since many spectral channels are collinear, responses of analytes typically overlap, and many interferences can be accurately modeled.[6] The goal of chemometrics is to estimate the concentration of a given analyte or analytes in an unknown mixture. In spectroscopic analysis, a spectra S of length n is commonly acquired, and thus a mapping needs to made between concentration and the acquired data, which can be expressed as,

$$c = f(S). \tag{3}$$

Such mappings have been explored extensively, and typically are found by using a partial least squares (PLS) regression on a calibration data set of known concentrations.[6] A regression vector, w, of length n is then found which can be multiplied by an unknown spectrum to obtain an estimated concentration, $\hat{c}$

$$\hat{c} = S \cdot w. \tag{4}$$

This regression vector is formed by a limited number of weighting vectors (components) which are found using an iterative algorithm. The number of components to use is typically found by cross-validation techniques.

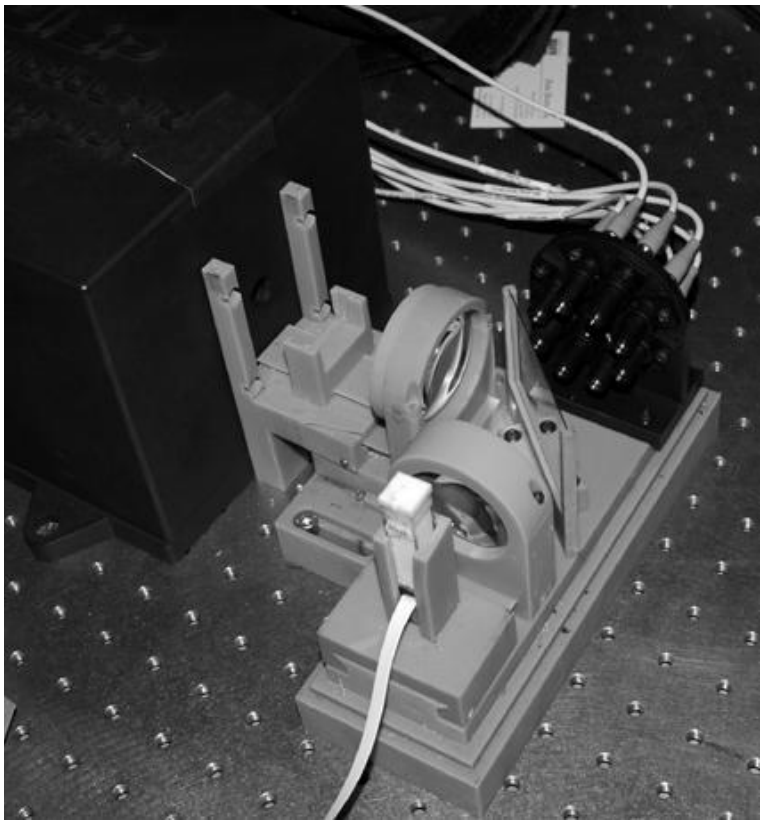

Figure 1. The illumination system. A ring of 8 lasers (800-808 nm) is used to excite the sample through a dichroic filter and focusing lens. Raman scattered photons are then collimated by the same lens, reflected by the dichroic filter, then imaged by another lens onto the input aperture of the spectrometer.

3. PROTOTYPE SYSTEM

To investigate the quantitative performance of the shifted-excitation technique, experiments were performed using a tissue phantom. Varying amounts of ethanol were added to a solution of Intra-Lipid with scattering properties similar to that of tissue in order to have a realistic phantom. Spectra were then collected and analyzed using multivariate techniques to assess the chemometric performance.

3.1. Optical system

A high throughput coded aperture spectrometer was used to obtain spectral measurements of the sources. Its description and performance have been described in an earlier work.[7] The illumination and collection system is designed for 8 excitation lasers, which are fiber-coupled and placed in a ring pattern as shown in Fig.1. The 8 lasers are butterfly packaged and temperature and drive current regulated to assure wavelength stability. The lasers are set at temperatures between 15 and 30°C in order to equally space their center wavelength between $\simeq 800 - 808$ nm in 1 nm increments with 30 mW of optical power on the sample for each laser. A small lens tube containing a collimating lens and bandpass filter is mounted in front of each fiber output.

The laser light passes through a dichroic filter, and is then focused onto the sample by an aspheric lens onto the sample. By translating the lens tube assemblies, all 8 lasers can be focused onto the same location on the sample. An aspheric lens (f=25 mm, D=45 mm) then collects the scattered light, which is then reflected off the dichroic filter towards the spectrometer. A longer focal length spherical lens (f=100 mm, D=50 mm) then forms a magnified image of the scattered light from the sample onto the aperture mask of the spectrometer. The

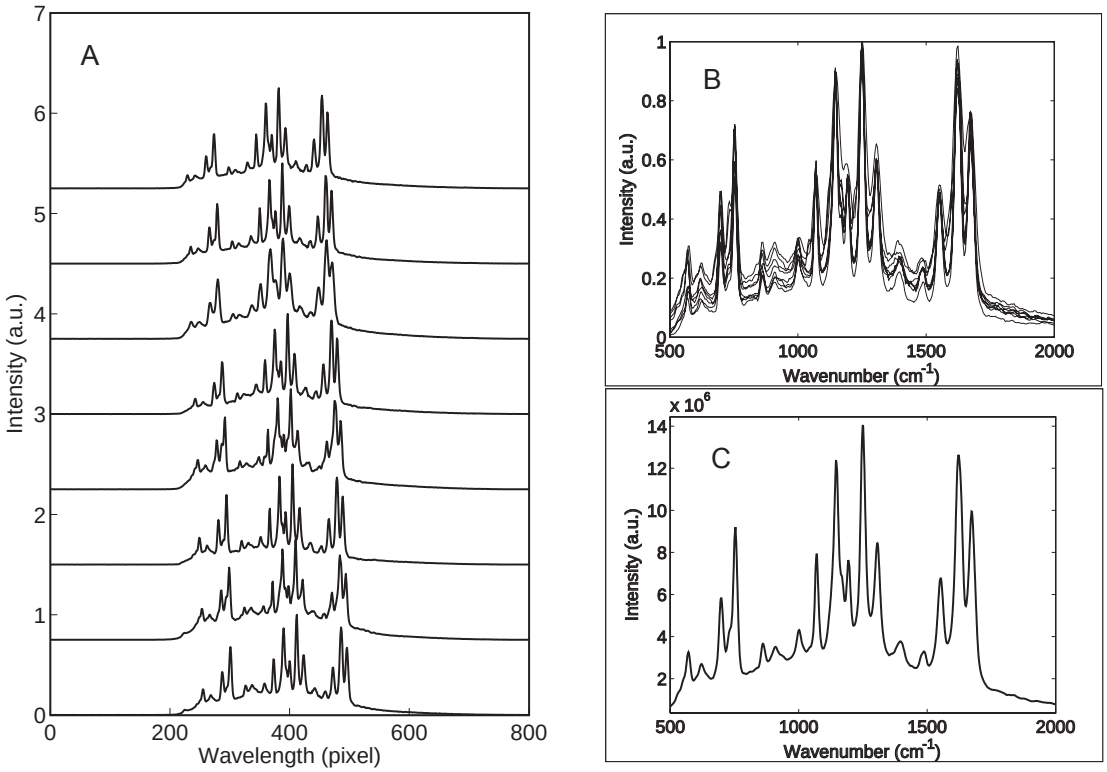

Figure 2. (A) Acetaminophen Raman spectra (5 s exposure) taken with each excitation laser, spectra separated for visual purposes only (B) Aligned spectra after fitting peak locations to known peaks of Acetaminophen, (C) Sum of all 8 remapped spectra

coded aperture spectrometer then measures the average spectrum impinging on the mask, and the resulting data is stored on a computer for later processing.

3.2. Spectral Pre-processing

In order to combine the spectra for the different excitation wavelengths, a calibration sample of acetaminophen is used since its Raman spectrum known to a very high accuracy and it has peaks which span the bandwidth of the instrument ($500 - 2000$ cm^{-1}). A spectrum is gathered by each of the lasers individually, as shown in Fig.2 A. Simply shifting the spectra by the necessary pixels is unable to account for the nonlinear relationship between wavelength and pixel which is inherent to spectrometers. To account for this, the pixel locations of the acetaminophen peaks for each laser are then found, and then matched to the known location of the peaks. A least squares linear fit can be then done between the measured peak locations and their wavenumber locations for each excitation laser. The spectra are then resampled using a cubic-spline interpolation onto a vector where the index of the vector corresponds to wavenumber. In this way, the Raman peaks from the 8 excitation lasers are all aligned, as can be seen in Fig.2 B. The described technique provides no significant loss of resolution across the entire wavenumber range as can be seen when the spectra are summed in Fig.2 C. The necessary resampling of the spectra is then saved, and then used on subsequent unknown spectra to align the Raman features.

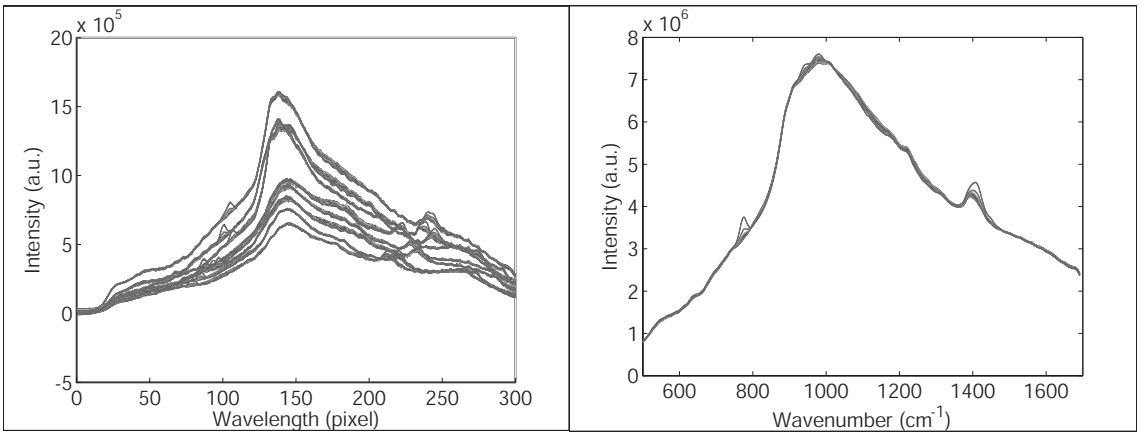

Figure 3. (Left) Entire data set (8 concentrations, 8 wavelengths, 3 exposures each), after multiplicative signal correction has been applied to each excitation wavelength set of spectra (Right) Shifted and summed data set after combining multi-wavelength spectra into a single spectrum for each sample.

4. RESULTS

In order to assess the performance of the shifted excitation technique, mixtures of Intra-Lipid and ethanol were prepared in 7 concentrations between 0.05 and 2% by weight (8.7-348 mmol/L) along with a sample of pure Intra-Lipid. Intra-Lipid is used as a tissue phantom, since its optical scattering and fluorescence properties are similar to that of human tissue. Each sample was individually measured by all 8 lasers, with 3 exposures of 60 seconds for each sample. Thus for all 8 samples, there are 3 spectra for each 8 excitation wavelength, leading to a data set with 192 spectra. To account for laser fluctuations, before the spectra are resampled and summed, a multiplicative signal correction is performed for each excitation laser's data set.[8] The resultant spectra are shown in Fig.3. The spectra are then resampled according to the calibration with acetaminophen, and then the data from each excitation wavelength is summed for each sample, leading to a dataset which has 24 spectra. This resultant 2-D data matrix, along with a vector containing the known concentrations of ethanol for each sample, can be analyzed using PLS-regression. Due to the limited amount of data, model formation and validation was performed on the same dataset. A contiguous block cross-validation was performed, wherein the three samples for each concentration are left out to form a model, and then the resultant model is tested on the left-out samples. This was used to provide a more realistic cross-validation model, where no measurements of the unknown sample are used in model formation.

To compare the performance of the technique against single-channel excitation, a previous dataset was included in the analysis. In this dataset, a single laser was used with identical sample conditions, however the excitation was doubled to 60 mW and 4 minute exposures were used. Thus, equivalent energy was used, since the multi-wavelength analysis contained 8 lasers at 30 mW. Similar contiguous block cross-validation was performed on the dataset and the performance was compared.

4.1. Discussion

To explore the performance of the system, the root-mean-squared-error-cross-validation (RMSECV) was used to both analyze the error and to determine how many latent variables are needed in the PLS-regression in order to form a robust model. The RMSECV versus number of components used to form the model is shown in Fig.4. The multi-wavelength dataset achieves a much lower error, 9.2 mmol/L, compared with 39.1 mmol/L for the single wavelength excitation. The number of components needed to reach the minimum error is also much less, 2 compared with 7, suggesting that its model is more robust. To verify that the regression models were indeed locking on to the Raman signature of ethanol, the regression vectors shown in Fig.4 were compared to the Raman spectrum of pure ethanol. The peaks in the ethanol spectrum are clearly present in the regression vectors, and

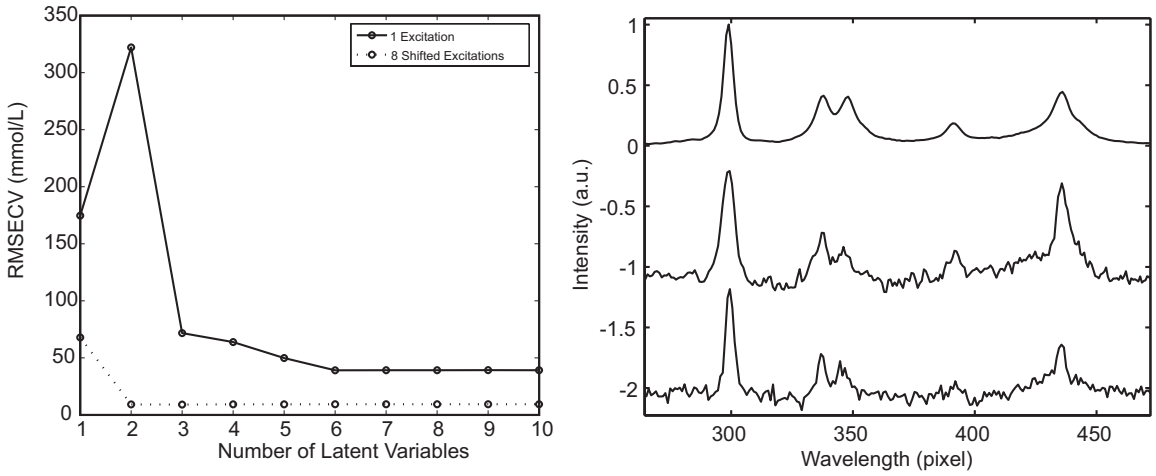

Figure 4. (Left) Root-mean-square-error-cross-validation with PLS-regression, solid line is single wavelength excitation, dashed line is 8 wavelength excitation with equivalent energy, (Right) Top, Raman spectrum of pure ethanol, middle, PLS-regression vector for multi-wavelength excitation formed with 2 principal components, bottom, PLS-regression vector for single-wavelength excitation formed with 7 principal components.

the right-most peaks in the spectrum are more clearly present in the regression vector for the multi-wavelength excitation.

To investigate the performance further, scatter plots were generated, shown in Fig.5. While at the higher concentrations, performance between the single and multi-wavelength excitation techniques are similar, at the lower end of the concentration scale the multi-wavelength estimations are much more accurate. The 0% values for the single-wavelength data were so far negative that they are not shown on the plots.

5. CONCLUSIONS

A multi-wavelength Raman spectroscopy system has been demonstrated and achieves cross-validation performance better than a single-wavelength excitation system for ethanol chemometrics in a lipid tissue phantom. By taking advantage of the shift-variance of the Raman effect, spectra recorded from different excitation wavelengths can be shifted and summed in order to emphasize the Raman portions of the signal. In cross-validation of the PLS regression models, this is demonstrated in the reduction of number of factors needed to form a model with minimum cross-validation error, from 7 to 2, as compared to a single-wavelength case.

The proposed technique should improve efforts to achieve *in vivo* molecular estimation. By using multi-wavelength Raman spectroscopy, systems could be more robust against fluctuations and interferences that are unavoidable in such applications. By further using more complicated excitation patterns and processing techniques, individual molecular signatures could be filtered instead of only Raman features.

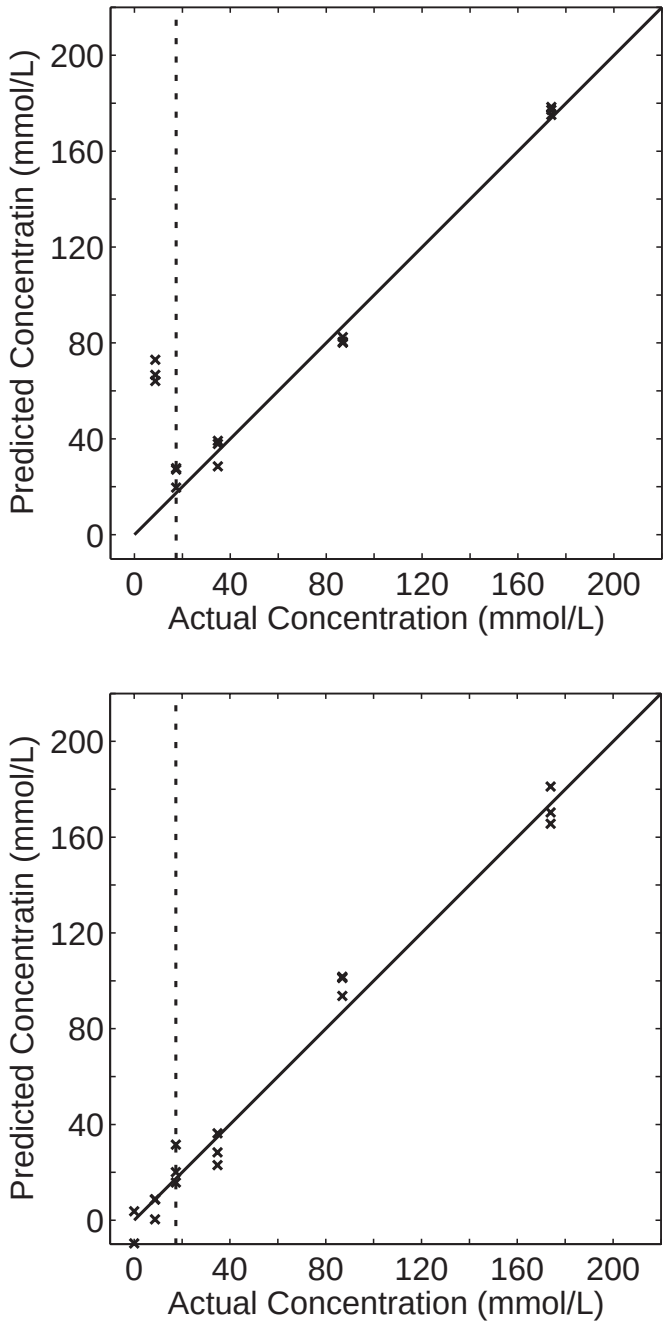

Figure 5. (Top) Cross-validation prediction with single-wavelength excitation using a 7 component model, RMS error of 39.1 mmol/L (Bottom) Cross-validation prediction with multi-wavelength excitation using a 2 component model, RMS error of 9.2 mmol/L. Solid lines are known concentration, not fits of data. Dashed line common legal intoxication limit for blood alcohol concentration.

6. ACKNOWLEDGMENTS

This work was supported by the National Institute on Alcoholism and Alcohol Abuse (NIAAA).

REFERENCES

1. E. B. Hanlon, R. Manoharan, T. W. Koo, K. E. Shafer, J. T. Motz, M. Fitzmaurice, J. R. Kramer, I. Itzkan, R. R. Dasari, and M. S. Feld, "Prospects for in vivo raman spectroscopy," *Phys. Med. Biol.* **45**(2), pp. 1–59, 2000.
2. X. Dou, Y. Yamaguchi, H. Yamamoto, S. Doi, and Y. Ozaki, "Quantitative analysis of metabolites in urine using a highly precise, compact near-infrared raman spectrometer," *Vib. Spectrosc.* **13**(1), p. 83, 1996.
3. 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**(8), pp. 1024 – 1031, 2005.
4. A. M. K. Enejder, T. G. Scecina, J. Oh, M. Hunter, W. C. Shih, S. Sasic, G. L. Horowitz, and M. S. Feld, "Raman spectroscopy for noninvasive glucose measurements," *J. Biomed. Opt.* **10**(3), p. 031114, 2005.
5. S. E. J. Bell, E. Bourguignon, and A. Dennis, "Analysis of luminescent samples using subtracted shifted raman spectroscopy," *Analyst* **123**, pp. 1729–1734, 1998.
6. H. Martens and T. Næs, *Multivariate Calibration*, John Wiley and Sons, Chichester, U.K., 1989.
7. S. T. McCain, M. E. Gehm, Y. Wang, N. Pitsianis, and D. J. Brady, "Coded aperture raman spectroscopy for quantitative measurements of ethanol in a tissue phantom," *Appl. Spec.* , 2006. Under Review.
8. H. Martens, J. P. Nielsen, and S. Engelsen, "Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures," *Anal. Chem.* **75**, pp. 394–404, 2003.