

NON-INVASIVE BLOOD ALCOHOL DETECTION USING NEAR INFRARED SPECTROSCOPY AND CHEMOMETRIC TECHNIQUES**Salvatore Brauer¹**¹Joyson Safety Systems, USAGary Ritchie¹¹Joyson Safety Systems, USALen Cech¹,¹Joyson Safety Systems, USAEmil Ciurczak²²Doramaxx Consulting, USAJohn Coates³³John Coates Consulting, USA**ABSTRACT**

This paper presents an update on the research, development, and manufacturing of a novel passive contact-based Near-Infrared Alcohol Sensor (NIR-AS) for non-invasively measuring Blood Alcohol Concentration (BAC) in human subjects and thus, provides potential for application in support of the new US Infrastructure Investment and Jobs Act bill, section 24220, signed into law on 11/15/2021, once it is enforced.

Alcohol-impaired driving remains a global problem. According to the most recent published report in 2020, U.S. motor vehicle crashes, alcohol-impaired fatalities represent over 30% of the total fatalities; a 14% increase over 2019 and a 29% increase relative to Vehicle Miles Traveled (VMT). The Infrastructure Investment and Jobs Act bill, section 24220, cites statistics on the societal and human costs of alcohol impaired driving and specified intent to make BAC sensors standard equipment in all new U.S. cars in the future. The NIR-AS design and process for analyzing performance in quantifying BAC builds on the R&D carried out in support of the Driver Alcohol Detection System for Safety (DADSS). The published research from DADSS provides valuable technical guidance and performance targets for BAC sensing in motor vehicles. Blood testing is the established gold standard for measuring driver BAC. Although blood testing is the most accurate reference for comparison against NIR-AS (or any new BAC sensor), it is highly invasive, time consuming, and cost prohibitive. Breathalyzers are well established sensors for estimating BAC, however, they also have performance limitations in practical, real-life conditions. Even so, based on published research, including DADSS, breathalyzers can provide an appropriate surrogate reference under controlled clinical and analysis conditions, for analyzing the performance of any new BAC sensor. The NIR-AS sensor described in this paper targets the passive detection performance requirements specified by DADSS.

An alcohol dosing Design of Experiments (DOE) was carried out using a set of Near Infrared Alcohol Sensor (NIR-AS) prototypes with human subjects using a repeat low level alcohol dosing protocol. BAC reference data was also collected using several law enforcement grade and commercial breath analyzers. NIR-AS spectra were processed and analyzed using commercially available and proprietary software.

The DOE resultant data was analyzed using commercially available software packages to produce chemometric models. The paper presents model performance statistics including root mean square standard error of calibration (RMSEC), root mean square standard error of prediction (RMSEP), and square of the correlation coefficient, R^2 , for the NIR-AS calibration. A global model employing multiple sensors was tested across the same DOE and performance statistics are presented. Using NIR-AS, it is shown that BAC can be measured at varying concentrations of alcohol within the human body, including low alcohol dosing levels. Further improvements on the NIR-AS design and function will also be presented.

Based on our results, there is significant correlation between BAC breathalyzer and NIR measurements at low dosing levels. The results demonstrate a high correlation between NIR-AS spectra and reference breathalyzers and achieve low RMSEP, RMSEC, and RMSECV. NIR-AS, with continued development, can be a potential tool for assessing driver alcohol impairment in support of ADAS and/or ADS countermeasures.

INTRODUCTION

The determination of the Percent Blood Alcohol Concentration (%BAC) in human subjects using Near Infrared (NIR) spectroscopy and the multivariate analysis technique known as chemometrics has been established [1]. In 2001, using NIR transreflectance spectroscopic measurement of analytes (e.g., fat) in milk samples, Norris declared that several criteria must be met for an accurate and precise measurement to be made by NIR [2]. Applying these concepts for the measurement of the %BAC in humans:

1. Alcohol must be able to be detected by NIR at a very low levels ($\leq 0.08\%$ or eight one hundredths of one percent, 800 ppm) means that there is approximately 0.08 g of alcohol for every 100 mL of blood. This is the legal definition of alcohol impairment in most states in the United States. Very low levels of an analyte can be detected and measured by NIR diffuse reflection spectroscopy (0.02% - 0.07% BAC) below the 0.08% legal limit.
2. As NIR is not a primary analytical method, but a correlation technique, accurate constituent data (ground truth) is required for developing NIR prediction models of low analyte values.
3. Sampling errors also must be overcome to obtain high accuracy.
4. Sampling errors can be greatly reduced by averaging the spectra from multiple samples of the same constituent level.
5. A narrow bandpass spectrometer is not essential to measure a narrow bandwidth constituent.
6. BAC must be uniquely separable against all other analytes (e.g., specificity).

This paper discusses two experiments (defined further in this paper as *Surrogate* and *Human*) which demonstrate that all six requirements can be met for the detection and the quantification of %BAC. By first analyzing a laboratory surrogate and then using two human subjects (palmar-side finger) and two NIR-AS spectrometers at low alcohol levels using chemometrics and as ground truth, a breathalyzer model Draeger Alcotest 5820 (Houston, TX). The paper includes procedural discussion, the principle components, loading plots, correlation, and regression coefficients from PCA and PLS, and the correlations and regression plots from a recursive chemometrics method called Derivative Quotient Math (DQM).

1. Laboratory surrogate (*Surrogate*)

In the first experiment, it will be shown that the surrogate, consisting of alcohol and water absorbed into a cotton matrix, ranging in concentrations from 0.01% to 0.10 % (~ 100 ppm to 1000 ppm) can be measured as % alcohol directly, using a wide bandpass spectrometer (32 nm) with a Signal to Noise (S/N) $\geq 1400:1$ @ 1700 nm, or 31.46 dB, and sufficiently resolved from water (the largest interfering absorber in human subjects) by the method of Derivative Quotient Math (DQM).

2. Human subject dosing (*Human*)

In the second experiment, an attempt to detect and estimate quantity of alcohol will be carried out using spectra from two different NIR-AS spectrometers (same design used in the *Surrogate* test) and quantified using an evidentiary breathalyzer at low levels 0.02% - 0.07% (200 ppm to 700 ppm) as %BAC in human subjects.

CHEMOMETRICS

Chemometrics was defined by Svante Wold in a 1971 grant application and mentioned again in 1972 [3], as “*The art of extracting chemically relevant information from data produced in chemical experiments is given the name of chemometric*” in analogy with biometrics, econometrics, etc.”

Nearly three decades elapsed between Wold’s definition and when Karl Norris studied NIR spectra with a chemometric application referred to as the Derivative Quotient Math (DQM) and concluded [1] that “*an optimized second derivative ratio makes it possible to obtain a linear correlation to an analyte from spectral data from diffuse transmission and diffuse reflectance measurements.*” This further implied that the DQM pre-treatment optimized the spectral data in such a way that the processed data then fits the Beer–Lambert–Bouguer law relationship [4].

INSPECTING ALCOHOL NIR SPECTRUM AND DERIVATIVES

The spectrum of an absolute alcohol¹ (100% ethanol) transreflectance reference spectrum is shown in figure 1, along with the spectrum second derivative annotated with prominent wavelength absorption bands shown [5].

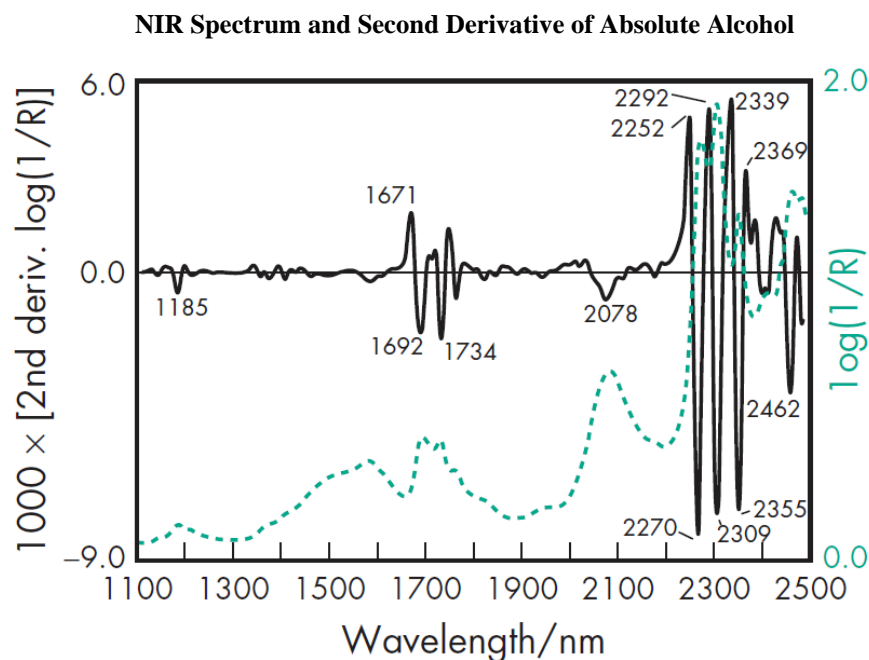


Figure 1. Plots of Absolute Alcohol (a) $\log(1/R)$ vs wavelength, dashed line, scale on right-hand side) and (b) second-derivative $\log(1/R)$ vs wavelength (solid line, scale on left-hand side). Near-infrared spectrum of alcohol (absolute) measured by transreflectance (1mm path-length).

¹ From the Handbook of Pharmaceutical Excipients Sixth Edition. Near-infrared (NIR) spectra of liquid samples were measured using a FOSS NIRSystems 6500 spectrophotometer (FOSS NIRSystems Inc., Laurel, MD, USA). Liquid samples were measured by transreflectance using a gold reflector (2 x 0.5mm optical path-length, FOSS) placed in a 45mm silica reflectance cell against air as the reference. Spectra are presented as plots of (a) $\log(1/R)$ vs wavelength (dashed line, scale on right-hand side) and (b) second-derivative $\log(1/R)$ vs wavelength (solid line, scale on left-hand side). R is the reflectance and $\log(1/R)$ represents the apparent absorbance. Second-derivative spectra were calculated from the $\log(1/R)$ values using an 11-point Savitzky-Golay filter with second-order polynomial smoothing.

The alcohol second derivative reference displays thirteen prominent bands in the region between 1100 nm and 2500 nm, where all these features are represented with a sensor resolution of 2 nm. This region is the C-H_n first overtone. The molecular formula for alcohol (ethanol) is CH₃CH₂OH. Therefore, the C-H_n first overtone corresponds to the CH₃ and CH₂ (methyl and methylene) group of the alcohol molecule [6].

Table 1 compares the reference standard bands with the five bands of the alcohol second derivative bands measured on the wide-band spectrometer shown in figure 2 (11-point Savitzky-Golay filter with second-order polynomial smoothing). The NIR-AS prototype devices, discussed in the rest of this paper, have a wavelength range of 1350 nm to 2550 nm, with a wavelength resolution of 32 nm.

The alcohol spectrum, 1st ($d^1A/d\lambda^1$) and 2nd derivative ($d^2A/d\lambda^2$) are shown in figure 2 from the wideband spectrometer. Second-derivative spectra were calculated using an 11-point Savitzky-Golay filter using a second-order polynomial smoothing, indicating the derivative of the optical values (usually the absorbance) with respect to wavelength ($d^nA/d\lambda^n$). The spectra taken with the spectrometer were converted from Wavenumber (cm^{-1}) to Wavelength (nm) and Transmission to Absorbance ($\log 1/T$).

Of the requirements previously mentioned, the one having the greatest significance to this research is number six: BAC must be uniquely separable against all other analytes (e.g., specificity). Water is the largest absorber interfering with BAC determination, and useful to investigate several chemometric approaches previously mentioned in other analogous applications. The authors are not aware of reports which investigate the use of the Derivative Quotient Math (DQM) for the determination of Blood Alcohol concentration (BAC) in humans. The algorithm possesses unique capabilities especially suited for measuring BAC in humans.

NIR SPECTROSCOPY SCATTER CORRECTION

The wavelength-dependent redirection of light scattered through a complex media can be defined simply as scatter in the NIR application discussed in this paper.

Davies [7], in presenting an explanation of the origin and use of derivatives in spectroscopy, concluded that, while derivatives are useful for removing extraneous signals from NIR spectra, the resulting spectra still contain multiplicative effects of scatter.

Derivative Quotient Math (DQM) Explanation

Karl Norris, “The Father of NIR,” explained the DQM mechanism in “Norris on Norris Regression,” described in reference 0 as follows:

“First, there are two distinct items involved. The first is the gap derivative (sometimes called the Norris derivative by mistake), the second is the “Norris Regression”, which may or may not use derivatives.

The “Norris Regression” is a regression procedure to remove the effects of varying pathlengths among samples because of scatter effects. This is accomplished by incorporating a divisor into the regression term. The divisor can be the absorbance at another wavelength, a difference between the absorbance at two wavelengths, a first derivative, or a second derivative. The single wavelength divisor does not work well in many cases because that signal contains offset variations as well as multiplier variations, and we only wish to sense the multiplier signal [Multiplicative Scatter].”

More recently, with the publication of the Fourth Edition of the Handbook of Near-Infrared Analysis [9], Hopkins offers further insight as to why applying derivative pre-treatments alone are insufficient for removing multiplicative effects from light scattering: He asks, “... why use Derivative Quotients? Simply stated, the ratios effectively cancel

the multiplicative effects caused by differences in scattering between samples. The ratios will also cancel differences between instruments that are due to differences in spectral bandwidth. In addition, it has been observed that DQM models generally require only 1 or 2 terms, possibly 3 terms. This may make such calibrations very robust...

When considering the use of derivatives to correct scattering, spectral pre-treatments based on derivatives remove most spectral baseline offsets. However, simple derivative-based pre-treatments cannot remove multiplicative effects. The second derivative-based pre-treatments can largely remove linearly (or nearly linearly) sloping baselines.

Again, considering Figure 2: the Alcohol reference NIR spectrum (Bottom), first derivative (Middle) and second derivative (Top) of alcohol measured by diffuse reflection on the wide bandpass spectrometer (NIR-AS-212), the highlighted lines represent peak find solutions using the chemometrics software tool Solo (Eigenvector Research Inc., Manson, WA).

The alcohol wavelengths and their positions are annotated textually. The wavelengths for the raw spectrum are positive, zero for the first derivative (points at zero), and negative for the second derivative.

A flow diagram outlining the DQM algorithm is available in reference 8. A MATLAB script, dqm1, was applied on the *Surrogate* and *Human* data sets. The program executes searches of gaps (segments) or smoothing point intervals (smt) that are tested over intervals individually selectable for the numerator and denominator.

The optimal gap segment size for the wavelength range, using the benchmark devices in the wavelength range between 1350 nm - 2550 nm, can be estimated on (or near) the band centered at 2295 nm (Figure 3). The number of points in the half-peak width of this segment is equal to six (6). The peak height has 12 points from peak to base line. Using these values for the gap and smoothing search, the DQM will not exceed 6 data points for the gap segment search, or 12 for the smoothing interval (see Figure 3).

Further from Hopkins: *“Derivatives can correct the offset and slope differences that are found in sets of spectra of diffusely scattering samples. However, they cannot remove the multiplicative effects. It was observed by Karl Norris that the ratio of derivatives terms can remove the multiplicative effects. He used simple multiple regression to find the terms of quotients employing optimal wavelengths for the numerators and denominators.”*

Hopkins explains [10], that choosing an optimal gap size can be approximated by *“selection of a convolution interval about the same size as the number of points in the half-band width of the sharpest band in the wavelength range in which you are working.”*

NIR spectrum, (Bottom) First Derivative (middle) and Second Derivative (Top) absolute alcohol from NIR-AS (unit #212) spectrometer.



Figure 2. Shows the alcohol NIR spectrum, (Bottom) First Derivative (middle) and Second Derivative (Top) of from NIR-AS (unit #212) spectrometer.

Table 1.

Compares the reference standard bands to the five bands of the alcohol second derivative bands measured on the wide band spectrometer shown in figure 2.

Alcohol NIR Second Derivative of Reference Wavelengths (nm)	Alcohol NIR Second Derivative of NIR-AS-212 Spectrometer Wavelengths (nm)
1185	
1671	1579
1692	1700
1734	
2078	2074
2252	
2270	
2292	2295
2309	
2339	
2355	
2369	
2462	2472

INSPECTING THE ALCOHOL NIR SPECTRA USING THE DQM APPLICATION

Having established the known alcohol NIR absorption bands from the wideband spectrometer above, the DQM program was applied to the *Surrogate* spectra in order to find the terms of derivative quotients employing optimal wavelengths for the numerators and denominators.

Experiment 1: Laboratory *Surrogate* (Low concentrations (0.01% - 0.1%) of alcohol and water)

Material and Equipment

The following chemicals were used to conduct the NIR measurements: pure lab grade Alcohol (PHARMCO-AAPER, Brookfield, CT) and distilled water. A micropipette, (Wilmed LabGlass, MED Plus, Vineland, NJ) was used to perform the serial dilutions. 100% cotton medical-grade gauze pads were used as the sampling matrix for these tests.

Procedure

Stock Standard

Serial dilutions were from 0.10% alcohol by volume down to 0.01% in steps of 0.01%. A 0.10% solution stock solution was prepared (500uL alcohol into 500mL water) with a micropipette. Serial dilutions were made with a 10 – 100uL micropipette.

NIR Measurement

To present the samples to the detector window, small uniform cotton squares were prepared and fit into a small holder on the NIR-AS sensor. A spectral sample of the cotton was measured as the reflectance background correction. After the dilutions were made, 300ul of solution was pipetted onto the cotton swab and then placed into the holder. Alcohol Spectra were collected (a total of 10) using 10-second dwell times. This process was repeated for each subsequent concentration.

Chemometric Analysis

The DQM parameters selected for searching the wavelengths from 2000 nm - 2550 nm (alcohol range) (at 32 nm, there are 80 wavelengths) used were:

1. # of Calibration Samples: N22
2. # of Validation Samples: N21
3. Number of terms (1 or 2): 2
4. Derivative order (d): Term #1: 1D, Term #2: 2D
5. Number of differential gaps (gap): 3
6. Number of smoothing points (smt): 6

Recall that, from the inspection of the absolute alcohol NIR spectra, it was determined that the optimum gap and smoothing for peaks in a spectrum is predicated on the half-width of a peak. For alcohol, it was determined that, based on the second derivative spectrum, the most intense absorption band is found at 2295 nm. Based on this fact, the number of data points in the spectrum (80), and studies on the NIR absorption of alcohol and water [11], the selection of DQM parameters were optimized.

With respect to alcohol detection, our research into the optimum range for the detection of alcohol in the presence of water and other absorbers that can interfere with the detection, showed that for the NIR-AS prototype the selection of the 2000 nm – 2550 nm (CH - CH combination) portion of the full range was required in order to identify principle components, latent variables, and to determine the correlation of breathalyzer to the NIR-AS spectra to generate the regression coefficient (PLS model) of the surrogate calibration data set specific for the determination of alcohol.

Second Derivative of the absolute Alcohol NIR Spectra (11, 2, 2D) @ 2295 nm

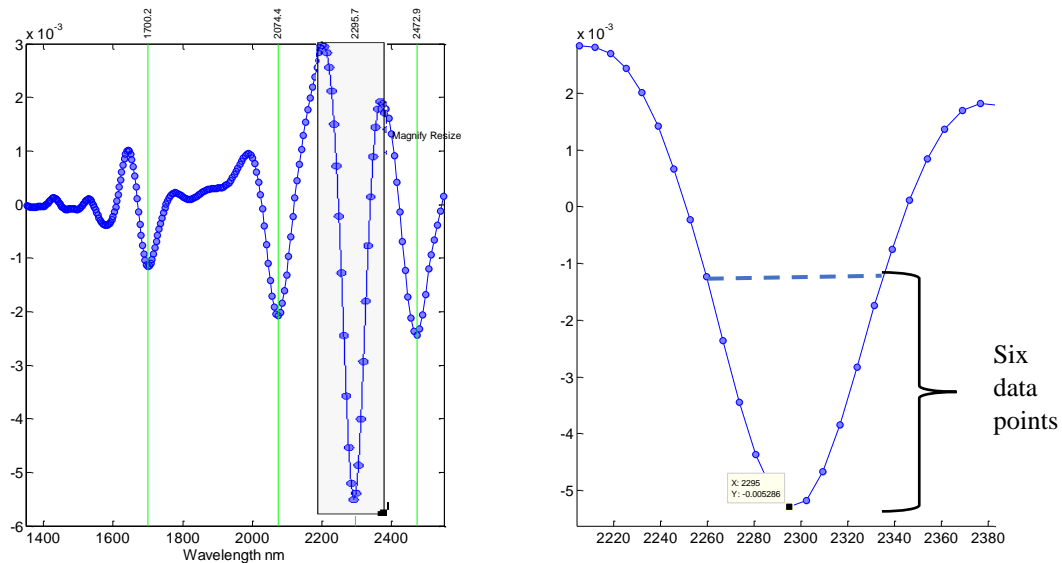


Figure 3. The optimum gap and smoothing for peaks in a spectrum is predicated on the half-width of a peak of interest. Alcohol peak at 2295 nm shows expanded view in order to count data points for band.

The *Human* experiment analysis will show that the range from 2000 – 2550 nm, **uses these same wavelengths for the selection of alcohol**. The final results following implementing the dqm1 program *Surrogate* experiment 1 (surrogate) were:

%Alcohol Regression Results

SEC = 0.008 RSQ = 0.9375 N = 22 RMSEP
 = 0.009 SEP = 0.009 bias = 0.000 RSQ = 0.9215 N = 21

2-Term Model: WAEXP2 N22 CAL 4.modl, Ders normalized

Term #1: (2D 2273.6842 NM, gap=1, smt =0)/(2D 2198.8024 NM, gap=3, smt =6)

Term #2: (2D 2392.1824 NM, gap=3, smt =6)/(2D 2028.7293 NM, gap=1, smt =5)

Coefficients B0, B1, B2, .. = -0.61818 0.54646 -0.017699

Figure 4 shows the surrogate NIR spectra and Extended Multiplicative Scatter Correction (EMSC) spectra used in experiment 1. Note that the spectral concentrations (0.01%, 0.03%, 0.05%, 0.07%, 0.09%) do not increase linearly from low to high alcohol concentration with wavelength, but changes due to scatter, instrument, sampling, and other effects previously mentioned.

The spectra are corrected for multiplicative scatter using Extended Multiplicative Scatter Correction (EMSC) [12] prior to analysis following a DQM analysis of unprocessed spectra, giving poor ratio quotient results for the % alcohol at five concentration levels.

Table 2 lists the DQM results for the 2-term and 1-term models for the determination of surrogate concentrations by NIR.

Figure 5 shows the optimal derivatives selected for alcohol by the DQM algorithm. The DQM Matlab application dqm1 (MathWorks, Inc., Natick, MA) was used to find the terms, quotients employing optimal wavelengths for the numerators and denominators terms. Figure 6 shows the calibration from the optimal wavelengths using DQM. The output for the Standard Error Calibration (SEC) and Regression Coefficient plots for the DQM results of the surrogate are shown in Figure 7.

%Alcohol NIR Spectra and Extended Multiplicative Scatter Correction (EMSC) (CAL N = 22)

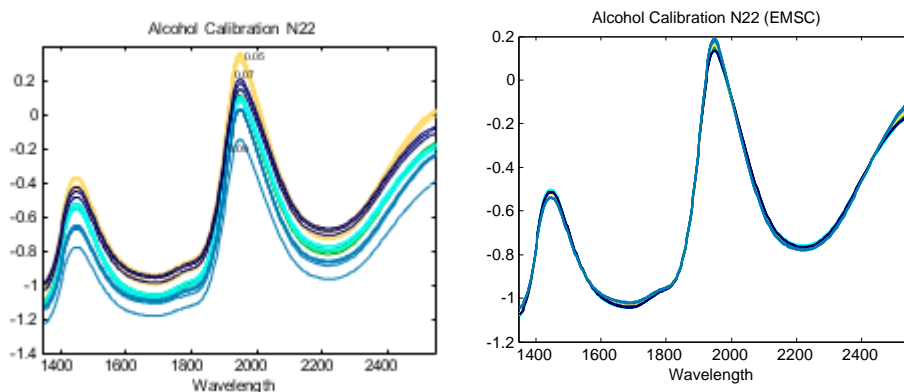


Figure 4. Shows the surrogate NIR spectra and Extended Multiplicative Scatter Correction (EMSC) spectra used in experiment 1.

Optimal Derivatives Selected For %Alcohol (Spectra (N22) By the DQM Algorithm (2-Term Model 2D/2D)

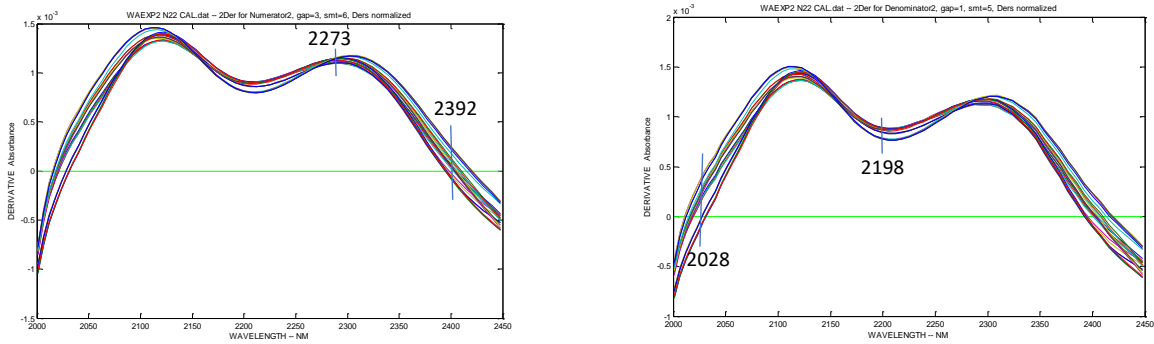


Figure 5. Optimal (2-term model 2d/2d) derivatives selected for %alcohol surrogate (N22) by the DQM algorithm. Note the prominence of regions of the spectra where wavelengths have been selected correspond to regions of no scatter: the cross-over regions of the spectra. Other wavelengths center on or near a region of the alcohol analyte where known NIR absorption is expected to occur.

Surrogate Concentration Scatter Plot CAL (N22) And VAL (N21)

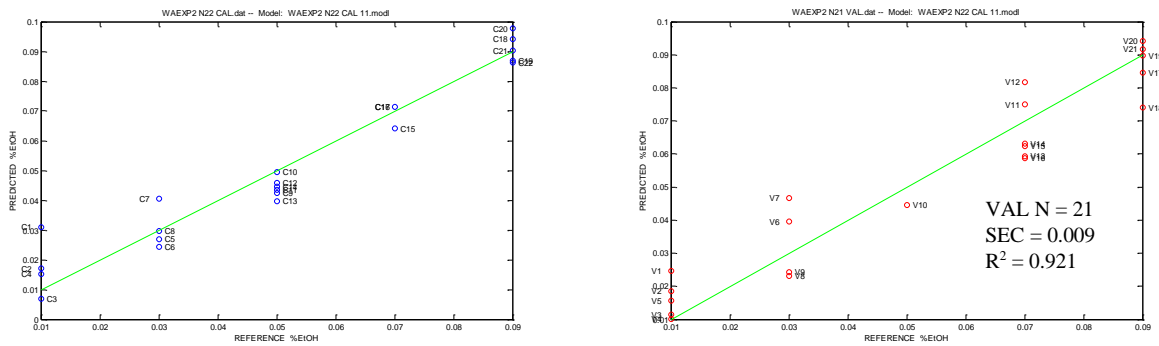


Figure 6. shows the surrogate concentration scatter plot from the DQM calibration and validation as a result of the gap, smoothing (smt) derivative quotient wavelength search.

TABLE 2.

DQM Results for the 2-Term and 1-Term Models for the Determination of Surrogate Alcohol Concentrations by NIR

Wavelengths	2-Terms	GAP								
Derivative Ratio	SEC (%BAC)	N1	smt	D1	smt	N2	smt	D2	smt	R²
1D / 1D Term 1 2384 / 2239 Term 2 2062 / 2062	0.012	2	0	3	5	3	0	2	0	0.846
1D / 2D Term 1 2062 / 2309 Term 2 2225 / 2392	0.011	3	6	1	1	1	0	1	1	0.876
2D / 1D Term 1 2309 / 2057 Term 2 2338 / 2017	0.010	1	1	3	6	1	5	3	6	0.883
2D / 2D Term 1 2273 / 2198 Term 2 2392 / 2028	0.008	1	0	3	6	3	6	1	5	0.938
	1-Term	GAP								
	SEC (%BAC)	N1	smt	D1	smt					R²
1D / 1D Term 1 2384 / 2239	0.013	2	0	3	5					0.798
1D / 2D Term 1 2062 / 2309	0.013	3	6	1	1					0.796
2D / 1D Term 1 2309 / 2957	0.013	1	1	3	6					0.816
2D / 2D Term 1 2273 / 2198	0.015	1	0	3	6					0.738

Standard Error Calibration (SEC) and Regression Coefficient Plots for the DQM results of the surrogate

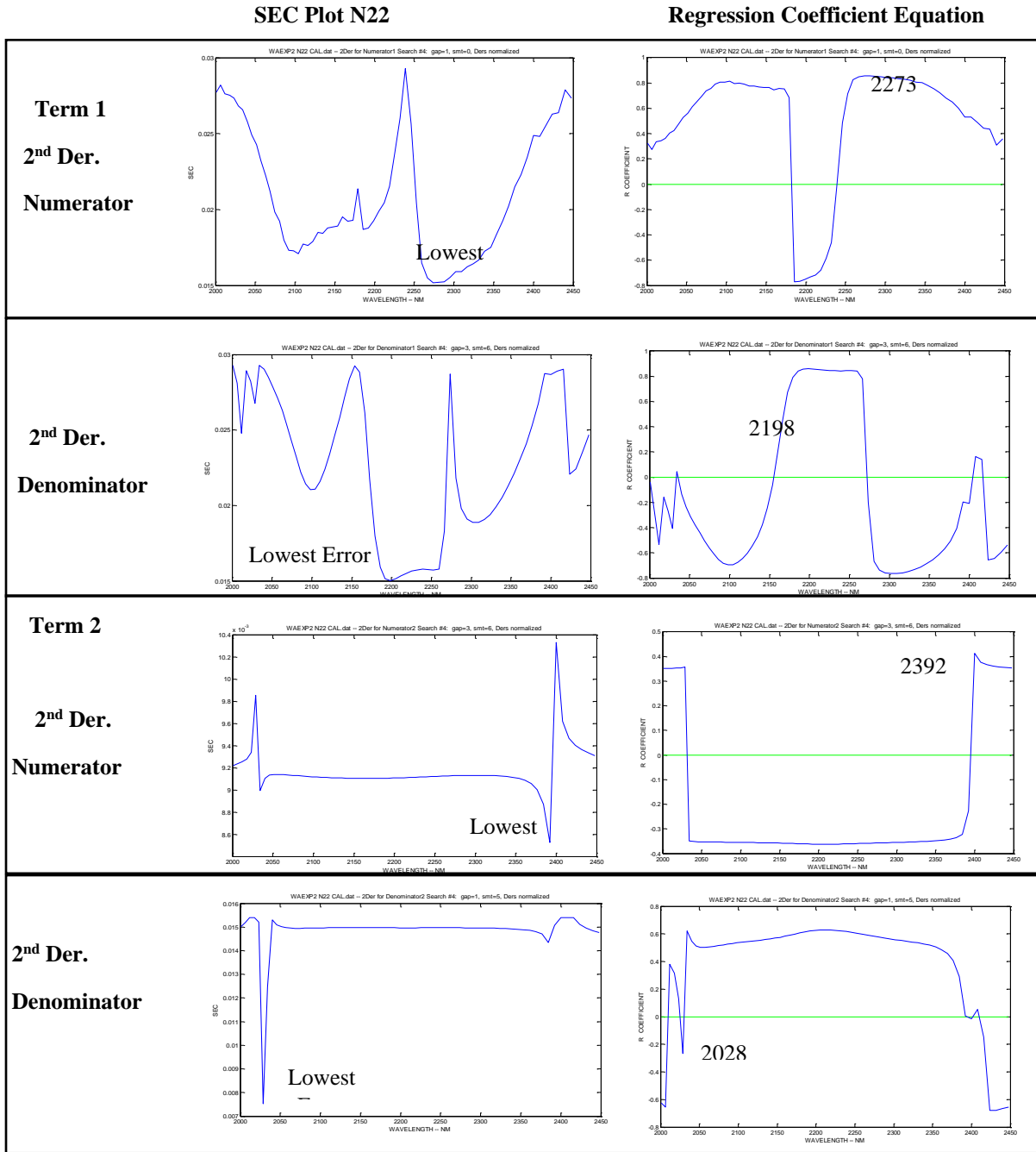


Figure 7. Standard Error Calibration (SEC) and Regression Coefficient Plots for the DQM results of the surrogate mixtures

The best model is found with the lowest standard error of the calibration (SEC) and a coefficient of determination (R^2) approaching 1.0. The SEC was 0.008 and R^2 was 0.938 for sample $N = 22$ for concentrations in the range 0.01%, 0.03%, 0.05%, 0.07% and 0.09% (Figure 6). The wavelengths selected to be specific for the alcohol should be on or close to those wavelengths specific for alcohol as shown in figure 2. For this analysis, those bands are the alcohol bands at 2273 and 2028 nm.

Based on the results of *Surrogate*, we will show that the results for the *Human* experiment over the calibration range from 2000 – 2550 nm is the only range showing specificity for alcohol using the NIR-AS

IDENTIFICATION OF ALCOHOL BY PRINCIPLE COMPONENT ANALYSIS

Since the DQM provides excellent correlation of the *Surrogate* data to the NIR spectra (low error, high coefficient of determination), it may be reasonable to assume that this may be also true for the *Human* data. However, prior to testing this hypothesis, a very useful tool to probe for alcohol following a NIR measurement and correlation with the corresponding breathalyzer values is to review the Principle Components (PCs), and residual analysis.

Principal Component Analysis (PCA)

The basic procedure fundamental to most chemometric analysis is the technique known as Principal Component Analysis (PCA). As reviewed by Wold *et al.* [13], and first explained by Pearson [14], the problem at hand is applicable for the use of PCA in determining the differences between groups of spectra. Wold *et al.* provide a general approach for extracting the dominant patterns from data matrices. They are:

1. Formulate the problem statement by asking why the data matrix (in this study, spectra and alcohol doses) were collected in the first place.
2. What is the purpose for the experiments and measurements?
3. Specify, before the analysis, what kinds of patterns are expected to be found.
4. The decomposition of the independent variables (X-wavelengths) and the dependent variables (Y-absorptions) matrices comprising the spectra results in a set of loading and scores describing the variance. (See section on Loadings and Scores).
5. In examining the resulting scores plot, look for outliers (spectra that do not fit any observed pattern), but do not remove outliers without understanding their underlying cause.
6. Use the resulting Principal Components to guide continued investigation or chemical experimentation, not as a result.

For the NIR-AS system under study the general approach is answered by the following statements.

Problem Statement: Development of a passive touch-based Near-Infrared (NIR) sensor for non-invasively measuring the Blood Alcohol Concentration (BAC) in the driver of a vehicle.

Purpose for the experiments and measurements: A NIR touch-sensor potentially offers non-invasive, non-destructive and rapid measurement times. This novel sensor is intended to meet the passive detection requirements of the Driver Alcohol Detection System for Safety (DADDS), improve driver safety by providing a non-intrusive means of notifying a driver or applying some other countermeasure when their estimated %BAC may exceed established threshold(s). When BAC values are modeled, it is intended that the model follows Beer–Lambert–Bouguer Law which describes a linear relationship between the spectral absorbance and the concentration, molar absorption coefficient and optical coefficient of a solution.

However, measuring BAC in human fingers is non-trivial. The finger is a highly scattering, chemically complex matrix, composed of tissues varying in thickness and densities, contributing to scatter. The purpose, therefore, is to find, through well planned and designed experiments, conditions that model this nonlinear phenomenon by reducing scatter, identify and minimize interfering absorbers (e.g., hemoglobin), and linearly correlate only the highly scattered spectra to the % BAC values obtained by suitable reference (in this study, an evidentiary breathalyzer, model Draeger Alcotest 5820 (Houston, TX). In future studies, there is a plan to use Headspace Gas Chromatography with Flame Ionization Detector (HS-GC-FID) on drawn blood taken simultaneously with the NIR-AS and breathalyzer measurements for validation of the NIR-AS touch sensor.

What kinds of patterns are expected to be found: Alcohol dosing curves (Y-matrix) obtained from breathalyzer measurements, when linearly correlated with the NIR-AS spectra (X-matrix) that are obtained simultaneously, result in correlation, score, and loading plots are a function of the relationship of the X – Y matrices. When corrected for variances from the NIR-AS spectra and breathalyzer values, using chemometric preprocessing techniques, the result is a linear regression curve from the calculated NIR-AS wavelength regression coefficients that can be useful for predicting %BAC.

Describe the expected results from the scores and loadings with respect to the observed variance: NIR %BAC results, when linearly modeled by correlating alcohol concentration absorbances and NIR spectral wavelengths, can be used to analyze future unknown NIR-AS spectra obtained through the finger for %BAC.

Describe any observed outliers and explain their root cause: NIR-AS Spectra and %BAC results not within the models 95% confidence limits must be investigated and root cause determined.

Propose future use of the resulting principal components to guide the ADS program: Unknown NIR-AS spectra may be compared with the model PCA scores to classify the spectra as either "free of alcohol" or "containing alcohol" and subsequently analyzed by regression analysis to estimate how much alcohol is or is not present.

Loading and scores: Plots of spectra, scores, loadings, and residuals may be used to extract the relevant information pertaining to the alcohol analyte and provide scientific evidence for the presence or absence of the compound in blood. As such, Nørgaard *et al.* [15] provide a methodology for achieving the objective as stated by Pearson, "to represent a system of points in plane, three, or higher dimensioned space by the "best-fitting" straight line or plane." Principal Component Analysis (PCA) can be used to "estimate the latent spectra (loadings) and determine the corresponding concentrations in the samples (scores) from the measured spectra."

The following analysis will be used to demonstrate if NIR spectroscopy can be used to detect the presence or absence of alcohol at low concentrations. Utilizing PCA of the spectra, Partial Least Squares (PLS) followed by DQM will be explored to analyze the spectra for the Percent Blood Alcohol Concentration (%BAC).

Experiment 2: Human (Methodology Described by Nørgaard *et al.*)

The data set used for accessing the utility of NIR spectroscopy for the analysis of %BAC is composed of two subjects on two benchmark devices (NIR-AS-103 and NIR-AS-215) and are presented in Table 3.

The %BAC ranges from 0.017% to 0.075%, from $N_{\text{Spectra}} = 1,171$, averaged (Coadd) = 586, and then split into calibration ($N = 291$) and validation ($N = 295$) samples. Outlier removal, using Robust PLS, was performed on the

calibration, while manual removal was used on the validation set. Wavelength ranges from 1350 nm - 2550 nm (257 variables), 32 nm bandpass. The Signal to Noise (S/N) was determined to be ($\geq 1400 : 1 @ 1700 \text{ nm}$) or 31.46 dB.

TABLE 3.

Design of Experiment for Two Human Subjects Measured on Two Benchmark Devices (N = 1,171)

Device	Subject 1 NIR-AS-103	Subject 2 NIR-AS-215
	Calibration (N = 683)	Calibration (N = 488)
%BAC	0.031% – 0.075%	0.017% – 0.074%
Coadd Data (Before Sorting)	Calibration (N = 586)	
N (X-Matrix) (Outliers removed)	(N = 206)	
%BAC	0.017% - 0.074%)	
N (Y-Matrix)	(N = 206)	
Wavelengths 1350 nm - 2550 nm	257	
Split from Coadd Data	Validation (N = 291)	
N (X-Matrix) (Outliers removed)	(N = 127)	
%BAC	0.019% - 0.075%	
N (Y-Matrix)	(N = 127)	
Wavelengths 1350 nm - 2550 nm	257	

Figure 8 shows plots of raw and Extended Multiplicative Scatter Correction (EMSC) from the *Human* data calibration spectra, measured against a 99% Spectralon® reflection standard, following outlier removal. Spectra are colored by the %BAC levels. The alcohol dosing range (after sorting on ascending values) is shown for each subject in Figure 9.

As seen in figure 8, the BAC observed in the raw and EMSC treated spectra are not varying with alcohol absorption as one goes from the lowest to the highest value BAC, but from the scatter created from the tissue comprising the finger, the measurement placement and other non-absorption sources.

As mentioned in the introduction to this section that, the loadings and the sample residual plots may be analyzed using the approach recommended by Nørgaard *et al.* for “estimating the latent spectra (loadings) and determines the corresponding concentrations in the samples (scores) from the measured spectra.”

Plot of Raw and Extended Multiplicative Scattered Correction Spectra From Humans Following Alcohol Consumption Calibration (N206) Validation (N127)

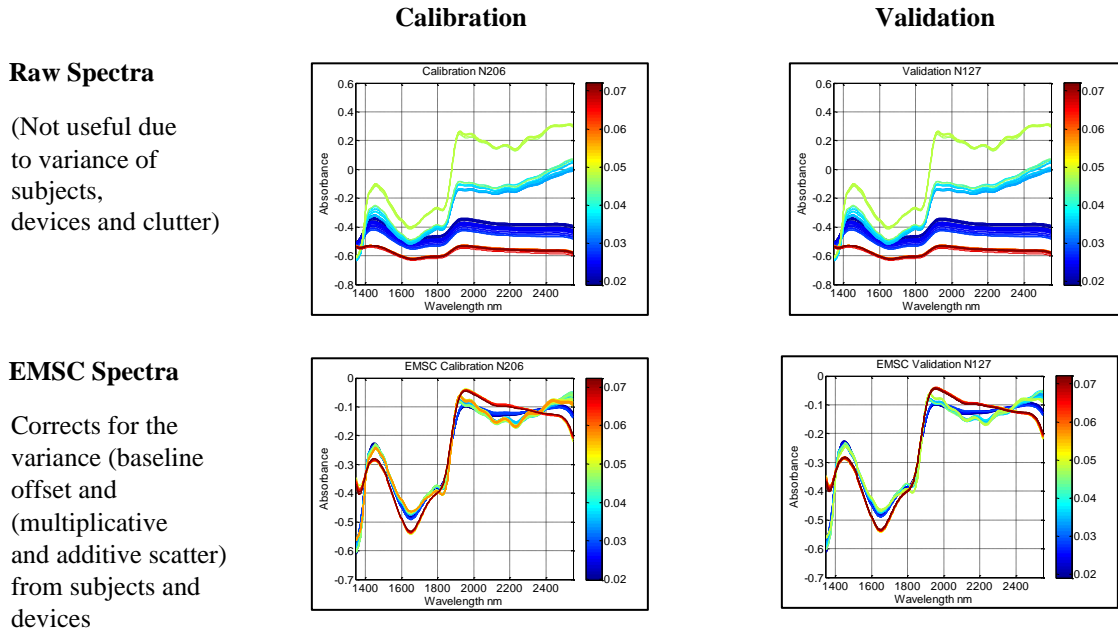


Figure 8. Plot of Raw and Extended Multiplicative Scattered Correction Spectra (Human)

Alcohol Dosing Plots Extended Multiplicative Scattered Correction Spectra After Sorting (Human)

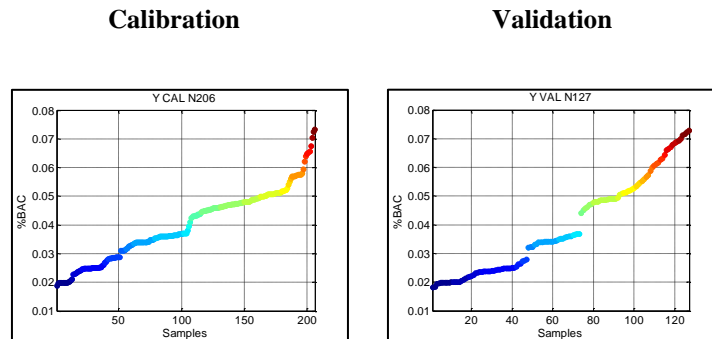


Figure 9. Sorted alcohol breathalyzer value plots for Calibration (N206) and Validation (N127)).

Figure 10 shows the PCs of the BAC values from human subject data of three selected samples, a low, middle, and high %BAC. Note that the PCA model is calculated on the calibration samples (N206) X-variables only (wavelengths); the Y-matrix breathalyzer values are not used. To the left, in column one, the raw spectra are shown for calibration sample 0.030 %BAC (#50), sample 0.050 %BAC (#149), and sample 0.070 %BAC (#241). Column two shows the mean spectrum over all calibration sample. The mean spectrum is identical for all samples.

The first loading vector is the spectral structure that is best at describing the variation in the EMSC data (Figure 8). No other structure can explain more of the variation in the data than the first loading vector. The first loading is common to all samples; what makes the samples different is the content or concentration of this structure in their spectrum. This concentration is called the score value. The score value for 0.030 %BAC is 22.6, for 0.050 %BAC is 47.5, and 0.070% is 10.0. The 288 remaining samples in the data set have different score values. Multiplying the

loading vector with the score values for samples #50, #149 and #241 are the best descriptions one can obtain for these samples, when the loading vector should also describe the other samples. Other observations are the following:

- The shape of the first loading vector is the inverse of the EMSC treated spectra. This explains the greatest variance of the spectra (N206), and is describing all non-absorbing phenomena, (i.e. physical offset of the spectra related to finger placement, differences in individuals, and different NIR device).
- The shape of the second loading vector resembles the human subject finger EMSC spectra. Prominent features of this spectrum are seen around the 1450 nm and 1940 nm wavelengths. These are known wavelengths for the NIR absorption of water.
- The residual spectrum shows, as one goes down the table with increasing alcohol concentration, the wavelengths between 1350 nm – 2000 nm increase at those wavelengths associated with water.
- The residual spectrum also increase between 2000 nm – 2550 nm as one goes from 0.05% to 0.1% alcohol.

Loading Plots of %BAC Alcohol Consumption (Range = 0.030% – 0.070%)

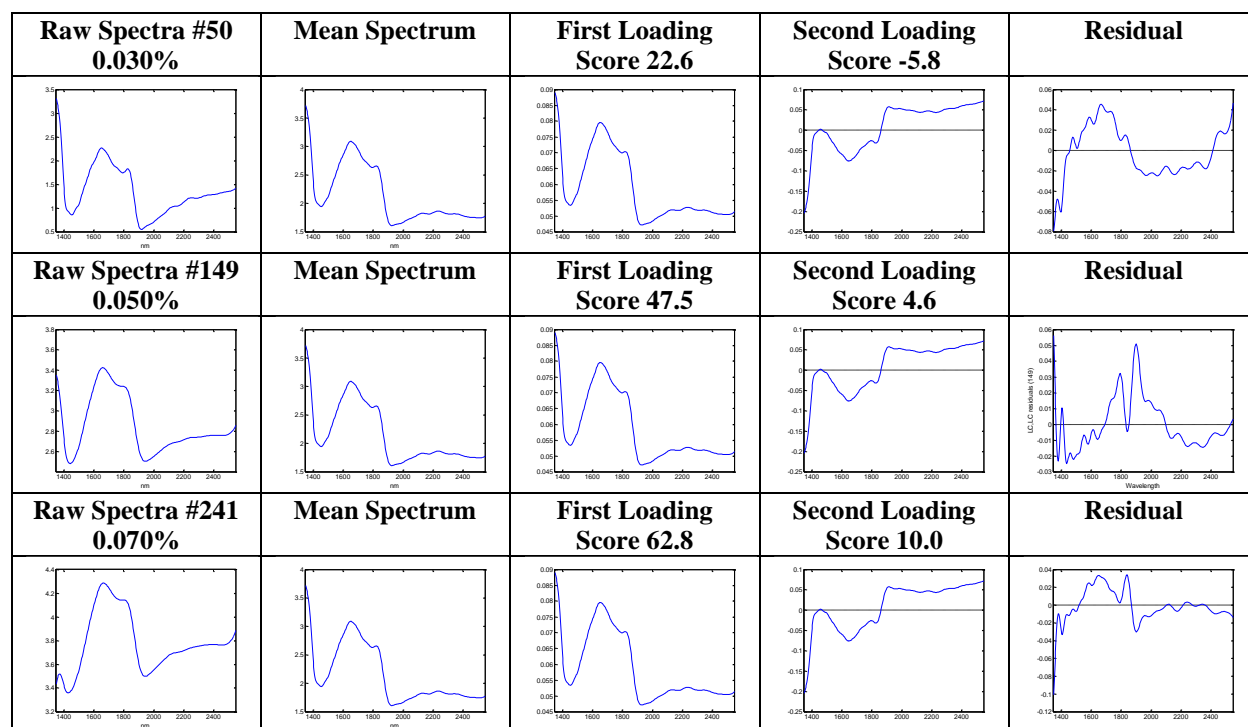


Figure 10. PCA of %BAC Alcohol Consumption (Range = 0.030% – 0.070%)

The second loading is the structure that describes the second most variation in the data set. The vector has the special property of being orthogonal (perpendicular) to the first loading. Once again, the sample diversity is reflected in the loading score value.

The part of the variation in the data set not described by the first two loading vectors is represented by the residuals (Figure 10, column five). The residuals are specific for each sample and can be used for the detection of deviating sample patterns. Note that, as one moves down the residual column, the features of the residual describing all other samples other than the sample selected look less like the loadings as they increase from 0.030% to 0.070%. By comparing the size of the residuals with the variation of the EMSC data, one can calculate the variance explained for each Principal Component.

An important observation in the residual plots as the concentration of alcohol is increasing moving down the residual column, the intensity of the peaks other than those of alcohol in the 1350 nm – 2000 nm region begin dominating the whole absorption spectrum. Prior to testing the DQM on experiment #1, the PLS calibration was tried on the set using the full wavelength range. The resulting derivative plots show the wavelength range being calibrated on is from 1350 nm – 2550 nm, and clearly shows the effect of water in this region around 1450 nm and 1940 nm (see figure 11).

So, while PCA is useful, in this case, finding loading vectors describing the absorption spectrum of water in the finger, there is no clear indication of where alcohol is in the spectrum using the whole wavelength range 1350 nm – 2550 nm.

Surrogate experiment, PLS calibration on %Alcohol NIR Extended Multiplicative Scatter Correction (EMSC) (N = 21) Set using the full wavelength range (1350 nm – 2550 nm)

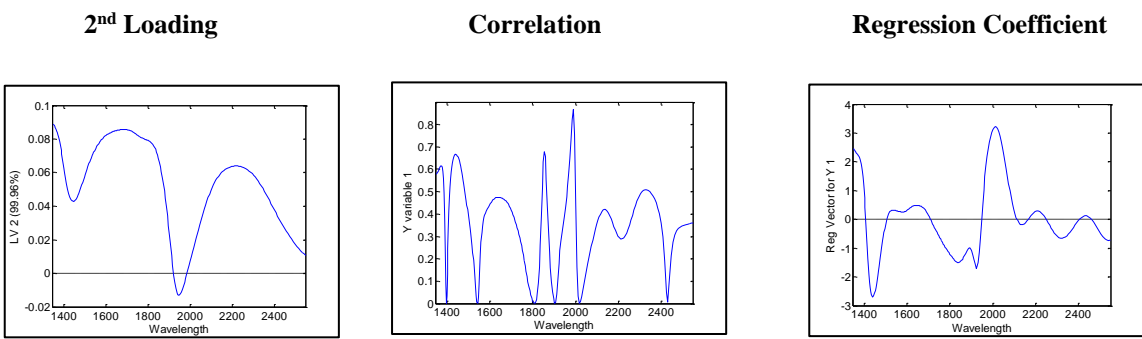


Figure 11. Surrogate Experiment, calibration full range 1350 nm – 2550 nm, shows 2nd loading explaining 99.96% model variation due to water, correlation plot showing water and the regression coefficient showing variables contributing most to the model. Note the large negative coefficient at ~1450 nm and large positive coefficient around ~1940 nm, associated to water.

As can be seen in figure 12, water absorption dominates the correlation and regression even though alcohol and other absorbers from hemoglobin, fat, and other analytes are present in the human. The relationship to these wavelengths is a function of the surrogate concentrations at 0.01%, 0.05% and 0.10%. Hence, the correlation and regression from the wavelengths associated with water strongly exhibit absorption effects (1350 nm – 2550 nm) greater than that of alcohol across the full spectrum.

Can the same patterns from the loading, correlation, and regression coefficient be observed in the Human experiment (BAC% estimation for two subjects and two devices) presented in table 3? Figure 12 shows that the same patterns seen from the surrogate data across the full range spectrum can also be seen in the %BAC human subject plots leading to the conclusion that the spectra from the **Human** data set behaves in the same manner as the spectra from the **Surrogate** model.

That is, the full wavelength range 1350 nm – 2550 nm is detecting and calibrating on water and other absorbing components of the fingers, not alcohol. The alcohol wavelength range 2000 nm – 2550 nm is detecting and calibrating on alcohol and other absorbing components of the fingers, but not calibrating on water since there are no NIR absorption bands of water in this range.

Experiment 2, 2. Human subject dosing PLS %BAC 2nd Loading, Correlation Plot & Regression Coefficient Calibration (N = 206) Wavelength Range (1350 nm – 2550 nm)

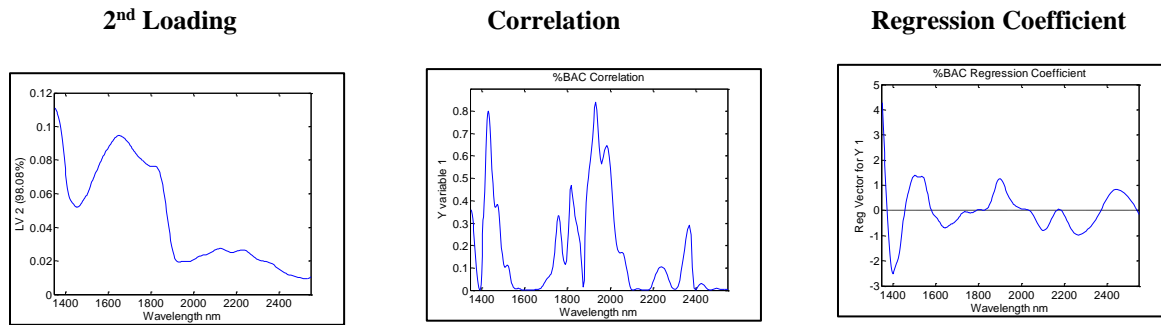


Figure 12. Experiment 2 Human subject dosing calibration range 1350 nm – 2550 nm. Compare this with figure 11. The 2nd loading explains 98.08% model variation due to water, correlation plot showing water and the regression coefficient showing variables contributing most to the model. Note the negative coefficients at approximately 2050 nm, 2300 nm, and a positive coefficient around 2450 nm, is undoubtedly an affect from the alcohol.

What is the takeaway from the *Surrogate* experiment and subsequent review of the *Human* experiment being analyzed in PLS using the full wavelength and alcohol range (2000 – 2550 nm) for calibration? Correlation is not causation. Figure 13 shows the regression scatter plots for the surrogate and human subject data sets from a full range and an alcohol range calibration.

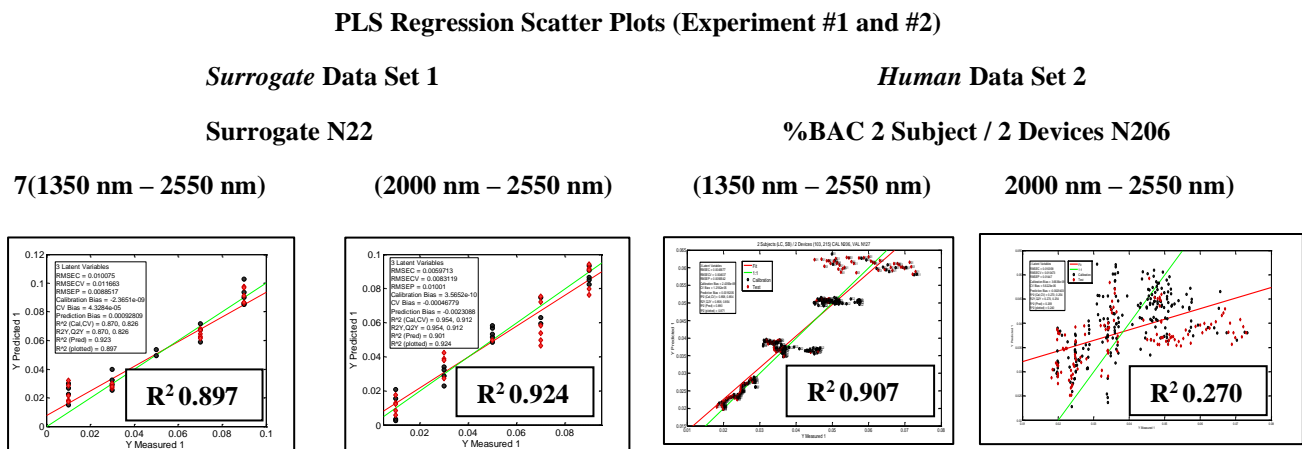


Figure 13. PLS regression plots of calibration (left) and validation (left middle) of surrogate data set, compared to calibration (right middle) and validation (right) of human data set.

The correlation for the 2000 nm – 2550 nm range for data set 2 falls from $R^2 = 0.907$ to $R^2 = 0.270$ when using the full wavelength range. This is in contrast to the *Surrogate* experiment, where the calibration correlation went up from $R^2 = 0.897$ to $R^2 = 0.924$. This leads to the conclusion that the correlation for %BAC (*Human*) was caused mainly by the water in the 1350 nm – 2550 nm range. When the calibration wavelength range was shortened to the alcohol range in the *Surrogate* experiment, alcohol was the dominant component contributing to the calibration as shown in figure 18. Note that alcohol now dominates the regression coefficient, as evidenced by the dramatic change in the first loadings. They no longer resemble water providing strong evidence that its influence has been reduced in the 2000 nm – 2550

nm range). The explanation for the cause of the correlation for both the *Surrogate* and *Human* spectra has now been correctly determined using loadings, correlation, and regression coefficient plots.

PLS on *Human* experiment, %BAC 2nd Loading, Correlation Plot & Regression Coefficient Calibration (N = 206) Wavelength Range (2000 nm – 2550 nm)

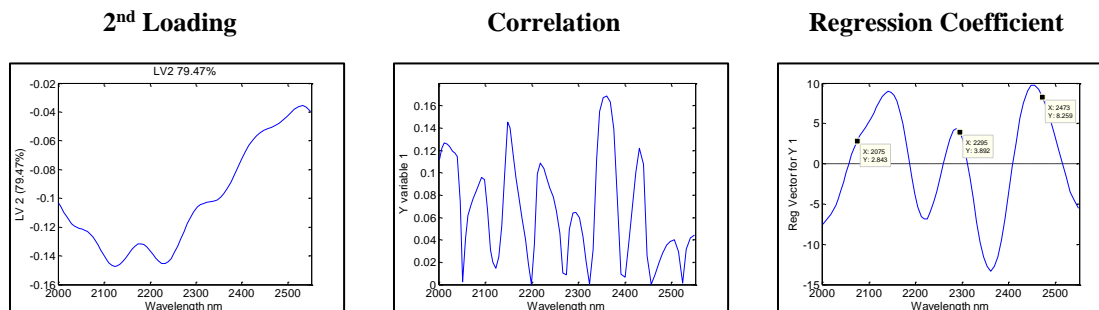


Figure 14. Shows the loadings, correlation, and regression vector plots for human subject data set 2 (2000 nm – 2550 nm) calibration range.

Figure 14 shows that while the regression coefficient is using the wavelengths of those associated with alcohol (see figure 2 and table 1). The second loading from figure 14 explains 74.47% of the model variation, but it does not follow the same pattern as in the surrogate or the %BAC human (N206) full-range calibration. It can only be concluded that other absorbers and non-absorbing phenomenon such as noise and scatter, are also contributing to the variance seen in the model. Moreover, the very low value and random appearing correlation plot may be showing that wavelength multicollinearity is impacting the calibration even more than that already present in the full 1350 nm -2550 nm range. This is not unexpected and can be explained by the reduced number of wavelength variables in the 2000 nm – 2550 nm range (80 compared to 257, when the full range is used), and the decrease of signal to noise.

Experiment 2 (*Human*): The DQM Equation Applied to %BAC Estimation for Human Subjects

Data from *Human* experiment (%BAC 2 Subject / 2 Devices) N206 can now be tested by the DQM regression program, following the parameters used on the *Surrogate* experiment, except for two differences. First, because the correlation obtained in the data set 2 %BAC 2 subject /2 devices (CAL N206) VAL (N127), it was postulated that using the large number of samples for the 2000 nm - 2550 nm range as was originally used for the full range calibration (N206) was introducing nonlinearities. The data set was trimmed down to CAL N159 and VAL (98) using combined leverage, KNN and manual outlier removal. Second, the spectra from the *Human* experiment have significantly lower signal to noise ratio than the *Surrogate* experiment, the gap / smoothing was set to a maximum of 6 and 12 respectively, based on the analysis of the alcohol peak at 2295 and measuring the number of data points in the peak at half-width (see figure 3).

The DQM parameters selected for searching the wavelengths from 2000 nm – 2550 nm (alcohol range) (at 32 nm resolution, there are 80 wavelengths) used were:

1. # of Calibration Samples: N159
2. # of Validation Samples: N98
3. Number of terms (1 or 2): 2
4. Derivative order (d): Term #1: 1D, Term #2: 2D
5. Number of differential gaps (gap): 6
6. Number of smoothing points (smt): 12

Figures 15, 16, and 17 give the correlation scatter plot, the derivatives, the SEC, and regression coefficients plots from the analysis. Table 4 provides the DQM final result for the 2-term model for the Determination of %BAC on two subjects and two devices. The final results following implementing the dqm1 program were:

%BAC Regression Results

SEC = 0.005 RSQ = 0.8230 N = 206

RMSEP = 0.004 SEP = 0.004 bias = 0.002 RSQ = 0.9501 N = 98

2-Term Model: %BAC S1 S2 N206 CAL 31.mod1, Derivatives normalized

Term #1: (1D 2198.8024 NM, gap=6, smt=10)/(2D 2147.3685 NM, gap=6, smt=12)

Term #2: (1D 2338.8535 NM, gap=2, smt=2)/(2D 2057.1429 NM, gap=4, smt=11)

Coefficients B0, B1, B2, .. = 0.064373 -0.002534 -0.00018087

**Optimal Derivatives Selected For %BAC Spectra CAL (N159) VAL (98) By The DQM Algorithm
(2-Term Model 1D/2D)**

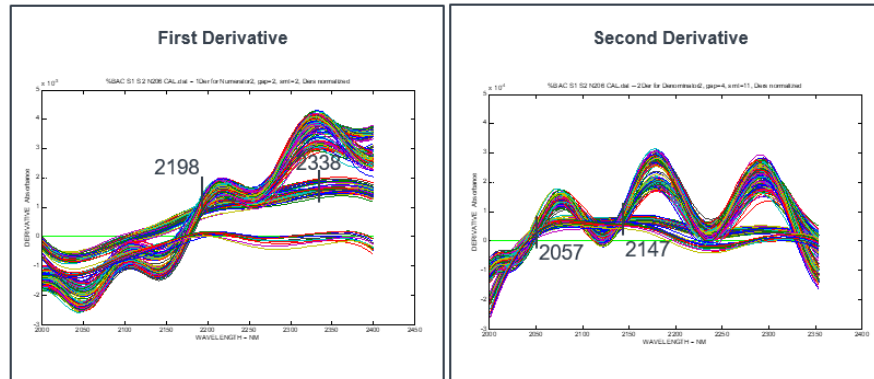


Figure 15. Optimal Derivatives Selected For %BAC Spectra CAL (N159) VAL (98) By the DQM Algorithm. As was shown for the surrogate DQM model experiment 1, note the prominence of regions of the spectra where wavelengths have been selected that correspond to regions of no scatter: the cross-over regions of the spectra. Other wavelengths center on a region of the alcohol analyte where known NIR absorption is expected to occur.

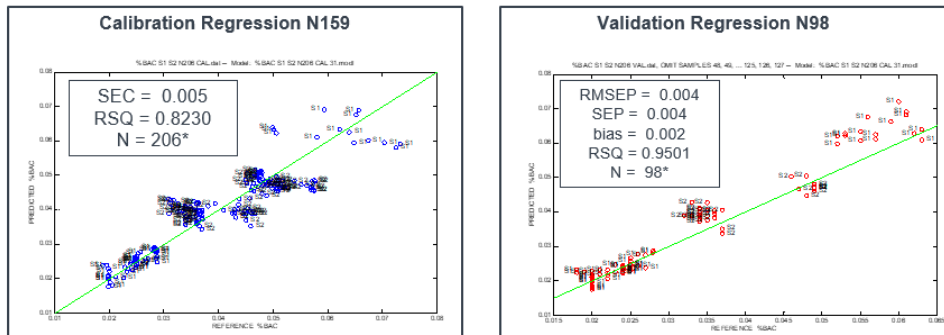


Figure 16. %BAC Spectra CAL (N159) VAL (98) Scatter Plot Standard Error Calibration (SEC) and Regression Coefficient Plots for the DQM results of the %BAC Spectra CAL (N159) VAL (98)

SEC Plot

Regression Coefficient Equation

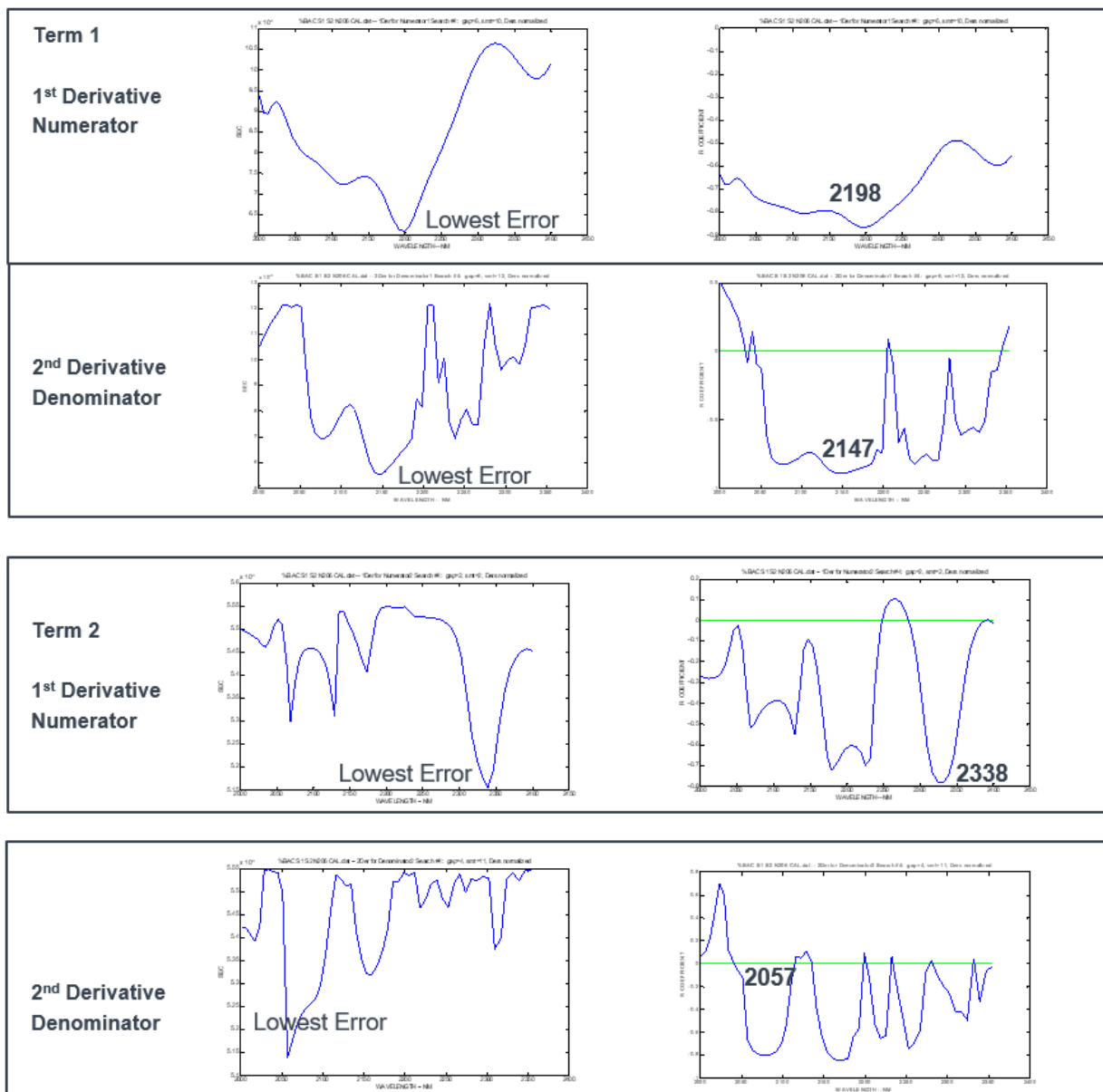


Figure 17. Standard Error Calibration (SEC) and Regression Coefficient Plots for the DQM results of the Human Subjects

TABLE 4.

DQM Results for the 2-Term and 1-Term Models for the Determination of %BAC in two subjects and two devices by NIR

Wavelengths	2-Terms	GAP								
Derivative Ratio	SEC (%BAC)	N1	smt	D1	smt	N2	smt	D2	smt	R ²
1D / 1D Term 1 2252 / 2034 Term 2 2172 / 2062	0.005	4	12	2	12	6	0	2	0	0.810
1D / 2D Term 1 2198 / 2147 Term 2 2338 / 2057	0.005	6	10	6	12	2	2	4	11	0.823
2D / 1D	0.006	5	0	3	0	5	2	5	1	0.767
2D / 2D	0.006	1	0	1	0	5	9	6	12	0.739
	1-Term	GAP								
	SEC (%BAC)	N1	smt	D1	smt					R ²
1D / 1D	0.006	4	12	2	12					0.754
1D / 2D	0.006	6	10	6	12					0.793
2D / 1D	0.008	5	0	3	0					0.597
2D / 2D	0.007	1	0	1	0					0.645

PLS calibration on %Alcohol NIR Extended Multiplicative Scatter Correction (EMSC) (N = 21) Set using the alcohol wavelength range (2000 nm – 2550 nm)

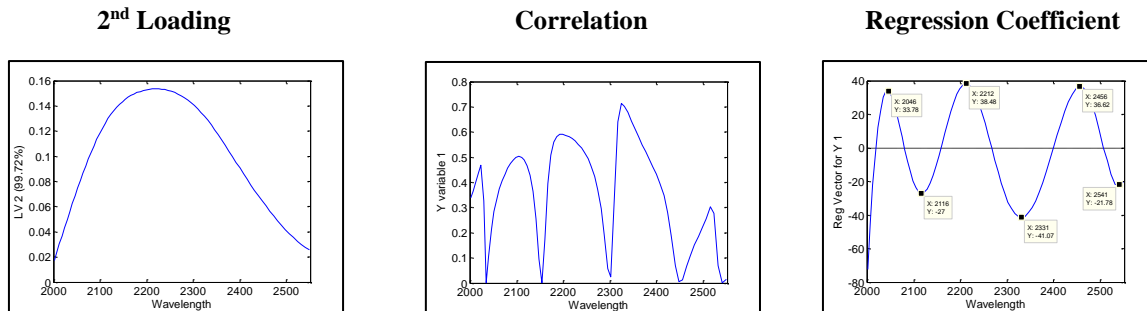


Figure 18. PLS calibration on %Alcohol NIR Extended Multiplicative Scatter Correction (EMSC) CAL (N22) Surrogate data set using the alcohol wavelength range (2000 nm – 2550 nm). The second loading explains 99.72% of the model variance and is now dominated by positive wavelength variables specific for alcohol. Note the coefficients around wavelengths closely associated with alcohol.

DISCUSSION

The PCA, PLS, and DQM analysis applied to NIR spectra obtained from surrogate mixtures of alcohol : water and two human subjects on two benchmark devices demonstrate several aspects of the measurement of %BAC that fulfill the requirements for an accurate and precise measurement to be made by NIR. However, challenges still remain considering instrument Signal to Noise S/N, limited spectral resolution, instrument variability, and measurement sampling variability contributing to additional scatter from nonlinear phenomenon related to biological tissue heterogeneity. Observations from these experiments and analysis using the NIR-AS prototype sensors are the following:

1. The correlation has been shown to be attributed to wavelengths that are specific for the absorption of alcohol in the wavelength range from 2000 nm – 2550 nm. The first and second latent variables of the calibration data set can be used to explain the high linear correlation and model wavelengths variance contribution to the model.
2. PCA of the residual spectra is useful for attributing the cause of the correlation from increasing concentration of the water or alcohol in these experiments.
3. The breathalyzer measurements are accurate and precise, at least sufficient to be used on several different subjects, correlated to different devices and are reproducible over time despite the high variance of the human data set.
4. The analysis presented in this paper is standard protocol for working through PLS model results. The analysis of loadings, correlation, regression coefficients, and residual plots are necessary for establishing the cause of the PLS correlation model. The difference between water NIR absorption versus alcohol NIR absorption show up in the loadings and regression coefficients indicating the source of the correlation of x-matrix wavelengths to Y-matrix breathalyzer values when either the full wavelength or the reduced wavelength scale are selected for calibration.
5. In this study, the basic experimental design of pairing laboratory *in vitro surrogate* (alcohol-water mixture) studies with *human* studies, in order to have a simple frame of reference (low noise, no scatter) for interpreting the spectra and resulting model information for analysis was the key to understanding the complex nature of %BAC determination in humans. This is a major milestone for this experiment and the adoption of this simple approach should be useful in guiding the next recommended R&D phase: human dosing and clinical analysis of breath and blood.
6. PLS analysis shows that there is a significant correlation between BAC breathalyzer values and NIR measurements at low dosing levels. While the use of PLS for calibrating multiple human subjects and NIR devices looks promising in the 2000 nm – 2550 nm wavelength range, further work is ongoing for improving the NIR-AS, design of experiments, alcohol dosing spectral preprocessing and ongoing model variance studies over time.
7. A correlation on the surrogate data set was made using a wideband NIR device (NIR-AS) to measure very low concentrations of alcohol in the range from 0.01% to 0.01%. The same range was used for the determination of the %BAC in humans and gave similar results. This was useful for studying and understanding the cause of the loadings, regression coefficient and correlation plots in a low noise and non-scattering environment. This approach is a significant finding that will aid in improving the human studies, device design and software performance for future NIR-AS devices.
8. Sampling averaging leads to improved correlations. The design of experiment for the human subject data set takes into consideration one of the requirements of measuring low amounts of a constituent on a wide band device.
9. Averaging the spectra from multiple samples of the same constituent level in order to reduce the sampling error, and hence the impact of spectra preprocessing and final model variance due to unwanted physical effects.
10. Attention paid to the handling of spectral outliers and the design of individual data sets are paramount if good correlation models are to be robust.

The introduction of the use of the DQM regression tool for both calibrating and diagnosing data sets used by both PLS and PCA approaches offers renewed insight into the causes of correlation in determining %BAC. The method can be

used to target the identification of derivative ratios at optimal wavelengths that are specific for the alcohol analyte by adjusting the wavelength range, derivative order, gap, and smoothing parameters. DQM offers a specific and sensitive (detection of the analyte in the presence of interfering absorbers and the elimination of multiplicative scatter) method for the accurate and precise determination of BAC in humans by NIR spectroscopy.

Norris [16] describes the utility of DQM moreover, he explains the method as it applies to single, and two term 1st and 2nd derivatives combined with gap and smoothing pretreatment and is excerpted below.

Near infrared spectra of diffusely reflecting samples [e.g. human skin] are characterized by:

- poor reproducibility
- poor linearity
- high noise and high sensitivity to sample measurement geometry.

Typical calibration procedures for such spectra involve pretreatment with multiplicative scatter correction or standard normal variant correction, first or second derivative, and partial least squares regression. In the case of determining BAC in NIR, spectral preprocessing using EMSC followed by DQM, has proven successful as shown above. Studies are ongoing for exploring the utility of the DQM for the determination of BAC using multiple subjects and devices for understanding how DQM can be used in conjunction with other techniques, to arrive at a better understanding of BAC measurements, correlations to breath, blood, or other biological markers identified for alcohol intoxication and the goal of achieving a chemometric predictive model for %BAC by NIR.

CONCLUSION

The determination of blood alcohol concentration by NIR has been shown to primarily the result of wavelengths in the 2000 nm – 2550 nm region of the spectrum. Based on the known NIR absorption bands from alcohol reference material, and the known NIR absorption spectrum of water, it was shown by exclusion and inclusion experiments of portions of the wavelength range that correlate solely to water (1350 nm – 2000 nm) and to alcohol (2000 nm – 2550 nm), that the cause of the correlation and regression equation could be attributed. More importantly, the PLS equation in the wavelength range from 2000 nm - 2550 nm was shown to not be able to regress on two subjects measured on two devices. The reasons for this are not immediately obvious, but sources of variance from multiple subjects and multiple devices as well as diminished signal to noise and absorption from other constituents (i.e., hemoglobin, protein, fat and other NR absorbers in human tissue) are likely the leading causes. This immediately led to the design of the experiments for including the surrogate for analysis and trying to understand the cause of this.

Again, the authors introduced DQM for the reason of trying to understand the cause of the alcohol correlation failing in the human study, but succeeding in the surrogate study using PLS chemometrics. The output of the DQM reveals several attributes of NIR-AS spectra.

1. Wavelength selection for the optimal detection of alcohol correspond to spectral regions of no scatter (the crossover regions) seen in the surrogate and human spectra.
2. The effect of the selection of wavelength range for inclusion for DQM analysis also shows that DQM will select regions of spectra that have a low error for wavelengths in the crossover regions of the spectra, not necessarily from the analyte of interest. For instance, limiting the range to 2000 nm – 2550 nm for gap, smoothing, derivative selection, and optimization will correlate and regress on the alcohol for both the surrogate and human data set. However, when the full wavelength range is selected (1350 nm – 2550 nm), water is selected for calibrating against the breathalyzer values just as in the PLS algorithm as shown for both the surrogate and human data.
3. The regression vector explaining the correlation of the final DQM model can be shown by the PLS analysis of the same data set using loadings, correlation, and regression coefficients that the analyte in the DQM model is in fact alcohol based on known NIR absorption of reference spectrum.

Further experiments have already been carried out and additional are underway for refining the surrogate experiment (i.e., controlling evaporation and understanding the contribution of scatter from the cotton matrix), controlling NIR measurement of the finger, and optimizing the NIR test device for enhanced signal to noise and resolution. The next

phase for beta testing the devices will focus on completing the pre-clinical study of multiple human subjects on multiple test devices and regression analysis of the data by DQM, PLS, and PCA. The goal of providing a global model for the determination of BAC in humans, which we believe is achievable based on this and other decisive and targeted studies designed to understand the cause of correlation and regression of the BAC model.

REFERENCES

- [1] 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Paper Number 17-0304, DRIVER ALCOHOL DETECTION SYSTEM FOR SAFETY (DADSS) – PRELIMINARY HUMAN TESTING RESULTS, Lukas, S., Zaouk, A., Ryan, E., McNeil, J., Sheperd, J., Willis, M., Dallal, N., and Schwartz, K., (2017)
- [2] K.H. Norris, NIR news Vol. 13 No. 3 (2002)
- [3] Wold S. Spline-funktioner-ett nytt vertyg i data-analysen. *Kemisk Tidskrift*. 1972; 84(3): 34– 35
- [4] Thomas G. Mayerhöfer, Susanne Pahlow, and Jürgen Popp, The Bouguer-Beer-Lambert Law: Shining Light on the Obscure *ChemPhysChem* 2020, 21, 1–19
- [5] Rowe, R.C., Sheskey, P.J. and Quinn, M.E. (2009) *Handbook of Pharmaceutical Excipients*. 6th Edition, Pharmaceutical Press
- [6] Ma, L., Peng, Y., Pei, Y. *et al.* Systematic discovery about NIR spectral assignment from chemical structural property to natural chemical compounds. *Sci Rep* 9, 9503 (2019)
- [7] A.M.C., Davies, Back to basics: spectral pre-treatments – derivatives, *Spectroscopy Europe*, pgs. 32-33. Vol. 19 No. 2 (2007)
- [8] [<http://www.impublications.com/discus/messages/5/237.html?1172691604>].
- [9] Ciurczak, E.W., Igne, B., Workman, Jr., J., & Burns, D.A. (Eds.), Chapter 14 Derivative Quotient Method Regression, or the Norris Regression, *Handbook of Near-Infrared Analysis* (4th ed.). CRC Press, pgs. 269 -287, (2021)
- [10] [<http://www.impublications.com/discus/messages/5/237.html?1172691604>].
- [11] Wanga, X., Baoa, Y., Liua, G., Li, G., Lin, L., Study on the Best Analysis Spectral Section of NIR to Detect Alcohol Concentration Based on SiPLS, *Procedia Engineering* 29, 2285 – 2290, 1877-7058 © 2011 Published by Elsevier Ltd. doi:10.1016/j.proeng.2012.01.302 (2012) Available online at www.sciencedirect.com
- [12] Martens, H., Nielsen, J. P., and Engelsen, S. B., Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures, *Anal. Chem.* 75, 394 – 404, (2003)
- [13] S. Wold, K Esbensen and P. Geladi, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 37-52
- [14] K. Pearson, On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, (6) 2 559-572, (1901)
- [15] L. Nørgaard, R. Bro and S. Balling Engelsen, *Principal Component Analysis and Near Infrared Spectroscopy*, A white paper from FOSS
- [16] K.H. Norris, NIR news Vol. 12 No. 3 (2001)