Published in final edited form as:
Anal Chem. 2010 December 1; 82(23): 9719–9726. doi:10.1021/ac101754n.

Development of robust calibration models using support vector machines for spectroscopic monitoring of blood glucose

Ishan Barman[*], **Chae-Ryon Kong**, **Narahara Chari Dingari**, **Ramachandra R. Dasari**, and **Michael S. Feld**[‡]
Laser Biomedical Research Center, G. R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Abstract

Sample-to-sample variability has proven to be a major challenge in achieving calibration transfer in quantitative biological Raman spectroscopy. Multiple morphological and optical parameters, such as tissue absorption and scattering, physiological glucose dynamics and skin heterogeneity, vary significantly in a human population introducing non-analyte specific features into the calibration model. In this paper, we show that fluctuations of such parameters in human subjects introduce curved (non-linear) effects in the relationship between the concentrations of the analyte of interest and the mixture Raman spectra. To account for these curved effects, we propose the use of support vector machines (SVM) as a non-linear regression method over conventional linear regression techniques such as partial least squares (PLS). Using transcutaneous blood glucose detection as an example, we demonstrate that application of SVM enables a significant improvement (at least 30%) in cross-validation accuracy over PLS when measurements from multiple human volunteers are employed in the calibration set. Furthermore, using physical tissue models with randomized analyte concentrations and varying turbidities, we show that the fluctuations in turbidity alone causes curved effects which can only be adequately modeled using non-linear regression techniques. The enhanced levels of accuracy obtained with the SVM based calibration models opens up avenues for prospective prediction in humans and thus for clinical translation of the technology.

1. Introduction

Disorders of glucose homeostasis, including types 1 and 2 diabetes, as well as gestational diabetes, represent a leading cause of morbidity and mortality worldwide. Failure to adequately regulate blood glucose leads to acute and chronic health complications[1, 2] making frequent monitoring of glucose levels imperative. To accomplish this in a painless and non-invasive manner, optical techniques such as NIR Raman spectroscopy, which combines excellent chemical specificity and adequate sampling volume, have been proposed.[3] Several research groups, including our own, have reported accurate results for Raman spectroscopy-based detection of glucose (and other blood analytes) in serum, whole blood and human aqueous humor[4-6]. Initial studies in human volunteers have also provided promising results when individual calibration schemes are developed on each volunteer ("local calibration model")[7]. However, when such a scheme is developed on multiple volunteers - simulating its performance over a larger population ("global calibration model") - it has failed to show adequate predictive ability. This is clearly undesirable, as it is important for the developed calibration model to be transferable.

[*] To whom correspondence should be addressed. ishan@mit.edu; fax: +1-617-253-4513.
[‡] Deceased

Multiple factors can be attributed to the current difficulty in achieving successful calibration transfer including variations in tissue absorption and scattering (turbidity)[8] , variations in fluorescence levels and associated photobleaching[9] and differences in glucose concentrations in the blood and interstitial fluid compartments[10,11]. Moreover, any spurious correlation introduced into the calibration model will adversely affect its prospective prediction capability. Over the years, investigators have developed various correction schemes, incorporating changes in system design as well as spectral processing, to account for specific sources of variability (such as methods to correct for the turbidity-induced changes in sampling volume[12,13]). Nevertheless, given the broad range of such sources of variance, it is crucial that the calibration model is robust to non-analyte specific variations in the spectral data. Unfortunately, conventional implicit multivariate calibration (MVC) methods are prone to be misled by spurious effects such as system drift and covariance among constituents. To provide a more enhanced robustness than that afforded by the conventional calibration techniques[14], such as partial least squares (PLS) and principal component regression (PCR), our laboratory has previously developed hybrid MVC methods, which incorporate prior information of the system (e.g. pure spectrum of the analyte of interest) into the implicit calibration framework[15,16].

Despite the improvements provided by the hybrid methods, all of the above schemes are limited by the underlying assumption that the relationship between the spectra and the property of interest (analyte concentration) is linear. This is a reasonable first approximation for tissue Raman spectra as it has been previously shown that the Raman spectrum of a mixture sample can be treated as a linear superposition of the mixture's component spectra[17,18]. However, the assumption of linearity may fail under the influence of fluctuations in process and system variables, such as changes in temperature, sampling volume and physiological glucose dynamics. It is worth mentioning that weak non-linearities can be modeled by the conventional MVC methods by retaining larger number of factors than is necessitated by the chemical rank of the system – thereby also risking the inclusion of irrelevant sources of variance and noise in the calibration model.

Alternatively, the curved effects could be modeled by non-linear methods. Support vector machines (SVM) provide a relatively new class of such methods, which can handle ill-posed problems and lead to unique global models. Since its initial formulation by Vapnik and co-workers[19,20], SVM has been used extensively for classification problems in bioinformatics[21] and chemometrics[22]. Moreover, SVM has been extended to develop non-linear regression models capable of quantitative prediction. For example, SVM calibration models have exhibited excellent potential for near-infrared (NIR) absorption-based concentration prediction in chemical mixtures, even when the acquired spectra are non-linearly affected by temperature fluctuations[23].

In this article, we first study the relationship between glucose concentrations and Raman spectral datasets; previously acquired from multiple human volunteers during the performance of oral glucose tolerance tests.[7] PLS implementation on the human subject dataset clearly shows the presence of curved effects in the aforementioned relationship. To account for the deviation from linearity, we employ non-linear SVM calibration. We observe that SVM models provides a significant improvement over the PLS models with a reduction in prediction error by at least 30%, when multiple human subject datasets are included in the analysis. Importantly, it also makes substantially better predictions in the hypoglycemic range.

Further, we hypothesize that one source of the observed non-linearity in the human subject dataset is the variations in tissue turbidity. To test our hypothesis in a controlled environment, we perform a set of physical tissue model (tissue phantom) studies with

varying turbidities and randomized concentrations of the analyte of interest (glucose) and spectral interferents. Analysis of the tissue phantom data also reveals that SVM outperforms PLS, and PLS used in conjunction with a turbidity-correction scheme (turbidity corrected Raman spectroscopy, TCRS)[13]. This suggests the ability of SVM to provide an accurate calibration model, even in the presence of such non-analyte specific variances. Taken together, these results have important implications for clinical translation of Raman spectroscopy-based glucose monitoring and for other biological applications where sample-to-sample variability could lead to significant non-linear relationships between the property of interest and the measured spectra.

2. Theory of support vector regression

The theory of SVM and its extension to regression analysis has been described in detail elsewhere in the literature[24, 25]. For orientation, the primary equations of ε-SVM (epsilon-SVM) regression are provided in the Supporting Information (Sec.-S1). The central idea of this method is to find a function $f(x)$ that has at most ε deviation from the actually obtained targets for all the calibration data (epsilon-tube criterion). In support vector regression, a Lagrangian problem is formulated that balances this criterion with the minimization of the regression coefficients (since large regression coefficients are prone to overfitting). Solution of this problem leads to the following regression model:

$$f(x) = \sum_i \left(\alpha_i - \alpha_i^*\right) K(x_i, x) + b \tag{1}$$

where x_i and x represent the calibration and prediction data, respectively. b is the bias and α_i, α_i^* are the Lagrange multipliers for the inequality constraints, given in Eq. (S1.1-S1.3) of Supplementary Information. The non-linear transformation is performed via the kernel function $K(x_i,x)$. The most common choices of non-linear kernels are the radial basis function $\left(\exp\left(-\|x_i - x_j\|^2/2\sigma^2\right)\right.$, where σ^2 is the width of the RBF) and the polynomial function ($<x_i,x_j>^d$, where d is the polynomial degree). The kernel parameters, ε and C are optimized by the user as per the training dataset. Details of our SVM implementation procedure are discussed in Sec. 3.2.

3. Materials and Methods

Experimental studies are undertaken to accomplish the following objectives: (1) compare the effectiveness of SVM and conventional PLS calibration models when multiple human subject datasets are incorporated in the analysis and (2) investigate the effect of turbidity-induced sampling volume variations on the relationship between the reference concentrations and the acquired Raman spectra. To fulfill objective (1), multivariate analysis is undertaken on datasets obtained from our laboratory's clinical studies on human volunteers. In addition, the study is used to understand the nature of non-linearity introduced, if any, as multiple human volunteers are included in the calibration model. To accomplish objective (2), a tissue phantom study is designed where the concentrations of the analyte of interest and spectral interferents as well as the turbidities of the phantoms are randomly varied. The tissue phantom study provides a degree of control and specificity, which is difficult to implement in a human subject or animal model study. In the latter cases, fluctuations in other parameters including tissue fluorescence, skin heterogeneity and physiological glucose kinetics would obscure the effect of turbidity variation on the relationship between the spectra and concentration datasets. In this study, Raman and diffuse reflectance spectra are collected from the tissue phantoms for comparison of the PLS, SVM

and turbidity corrected PLS calibration models. Correction for turbidity variation is performed in accordance with the TCRS procedure.

3.1 Experimental section

3.1.1 Human volunteer study—The acquisition of the clinical data from the human volunteers was originally described in one of our laboratory's previous publications[7]. Briefly, Raman spectra were acquired from the forearms of healthy Caucasian and Asian volunteers undergoing oral glucose tolerance tests (OGTT). An 830 nm diode laser was used as the Raman excitation source and a gold-coated half-paraboloidal mirror was employed to enhance the collection efficiency of the system. The collected light was fiber-coupled to a f/1.8 spectrograph for dispersion before spectral acquisition using a liquid nitrogen cooled deep depletion CCD. In accordance with a typical OGTT protocol, each volunteer was given a glucose-rich drink. Following this, spectra were collected every 5 minutes, with an acquisition time of 3 minutes, over a 2-3 hour measurement period. In addition, finger prick measurements of blood glucose levels were performed at 10 minute intervals using a HemoCue glucose analyzer. The human volunteer studies were approved by the Massachusetts Institute of Technology Committee On the Use of Humans as Experimental Subjects. Informed consent was obtained from all subjects prior to their inclusion in the OGTT study.

3.1.2 Tissue phantom study—For this study, the basic experimental setup is kept unchanged except for the introduction of a broadband source (tungsten-halogen lamp) for collection of diffuse reflectance spectra. The excitation-collection geometry for the acquisition of the diffuse reflectance spectra is identical to that for the Raman spectra. Appropriate bandpass filters are added in the path of the broadband source so that the diffuse reflectance spectra are acquired over the same wavelength range as the Raman spectra. For this study, 50 tissue phantoms are formed by pipetting out randomized concentrations of glucose (analyte of interest), creatinine (Raman spectral interferent), intralipid (anisotropic elastic scatterer) and India ink (absorber) into distilled water. The elastic scattering (μ_s) and absorption (μ_a) values in these tissue phantoms are varied over the ranges typically observed in skin tissue (40 to 100 cm^{-1} and 0.05 to 0.2 cm^{-1}, measured at 830 nm)[26]. The glucose and creatinine concentrations are varied in the range of 4-32 mM. Spectroscopic measurements are performed on aliquots of these tissue phantoms in a fused silica cuvette. Between each set of measurements, the cuvette is washed three times with distilled water. Raman and diffuse reflectance spectra were acquired alternately for 20 seconds each from the samples. The ambient temperature in this study is maintained at 73° ± 0.5°F.

3.2 Data analysis

For both human volunteer and tissue phantom studies, conventional linear regression (PLS) and non-linear SVM regression methods are employed.

3.2.1 Human volunteer study—Of the 17 volunteers on whom the OGTT were successfully performed, 4 volunteers exhibited impaired glucose tolerance profiles. These 4 volunteer datasets are excluded from further analysis, given the risk of incorporating temporal spurious correlations between the nearly monotonic glucose profiles and spectral altering parameters (such as fluorescence quenching and system drift)[27]. Spectra from 355-1545 cm^{-1} are used in all data analysis after image curvature correction, vertical binning and cosmic ray removal. The datasets from the human volunteers were directly input for development of the MVC calibration models without any data manipulation such as removal of fluorescence.

The 13 volunteer datasets constituting a total of 350 data points (spectra and concentrations) are first analyzed using a PLS leave-one-out cross-validation (LOOCV) routine to construct a “global” calibration model. The purpose of the LOOCV process is to optimize the parameters, namely loading vectors for PLS and C and σ^2 for SVM. As in a typical LOOCV procedure, one data point is left out at a time from the calibration data and the developed model is used to compute the concentration of the left out data point. This procedure is then repeated until all data points have been left out in turn. The resultant root mean square error of cross validation (RMSECV) is used as a metric for tuning the parameters of the MVC models. For the PLS regression analysis, the number of loading vectors which provide the minimum RMSECV is first determined. Next, to reduce chances of overfitting, the calibration model corresponding to the minimum number of loading vectors, which gives a less than 5% deviation from the previously obtained minimum RMSECV, is selected. (Moreover, for all PLS analysis we satisfied the criterion that the number of calibration samples is at least 3 times larger than the rank of the calibration model, i.e. the number of loading vectors[28].) In addition to global calibration, PLS “local” calibration is performed by using the LOOCV protocol on each of the 13 volunteer datasets individually.

The calculations pertaining to the multivariate non-linear calibration are performed using a SVM MATLAB toolbox[29]. The RBF kernel is used for non-linear regression. Prior to SVM leave-one-out cross-validation, the Raman spectra are linearly scaled such that the intensity values were distributed between 0 and 1. The scaling is performed by dividing each spectrum by the maximum intensity value across the entire spectral range under consideration. The scaling step is performed to avoid the dominance of specific pixel intensity values that reside in greater numeric ranges over those having smaller numeric values. The optimal model parameters C and σ^2 (i.e. the parameters that provide the smallest RMSECV) are obtained using a grid search algorithm in the range of 1 to 10000 (C) and 0.01-10 (σ^2), respectively. Based on the results of prior studies, the ε parameter is kept constant at 0.001. SVM regression is used to generate only a global model – in contrast to PLS implementation where a global as well as multiple local models are generated. In addition, a sample out cross validation study for SVM is performed by omitting each human subject's dataset in turn from calibration and using it for prospective prediction. For this study, we allow a slope and bias term to be fit for each subject (due to the limited number of samples available for calibration). The objective of this specific study is to explore the possibility of prospective prediction in human population.

3.2.2 Tissue phantom study—For the tissue phantom study, the experimental dataset, consisting of 50 samples in all, is randomly split into 36 samples for calibration set and 14 for prediction. PLS and SVM models are first generated on the calibration set based on a leave-one-out cross-validation procedure (similar to that mentioned above in Sec. 3.2.1). For the SVM model, a grid search is performed over the following parameter space: 1 to 20000 (C) and 0.01-10 (σ^2); whereas for the PLS model the loading vectors are varied from 3 to 20. The developed calibration models are then used prospectively on the prediction set to determine the glucose concentrations. Overfitting in this study is avoided by invoking an independent prediction set (i.e. not present in the calibration model) to report the prospective errors (root-mean-square error of prediction, RMSEP). The creation of an independent prediction set is a standard approach to mitigating and/or testing for the presence of spurious correlations[14]. If the data of the calibration set is overfit with a large number of factors, its ability to predict prospective samples will be seriously compromised due to the incorporation of noise and spurious elements. The entire analytical procedure (i.e. splitting of the experimental dataset, followed by calibration and prediction) is iterated 100 times to evaluate an average prediction error value.

Additionally, as mentioned above, turbidity correction is undertaken on the same calibration and prediction datasets using the TCRS formalism. TCRS is based on the principle that Raman and diffusely reflected photons (at the same wavelength) undergo similar scattering and absorption events in turbid media. Thus, by utilizing the information present in the diffuse reflectance spectra, the removal of turbidity-induced Raman spectral distortions can be achieved. Following the TCRS formulation, the Raman intensity at emission wavelengths is divided by the square root of the product of the diffuse reflectance intensities at the laser excitation and Raman emission wavelengths to obtain the "intrinsic" (turbidity-free) Raman spectra. The resulting intrinsic Raman spectra are then analyzed using the PLS procedure. In each of the 100 iterations mentioned above, the calibration and prediction sets for PLS (raw and TCRS-applied) and SVM were identical.

4. Results and discussion

4.1 Human volunteer study

A representative set of tissue Raman spectra and the corresponding blood glucose concentration profile acquired from one of the human volunteers undergoing an OGTT can be seen in Fig. 4 of Barman *et. al*.[11] The spectral features of multiple Raman-active tissue constituents (glucose constitutes only about 0.3% of the observed signal[30]) can be observed, along with the strong autofluorescence background. The blood glucose profile during a typical OGTT exhibits a characteristic rise in glucose levels due to ingestion of the sugar rich drink followed by a return to normal levels mediated by the person's insulin response.

Glucose cross-validation results on the human volunteer datasets - using PLS global (a), PLS local (b), and SVM global (c) - are shown in Fig. 1 plotted on the Clarke error grid[31], a widely used method for evaluating the clinical usefulness of glucose predictions. Predictions in zones A and B are considered acceptable, and predictions in zones C, D, and E are potentially dangerous if used in clinical judgment. From the figure it is evident that application of the PLS global model exhibits significantly larger errors and worse correlations (RMSECV = 28 mg/dl, $R^2 = 0.67$) compared to the SVM global model (RMSECV = 10.87 mg/dl, $R^2 = 0.96$). In fact, the glucose predictions generated with the SVM global model show a better match to the reference glucose concentrations, even compared to those generated using PLS local calibration (RMSECV = 15.45 mg/dl, $R^2 = 0.92$). In other words, the SVM global calibration provides an improvement of *ca*. 61% and 30% over PLS global and PLS local calibration, respectively. It is also worth mentioning that the beneficial impact of nonlinear regression is particularly pronounced in the low analyte concentration region (when the reference glucose concentration is in the range of 4-6 mM). This is a crucial advantage as a major application of a continuous (non-invasive) glucose monitoring system is the accurate diagnosis of hypoglycemic states.

To understand the root cause of this disparity in model performances, the development of the PLS global model is investigated. We find that the PLS model employs increasingly larger number of more loading vectors as more datasets are incorporated in the analysis (Fig. S1, Supporting Information). Clearly, the nearly monotonic rise in the number of optimal loading vectors (with the inclusion of additional volunteers) cannot be attributed to the chemical rank of the system as human skin tissue has a unique and consistent set of principal chemical components[7]. While a small increase in the number of loading vectors over that obtained for a single individual may be explained by the presence of instrumental drift and variable baselines, an increase from 6 LV for 1 human subject to 22 LV for 13 human subjects cannot be ascribed to these factors. This behavior implies the presence of significant curved effects in the relationship between Raman spectra and the reference glucose concentrations. Obviously, the retention of more regression factors in the PLS global model than that warranted by the chemical rank comes at the cost of including

irrelevant sources of variance and noise (as verified by Fig. S2, Supporting Information). In contrast to PLS global calibration, the PLS local calibration models developed individually on each volunteer dataset only employ 5-8 loading vectors. Nevertheless, the PLS local models themselves are susceptible to spurious correlations in the individual dataset that prevents calibration transfer to another individual. For the SVM global model, the optimal combination of C and σ^2 was established to be 1000 and 0.05, respectively. Importantly, this specific combination of C and σ^2 resulted in a less than 2% change in the calculated RMSECV, regardless of the number of volunteers included in the analysis, showing the relative constancy of the optimal parameter combination in clear contrast to the PLS global model.

Further analysis with the inclusion of the previously excluded human subject datasets (i.e. from the volunteers exhibiting impaired glucose tolerance profiles) showed negligible change in the error and correlation values for any of the MVC calibration approaches. Finally, from the SVM sample out cross validation procedure, we observe that, on average, a correlation coefficient of ~0.7 is obtained between the predicted and the reference glucose concentrations, with a maximum of 0.92. This provides strong auxiliary evidence of the SVM model being predictive of the analyte of interest (glucose).

While kernel-based support vector regression enables the non-linear modeling between the spectra and concentration datasets, the sources of the non-linear interferences in this relationship needs to be examined. We attribute these non-linear interferences to fluctuations in process and system variables, which are not specific to the analyte of interest but alter the spectral measurements. The non-linear influence of external process variables such as temperature on vibrational spectra is well-known[32, 33] and is particularly relevant when measurements are performed in an actual process environment where well-controlled laboratory conditions are not achievable. Furthermore, the influence of system (i.e. human subject) parameters such as tissue turbidity and physiological glucose dynamics may also non-linearly affect the spectroscopic measurements in a human population. The effect of tissue turbidity, in particular, has been previously studied in connection with changes in sampling volume. As a direct consequence, the observed spectra from biological tissue are often significantly altered, especially in terms of spectral intensity scaling and peak distortion. (The tissue phantom study results, detailed in Sec.-4.2, elaborate on the influence of turbidity-induced distortions in Raman spectra and the consequences for the predictive ability of multivariate models.) Further information on the influence of physiological glucose dynamics on our human subject results is provided in Supporting Information (Sec.-S2).

Taken together, the fluctuations in these process and system variables sufficiently explain the superior predictive performance of the SVM global model compared to the PLS global model. The improvement in SVM global model performance, compared to that of PLS local calibration, can be mostly attributed to the variation in process variables, given the relative constancy of tissue turbidity and other sample parameters in a specific individual.

Finally, it is worth noting that one of the primary challenges in demonstrating calibration transfer is the establishment of causation between the glucose levels and the optical signals (i.e. whether the source of the signal is truly a change in glucose). For all calibration models (linear or non-linear), spurious effects may arise from different sources that correlate with changes in glucose. Some of these sources may stem from the glucose tolerance test itself, e.g. due to a small physiological change or a change in room temperature which could correlate with observed glucose changes. Here, such correlations are largely mitigated due to the excellent chemical specificity of Raman spectroscopy, as shown by the aforementioned results of the sample out cross validation study. To investigate this issue of causation

further, we employ a tissue phantom study (Sec. 4.2 below), where such non-analyte specific sources can be eliminated due to the controlled environmental settings.

4.2 Tissue phantom study

Effect of turbidity on Raman spectral distortions—Figure 2 represents spectra acquired from tissue phantoms having a fixed concentration of the analyte of interest (Raman scatterer) but different absorption coefficients. Panel (A) and (B) shows the spectra obtained at two different values of the scattering coefficient of the tissue phantom, μ_s = 24 cm^{-1} and μ_s = 130 cm^{-1}, respectively. (These spectra were acquired as part of a 49 tissue phantom study where the Raman scatterer concentration was kept constant and the phantom absorption and scattering coefficients were varied over the range of values observed in human skin tissue. This dataset was originally described in one of our laboratory's previous publications[12].) First, it is clear that the presence of differing amounts of turbidity causes a significant spread in the observed spectral profiles in both (A) and (B). In the absence of multiple scattering and absorption events, the Raman spectra would be expected to be the same for all the samples, as dictated by the constant concentration of the Raman scatterer. Second, we observe that the observed spectra from the tissue phantoms having the higher scattering coefficient (B) show an approximately linear scaling over the entire wavelength range mediated by the value of the absorption coefficient in the tissue phantom. In contrast, at the smaller scattering value of the tissue phantoms in (A), the observed Raman spectra exhibit a more complex scaling behavior - which is more pronounced at the higher absorption values.

These findings can be attributed to the relative interplay of the elastic scattering and absorption in the tissue phantoms. It has been previously observed[13, 34] that when the reduced scattering coefficient $\left(\mu_s^{'}=\mu_s^{*}(1-g)\right)$ is significantly larger than the absorption coefficient (diffusive regime), the observed spectrum (fluorescence or Raman) could be linearly scaled by the diffuse reflectance intensities to extract the intrinsic spectrum of the sample. However, this simple relationship fails when the diffusion approximation is no longer valid, as pointed out by Zhang *et al.*[35]. Our observations in Fig. 2 are in line with such an understanding. The spectra plotted in Fig. 2(A) come from tissue phantoms that do not satisfy the diffusion approximation, especially at μ_a = 0.36, 0.5 and 0.95 cm^{-1} and therefore show a non-linear intensity attenuation. On the other hand, the tissue phantoms of Fig. 2(B) have considerably higher reduced scattering coefficients than the absorption coefficients and thus demonstrate an approximately uniform intensity scaling.

This fact is further borne out when ordinary least squares (OLS) is performed on the raw and TCRS-applied spectral dataset using the model components as the basis spectra. Figure 3 plots the ratio of the observed concentrations to the reference concentrations of the Raman scatterer (C_{obs}/C_{ref}) in the 49 tissue phantoms as a function of μ_s'/μ_a in the tissue phantoms. The dotted black line at C_{obs}/C_{ref} = 1 indicates the position where the glucose concentrations computed from OLS analysis are equal to the reference glucose concentrations in the samples. The red circles indicate the values obtained with raw spectral analysis and the blue circles correspond to those computed with TCRS-applied spectra. It is clear that the OLS analysis of the raw spectra provides incorrect estimates of the concentrations over the whole range of turbidity values observed in skin tissue. (The individual influence of the scattering and absorption parameters on the observed Raman signal, with respect to the actual signal, has been previously discussed by Bechtel *et. al.*[12]) In contrast, the TCRS-applied spectra yield substantially more accurate predictions, especially if the ratio of $\mu_s^{'}/\mu_a$ is greater than 20. Below this threshold ratio, however, both raw and TCRS-applied spectra are ineffective in correctly predicting the analyte concentrations. In other words, the intensity attenuation in the observed Raman spectra does not have a straightforward relationship with the diffuse

reflectance intensities when the tissue absorption is high relative to its scattering. Clearly, this necessitates the incorporation of a non-linear methodology to account for the high absorption regime turbidity variations.

Our observations here are analogous to the findings from Wülfert's seminal study on the effect of temperature on NIR absorption spectra[32]. In this study, the researchers observed that variations in temperature translate via the changes in intermolecular forces into non-linear modifications in the vibrational spectra. Using numerical simulations and experiments with alcohol mixtures, it was shown that this effect can be manifested in terms of a change in peak area, a frequency shift or changing width of the peaks. For our turbidity study, the change in peak area is the predominant concern followed by the changing width of the peaks. It is worth mentioning that the changing peak width is of greater importance in studying fluorescence, which exhibits broader features, than Raman, which has significantly sharper peaks and therefore is less affected by the small local variations in tissue absorption especially in the NIR region. Put another way, the observed Raman spectra primarily undergo scaling distortions and minor local distortions as opposed to fluorescence spectra, which may undergo significant shape variations as well[34-36].

Consequences of turbidity-induced variations on multivariate calibration models—From Fig. 3, we get a clear indication about the detrimental impact of turbidity variation on the concentration estimate of the analyte of interest. Intuitively, one can understand that if the absorption is high then less Raman photons are back-scattered leading to an underestimate on the true concentration of Raman scatterers. While TCRS corrects this by using diffuse reflectance measurements, we find that it is limited by the boundaries of the diffusion approximation. Thus, one needs to incorporate some form of non-linear modeling when the tissue absorption is high relative to its scattering value. There are two approaches which can be used to this end: (i) introduction of a non-linear function which accurately encapsulates the changes in Raman spectra with respect to the diffuse reflectance spectra in the high absorption regime and (ii) introduction of a non-linear multivariate calibration model robust to sample turbidity-induced variations.

In the first approach (i), one can express the intrinsic Raman spectrum as a ratio of the observed spectrum to some pre-calibrated (non-linear) function of the diffuse reflectance spectrum (such as a power law or exponential function)[37]. This function would, in turn, depend on the sample turbidity parameters and excitation-collection geometry. Unfortunately, this would necessitate the explicit determination of the optical properties. In addition, there exists significant uncertainty about the reproducibility of such a pre-calibrated function especially *in vivo*. As a consequence, we propose to follow the second approach (ii) by introducing non-linear kernel-based support vector regression, which may potentially be robust to the turbidity-induced variations.

For our tissue phantom study (Sec.-3.1.2 and 3.2.2), standard PLS procedure was carried out on the raw and TCRS-applied Raman spectra using an optimal of 5 loading vectors. Support vector regression was carried out only on the raw spectra with an optimal (C, σ^2) parameter combination of (10000, 1). Figure 4 shows the boxplot results of the root-mean-square error of prediction (RMSEP) values obtained for the glucose concentrations. The mean prediction errors were 0.55 mM, 0.5 mM and 0.44 mM for the PLS (raw), PLS (TCRS-corrected) and SVM (raw) data, respectively. This demonstrates a 20% reduction in prediction error on application of SVM in the raw case. Importantly, the SVM case improves the prediction accuracy even when PLS is employed on the TCRS-applied spectra by about 12%. This improvement in predictive ability of glucose was observed largely for those tissue phantoms with high absorption values. Given that this study was performed under well-controlled laboratory conditions with temperature stabilization, we can conclude that SVM is able to

deal with the curved effects introduced by the turbidity variations of the various tissue phantoms. Similar results were observed by Thissen *et al*., where they showed that support vector regression provides superior performance and robustness in modeling temperature affected NIR spectra of alcohol and water mixtures.[23] Further, when SVM was applied to TCRS-corrected spectra for the tissue phantom study, we observed negligible alteration (<1%) from the results of the SVM application on the unprocessed spectra. This negligible change in resulting prediction errors can be attributed to the fact that SVM already accounts for the turbidity-induced non-analyte specific variances.

The enhanced robustness provided by SVM models, as compared to PLS, can be attributed to two primary factors in addition to its capability to model non-linear datasets. First, the assignment of relative weights (importance) of training samples using Lagrange multipliers in conjunction with the presence of the ε-insensitive loss function forces a strict distinction between important training samples and irrelevant ones. This ensures that the developed model is able to account for those training samples that would ordinarily provide significantly inaccurate estimates such as tissue phantoms having high absorption values. Second, the regularization term introduced in the formulation penalizes the large regression weights thereby ensuring that small variations in input concentrations and spectra do not lead to large variations in the regression model. Such an approach has been extensively used in ill-conditioned problems (e.g. Tikhonov regularization) where the noisy nature of the data necessitates a robust solution methodology.[38]

Finally, it should be noted that although application of SVM enhances prediction accuracy and robustness, it remains a mathematical approach which does not specifically address the root cause of any particular curved effect such as those introduced by turbidity. Given the large number of variable parameters in non-invasive glucose studies in a human population, we recommend that SVM should be used in conjunction with other approaches such as TCRS that specifically accounts for the variations due to turbidity.

5. Conclusion

While several researchers have reported promising results for blood analyte detection using Raman spectroscopy, calibration transfer in human population has proven to be an outstanding challenge. Given the large number of subject parameters that may vary substantially across a population, robustness of the spectroscopy-based calibration model to such fluctuations is critical to successful prospective application. Linear MVC methods can often be misled by spurious correlations and covariance among constituents and tend to incorporate larger number of factors, especially in the presence of non-linearities. Additionally, large factor models are problematic from a SNR, calibration maintenance and calibration transfer perspective. In this paper, we have proposed the application of non-linear support vector machine regression to enhance the robustness and prediction accuracy of the calibration model. We have demonstrated that SVM provides nearly 30% enhancement in prediction accuracy over PLS calibration models, when measurements from multiple human volunteers are considered. Furthermore, we have shown that variations in turbidity alone may introduce non-linearity into the spectra-concentration relationship, especially when tissue absorption is comparable to its scattering.

We believe that the improvements shown in this article open up the possibility of developing an accurate prediction algorithm for a human population. To fully evaluate the benefits of SVM and other correction approaches in performing prospective prediction in human population, we have initiated (in collaboration with the MIT Clinical Research Center) a larger clinical study, with approximately 100 normal and diabetic volunteers across different age groups. In addition, we are working on incorporating least-squares SVM (LS-SVM) [39],

[40], which significantly reduces the computational time by transforming the principal quadratic programming problem into one of linear programming. Investigating this issue in combination with wavelength subset selection and incorporation of prior information will be the focus of our future research. We expect that the corresponding increase in robustness would also be beneficial in a variety of other biological and industrial applications such as detection of cancer in breast and cervical tissue and real time process monitoring.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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. IB acknowledges the support of the Lester Wolfe Fellowship from the Laser Biomedical Research Center.

References

1. Brownlee M. Nature 2001;414:813–820. [PubMed: 11742414]
2. Ross SA, Gulve EA, Wang M. Chem. Rev 2004;104:1255–1282. [PubMed: 15008623]
3. Khalil OS. Diabetes Technol. Ther 2004;6:660–697. [PubMed: 15628820]
4. Berger AJ, Koo TW, Itzkan I, Horowitz G, Feld MS. Appl. Opt 1999;38:2916–2926. [PubMed: 18319874]
5. Enejder AMK, Koo T, Oh J, Hunter M, Sasic S, Feld MS, Horowitz GL. Opt. Lett 2002;27:2004–2006. [PubMed: 18033426]
6. Lambert JL, Pelletier CC, Borchert M. J. Biomed. Opt 2005;10:1–8.
7. Enejder AMK, Scecina TG, Oh J, Hunter M, Shih W, Sasic S, Horowitz GL, Feld MS. J. Biomed. Opt 2005;10:031114–9. [PubMed: 16229639]
8. Arnold MA, Small GW. Anal. Chem 2005;77:5429–5439. [PubMed: 16131049]
9. Barman, I.; Kong, CR.; Singh, GP.; Dasari, RR. J. Biomed. Opt. ASAP;
10. Steil GM, Rebrin K, Hariri F, Jinagonda S, Tadros S, Darwin C, Saad MF. Diabetologia 2005;48:1833–1840. [PubMed: 16001232]
11. Barman I, Kong CR, Singh GP, Dasari RR, Feld MS. Anal. Chem 2010;82:6104–6114. [PubMed: 20575513]
12. Bechtel KL, Shih WC, Feld MS. Opt. Exp 2008;16:12737–45.
13. Barman I, Singh GP, Dasari RR, Feld MS. Anal. Chem 2009;81:4233–4240. [PubMed: 19413337]
14. Brereton, RG. Applied Chemometrics for Scientists. John Wiley & Sons Ltd.; Chichester, West Sussex, England: 2007.
15. Berger AJ, Koo TW, Itzkan I, Feld MS. Anal. Chem 1998;70:623–627. [PubMed: 9470489]
16. Shih W-C, Bechtel K, Feld MS. Anal. Chem 2007;79:234–239. [PubMed: 17194145]
17. Hanlon EB, Manoharan R, Koo TW, Shafer KE, Motz JT, Fitzmaurice M, Kramer JR, Itzkan I, Dasari RR, Feld MS. Phys. Med. Biol 2000;45:R1–R59. [PubMed: 10701500]
18. Manoharan R, Shafer K, Perelman L, Wu J, Chen K, Deinum G, Fitzmaurice M, Myles J, Crowe J, Dasari RR, Feld MS. Photochem. Photobio 1998;67:15–22.
19. Cortes C, Vapnik V. Machine Learning 1995;20:273–297.
20. Vapnik, V. The Nature of Statistical Learning Theory. Springer-Verlag; 1995.
21. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang C, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR. Proc. Natl. Acad. Sci 2001;98:15149–15154. [PubMed: 11742071]
22. Widjaja E, Lim GH, An A. Analyst 2008;133:493–498. [PubMed: 18365119]
23. Thissen U, Ustun B, Melssen WJ, Buydens LMC. Anal. Chem 2004;76:3099–3105. [PubMed: 15167788]

24. Scholkopf, B.; Smola, AJ. Learning with Kernels. MIT Press; Cambridge, MA: 2002.
25. Christianini, N.; Shawe-Taylor, J. An introduction to support vector machines. Cambridge University Press; New York: 2000.
26. Tuchin, VV. Tissue optics : light scattering methods and instruments for medical diagnosis. SPIE Press; Bellingham, Wash: 2000. Society of Photo-optical Instrumentation Engineers..
27. Liu R, Chen W, Gu X, Wang RK, Xu K. Journal of Physics D: Applied Physics 2005;38:2675–2681.
28. Qi D, Berger AJ. Appl. Opt 2007;46:1726–1734. [PubMed: 17356615]
29. Chang CC, Lin C-J. LIBSVM: a library for support vector machines. 2001
30. Roe JN, Smoller BR. Crit. Rev. Ther. Drug 1998;15:199–241.
31. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Diabetes Care 1987;10:622–628. [PubMed: 3677983]
32. Wülfert F, Kok WT, Smilde AK. Anal. Chem 1998;70:1761–1767.
33. Despagne F, Massart DL, Chabot P. Anal. Chem 2000;72:1657–1665. [PubMed: 10763266]
34. Wu J, Feld MS, Rava RP. Appl. Opt 1993;32:3585–3595. [PubMed: 20829983]
35. Zhang QG, Muller MG, Wu J, Feld MS. Opt. Lett 2000;25:1451–1453. [PubMed: 18066245]
36. Biswal NC, Gupta S, Ghosh N, Pradhan A. Optics Express 2003;11:3320–3331. [PubMed: 19471461]
37. Shih WC, Bechtel KL, Feld MS. Opt. Exp 2008;16:12726–36.
38. Hansen, PC. Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion. Society for Industrial and Applied Mathematics; 1998.
39. Suykens, JAK.; Van Gestel, T.; De Brabanter, J.; De Moor, B.; Vandewalle, J. Least squares support vector machines. World Scientific; Singapore: 2002.
40. Borin A, Ferrao MF, Mello C, Cordi L, Pataca LCM, Duran N, Poppi RJ. Anal. Bioanal. Chem 2007;387:1105–1112. [PubMed: 17171559]

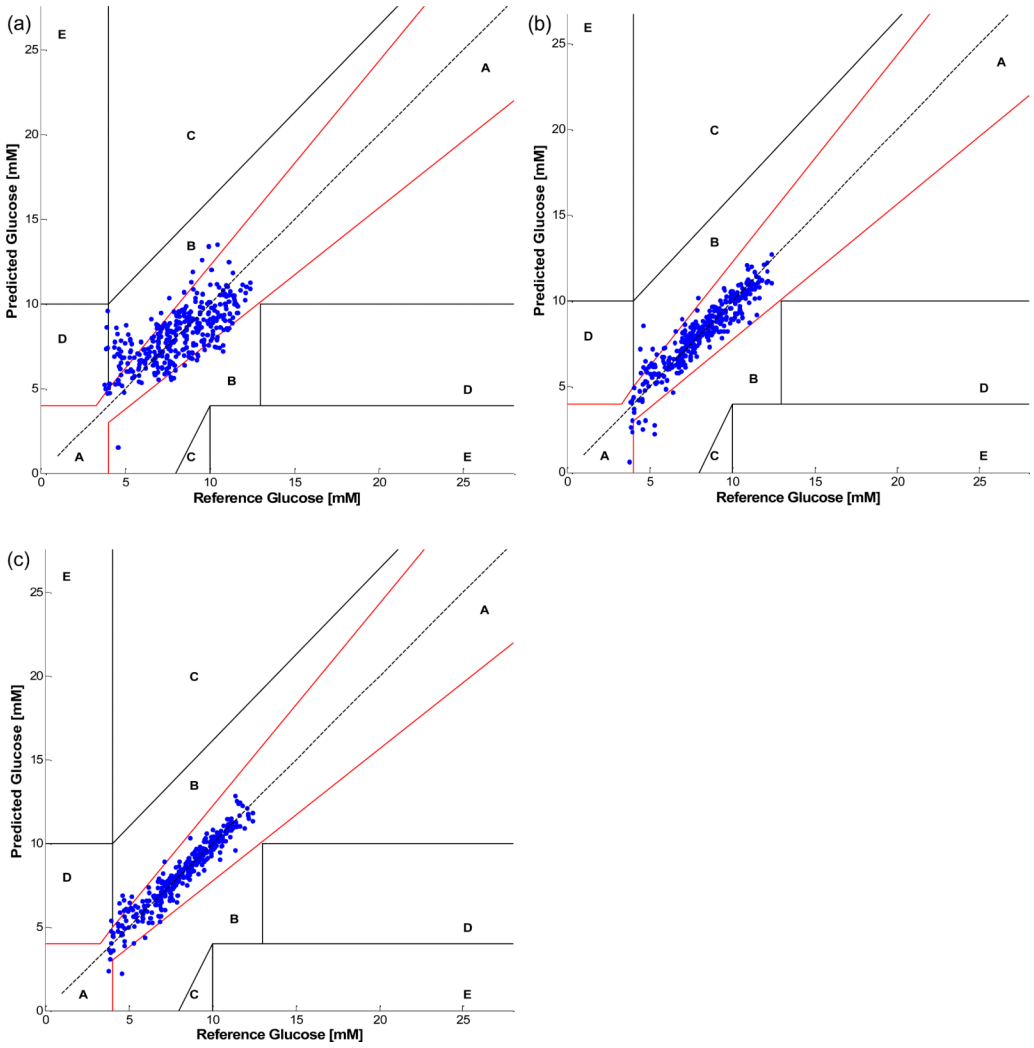

Fig. 1.
Blood glucose predictions of different approaches shown on the Clarke Error Grid: (a) PLS global; (b) PLS local; and (c) SVM global.

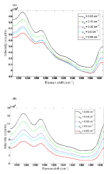

Fig. 2.

Observed Raman spectra from two sets of five tissue phantoms showing the typical spread. Each set has same scattering coefficient but different absorption coefficients. The scattering coefficient of the tissue phantoms in (A) and (B) are 24 and 130 cm^{-1}, respectively. The absorption coefficients are marked in the legend.

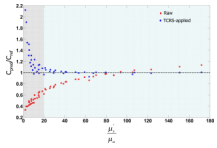

Fig. 3.

Ratio of predicted to actual values of the analyte of interest at constant concentration plotted as a function of $\mu_s^{'}/\mu_a$. The red circles give the OLS prediction values using acquired spectra. The blue circles indicate the prediction values obtained with OLS on TCRS-applied spectra.

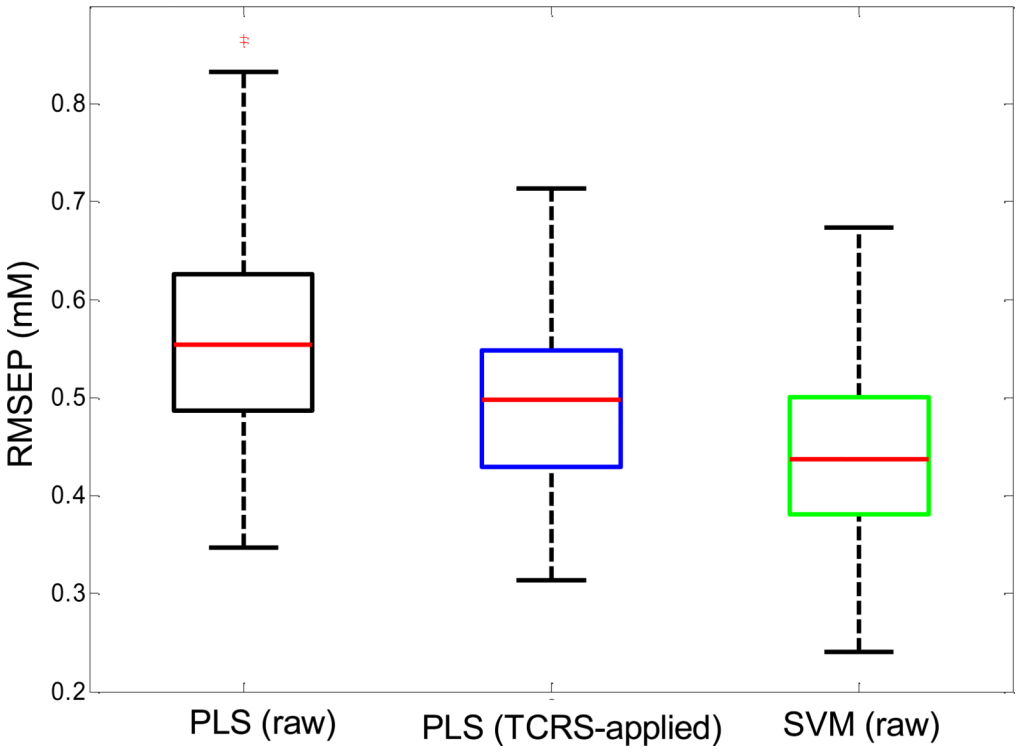

Fig. 4.
Boxplot of RMSEP obtained for glucose concentrations from 100 iterations using PLS on acquired spectra (raw), PLS on TCRS-applied spectra and SVM on raw spectra.