Glucose Monitoring in Various Matrices with Near-Infrared Spectrometry and Chemometrics

Jue Qian
University of Iowa

Copyright 2013 Jue Qian

This dissertation is available at Iowa Research Online: https://ir.uiowa.edu/etd/4901

Recommended Citation
https://ir.uiowa.edu/etd/4901.
GLUCOSE MONITORING IN VARIOUS MATRICES WITH NEAR-INFRARED SPECTROMETRY AND CHEMOMETRICS

by

Jue Qian

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Chemistry in the Graduate College of The University of Iowa

August 2013

Thesis Supervisor: Professor Mark A. Arnold
This is to certify that the Ph.D. thesis of

Jue Qian

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Chemistry at the August 2013 graduation.

Thesis Committee: __________________________
Mark A. Arnold, Thesis Supervisor

______________________
Gary W. Small

______________________
Claudio J. Margulis

______________________
James B. Gloer

______________________
Julie L. P. Jessop
To my parents, my wife Xiaoxiao and my lovely kids Angela and Amy
ACKNOWLEDGEMENTS

I would like to sincerely thank all people who have supported and inspired me during my doctoral study and research.

My foremost and deepest gratitude goes to my research advisor, Dr. Mark A. Arnold. I really appreciate him for his patience and encouragement that carried me on through all research projects since I joined his group in 2006. His enthusiasm and knowledge shaped and improved my research skills, and for his strict, extensive and creative comments guided me in my career. Without him, it is impossible for me to accomplish my dissertation. Furthermore, he is an outstandingly successful scientist and a very nice person to work with.

Another extreme gratitude is for Dr. Gary Small, as my mentor for chemometrics. Dr. Small led me into the world of chemometrics when I took his lecture-Chemometrics and provided me invaluable discussions and instructions for my research projects as well as priceless comments and suggestions as committee member.

It is my pleasure to give my special thanks to my other committee members: Dr. Claudio J. Margulis, Dr. James B. Gloer and Dr. Julie L.P. Jessop, who took effort in reading and providing me with valuable comments on my thesis.

I also would like to thank my colleagues who always gave me great encouragement and suggestions. I am honored to express my sincere acknowledgement to Dr. Jonathon T. Olesberg for helping with optical setup and configuration; Dr. Min Ren, for teaching me NIR spectra collection and collecting bovine blood ultrafiltrate; Dr. Chuannan Bai, for sharing me chemometrics knowledge and Matlab codes; Dr. Natalia Alexeeva and Ryan Smith, for discussing on chemometrics algorithms; Terry Graham, for animal caring and experiments, and the other members in Dr. Arnold’s group.

Last but not least, I would like to thank my family for their selfless support all through my doctoral studies.
The long-term complications of diabetes can be dramatically reduced with tight glycemic control. Although the current invasive technology for measuring blood glucose is effective, it is not well suited for the real-time measurements necessary for tight control. Near infrared (NIR) absorption spectroscopy, coupled with multivariate calibration modeling, can potentially provide portable, rugged and low-cost instrumentation for continuous glucose sensing. An optical microsensor that can be used in conjunction with an ultrafiltration sampling probe is under development for continuous glucose measurements in interstitial fluid (ISF) collected from subcutaneous tissue.

The first part of this research focused on the development of an algorithm for eliminating the effect of temperature variance on NIR glucose measurements. Spectra of 80 bovine blood ultrafiltrate samples were collected under 5 different temperatures by using a Fourier transform (FT) NIR spectrometer. Based on the fundamental properties of digital Fourier filtering, baseline variations created by differences in the temperature of the blood ultrafiltrate samples were shown to be eliminated by using an optimized Gaussian shape filter response function. Partial Least Square (PLS) calibration models combined with digital Fourier filtering provided standard errors of prediction in the range of 0.3-0.4 mM for samples with temperatures between 25-40 °C.

Before applying the microsensor to animal or human measurements, a testing platform was designed and constructed for the eventual purpose of evaluating the ability of the microsensor to follow glucose concentration transients. A series of computer-controlled pumps were used in combination with an ultrafiltration probe to create glucose transients and deliver the corresponding samples to the spectrometer for analysis. NIR spectra were collected continuously as the concentrations of glucose, urea, and lactate were varied independently. Glucose transients were followed over periods of days by using either partial least squares (PLS) or net analyte signal (NAS) calibration methods.
The NAS calibration method and a modified hybrid linear analysis (HLA) method were investigated for monitoring the concentrations of glucose and lactate during microbial fermentations. An Ultrafiltrate (UF)-sampling probe was used to collect samples of the fermentation broth and deliver these samples to the spectrometer for continuous analysis. The established NAS and modified-HLA calibration models provided glucose and lactate concentration measurements with mean percentage errors of 2 and 3%, respectively. These calibration functions were demonstrated capable of accurate concentration measurements several days beyond the formal calibration process.

Lastly, NIR spectra of whole bovine blood samples were used to demonstrate the ability to measure glucose in blood with different levels of hematocrit. Calibration functions were based on PLS modeling, and the effective models were developed for measurements from absorbance and single-beam NIR spectra. The method of multiplicative scatter correction was found to be particularly effective in reducing the impact of light scattering caused by the red blood cells at different hematocrit levels. These findings imply that nondestructive NIR spectroscopy has the potential to measure glucose without consuming blood, thereby reducing phlebotomy blood loss in neonates and potentially decreasing the frequency of red blood cell transfusions for this fragile patient population.
# TABLE OF CONTENTS

LIST OF TABLES .............................................................................................................. ix
LIST OF FIGURES ............................................................................................................ x
LIST OF ABBREVIATIONS ............................................................................................... xv

CHAPTER

I INTRODUCTION ............................................................................................................. 1

Diabetes .......................................................................................................................... 1
Continuous Glucose Monitoring (CGM) ................................................................. 2
Noninvasive Glucose Sensing .................................................................................... 4
Near Infrared Spectroscopy ....................................................................................... 10
Fourier Transform Spectrometers .......................................................................... 13
Multivariate Calibration Analysis .......................................................................... 15
  Classical Least Squares ......................................................................................... 16
  Inverse Least Squares ............................................................................................ 17
  Multiple Linear Regression .................................................................................... 18
  Principal Component Regression .......................................................................... 19
  Partial Least Squares Regression .......................................................................... 21
  Net Analyte Signal .................................................................................................. 23
  Hybrid Linear Analysis ........................................................................................... 25
  Digital Fourier Filtering .......................................................................................... 26
  Overview of Thesis .................................................................................................. 27

II TEMPERATURE-INSENSITIVE NEAR-INFRARED SPECTROSCOPIC MEASUREMENT OF GLUCOSE IN BOVINE BLOOD ULTRAFILTERATE ........................................................................ 29

Introduction .................................................................................................................. 29
Experimental Section .................................................................................................... 31
  Sample Preparation .................................................................................................. 31
  Instrumentation and Apparatus ................................................................................ 32
  Spectral Collection .................................................................................................... 33
  Data Processing ......................................................................................................... 33
Results and Discussion ................................................................................................. 36
  Temperature and Absorption Band .......................................................................... 36
  Digital Fourier Filtering ........................................................................................... 38
  PLS Calibration Models ........................................................................................... 42
  PCA Analysis ............................................................................................................. 44
  Effect of Spectral Resolution .................................................................................... 48
Conclusions ..................................................................................................................... 51

III CREATION AND MEASUREMENT OF IN VITRO GLUCOSE TRANSIENTS FOR MICROSENSOR DEVELOPMENT ........................................................................... 56

Introduction ..................................................................................................................... 56
Experimental Section ..................................................................................................... 58
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-1.</td>
<td>Currently available continuous glucose monitoring devices</td>
<td>6</td>
</tr>
<tr>
<td>II-1.</td>
<td>Error summary from the PLS models with and without digital Fourier filter</td>
<td>46</td>
</tr>
<tr>
<td>II-2.</td>
<td>Results from the PLS models based on temperature mismatch absorbance spectra</td>
<td>52</td>
</tr>
<tr>
<td>III-1.</td>
<td>Pump calibration slopes determined for implementation of 22 calibrations for each of five pumps over a 30-day period</td>
<td>69</td>
</tr>
<tr>
<td>III-2.</td>
<td>Delay time measurements based on cv-PLS models</td>
<td>72</td>
</tr>
<tr>
<td>III-3.</td>
<td>Summary of the within calibration performance of PLS calibration models for glucose</td>
<td>77</td>
</tr>
<tr>
<td>III-4.</td>
<td>Comparison of PLS Glucose Concentration Prediction Errors Outside and Within the Calibration Process</td>
<td>82</td>
</tr>
<tr>
<td>III-5.</td>
<td>Summary of NAS Model Predictions within Calibration Days and Comparison to within Calibration Day Prediction Errors for PLS Models</td>
<td>83</td>
</tr>
<tr>
<td>III-6.</td>
<td>Comparison of NAS and PLS Glucose Concentration Prediction Errors Outside the Calibration Process</td>
<td>88</td>
</tr>
<tr>
<td>IV-1.</td>
<td>NAS and PLS calibration model results for tracking the concentration of glucose during a single fermentation</td>
<td>107</td>
</tr>
<tr>
<td>IV-2.</td>
<td>Comparison of NAS, modified-HLA and PLS calibration models for lactate concentration measurements within a single fermentation</td>
<td>119</td>
</tr>
<tr>
<td>IV-3.</td>
<td>NAS and PLS calibration model results for glucose concentration prediction across fermentations</td>
<td>125</td>
</tr>
<tr>
<td>IV-4.</td>
<td>Modified-HLA and PLS calibration model results for lactate concentration prediction across fermentations</td>
<td>128</td>
</tr>
<tr>
<td>V-1.</td>
<td>Average RMS noise levels for whole blood samples composed of different hematocrit levels</td>
<td>142</td>
</tr>
<tr>
<td>V-2.</td>
<td>cv-SEP values from the PLS glucose calibration model based on absorbance spectra for bovine blood samples with different hematocrit levels</td>
<td>150</td>
</tr>
<tr>
<td>V-3.</td>
<td>cv-SEP values from PLS model based on single-beam spectra after MSC pre-processing for different hematocrit levels</td>
<td>156</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-1</td>
<td>Compartmental model for the blood-ISF glucose relationship.</td>
<td>5</td>
</tr>
<tr>
<td>I-2</td>
<td>NIR spectra of glucose (black) and other major biochemical analytes (See legend for details) collected over the combination region of the NIR spectrum.</td>
<td>9</td>
</tr>
<tr>
<td>I-3</td>
<td>Schematic diagrams of transmission and diffuse reflectance NIR measurements.</td>
<td>12</td>
</tr>
<tr>
<td>I-4</td>
<td>Schematic diagram of the Michelson interferometer.</td>
<td>14</td>
</tr>
<tr>
<td>I-5</td>
<td>Illustration of NAS projection.</td>
<td>24</td>
</tr>
<tr>
<td>II-1</td>
<td>Sampling system coupled with a thermoelectric cooler (A) and temperature controller (B).</td>
<td>34</td>
</tr>
<tr>
<td>II-2</td>
<td>Absorption spectra of distilled water at temperature of 20, 25, 30, 35, 40 °C (A); Synchronous variable-variable two-dimensional correlation spectra (B), generated from the absorption spectra shown in (A).</td>
<td>37</td>
</tr>
<tr>
<td>II-3</td>
<td>Temperature mismatch absorbance spectra calculated with reference spectra collected at 35 °C (A); Concentration correlation plot of the predicted glucose concentrations of PLS model (B), based on the temperature mismatch absorbance spectra shown in (A).</td>
<td>39</td>
</tr>
<tr>
<td>II-4</td>
<td>Absorbance spectra in frequency domain(A), and optimized Gaussian shape digital filter (solid blue line). Response surface map for optimization of mean and standard deviation of digital filter based on 1/SEP (B).</td>
<td>41</td>
</tr>
<tr>
<td>II-5</td>
<td>Errors in predictions from PLS models generated with single temperature spectra and combined temperature without digital Fourier filtering.</td>
<td>45</td>
</tr>
<tr>
<td>II-6</td>
<td>Temperature mismatch absorbance spectra after digital Fourier filtering (A); Concentration correlation plot of the predicted glucose concentrations of PLS model (B), based on absorbance spectra shown in (A).</td>
<td>47</td>
</tr>
<tr>
<td>II-7</td>
<td>Score plots from PCA models without (A) and with (B) digital Fourier filtering defined by a mean position of 0.0097 f and standard deviation width of 0.0037 f.</td>
<td>49</td>
</tr>
<tr>
<td>II-8</td>
<td>First loading of PCA models with (red) and without (blue) digital filtering.</td>
<td>50</td>
</tr>
<tr>
<td>II-9</td>
<td>Frequency domain spectra (left y-axis) and optimized Gaussian shape digital filter (right y-axis) under resolution 2 cm-1 (A) and 4 cm-1 (B).</td>
<td>54</td>
</tr>
</tbody>
</table>
Correlation plots of resolution and optimized parameters for digital filtering for PLS models with absorbance spectra of temperature match (A) and mismatch (B). ................................................................. 55

Schematic (A) and assembled view (B) of a prototype microsensor designed for in vivo glucose sensing. ................................................................. 59

Schematic of the ultrafiltration process where molecular flow is driven by a negative pressure across a semi-permeable membrane. .................. 60

Schematic layout of the experimental design for generating and measuring glucose transients................................................................. 62

Four designed concentration profiles (A, B, C, and D) with different varying trends and concentration levels for each analyte. ................. 64

Linear relationship between flow rate and pump frequency for calibration of pump 1 ................................................................. 67

Calibration results of 22 trials over a period of 30 days for the peristaltic pump system ................................................................. 68

Identification of delay time based on cv-SEP values from cross-validation PLS models while systematically varying delay time. ............ 71

Pure component spectra for 1 mM solutions of glucose, urea and lactate (A). Examples of sample spectra collected during a single day experiment (B) ................................................................. 74

Trend of SEP values versus number of factors used in PLS modeling... .... 76

Glucose concentration time profile as defined by the reference (blue), calibration (green) and prediction (red) measurements (A). Concentration correlation plot for glucose PLS model (B) ................................................................. 78

Glucose concentration transient profiles for reference (blue), PLS calibration (green) and PLS predicted (red) concentrations within and outside the calibration process (A); and concentration correlation plot for PLS glucose concentration predicts within and outside the calibration process (B) ................................................................. 80

Predicted glucose concentrations versus time, varying follow reference concentrations (A). Concentration correlation plot for glucose NAS model during same day as calibration spectra (B) ................................................................. 85

Comparison of calibration vectors for PLS and NAS models... ................. 86

Glucose concentration transient profiles for reference (blue), and NAS predicted (red) concentrations within and outside the calibration process (A); and concentration correlation plot for NAS glucose concentration predicts within and outside the calibration process (B)...... 87
### IV-1. Diagrammatic representation of the different sample presentation modes and measurement approaches used for bioprocess monitoring with NIR spectroscopy ................................................................. 92

### IV-2. Schematic layout of experimental design for fermentation process monitoring .................................................................................................................................................................................. 95

### IV-3. Growth curve of the cells by measuring OD$_{600}$ of the fermentation broth sample during the experiment ......................................................................................................................................................................... 98

### IV-4. Single-beam spectral intensities (A) and RMS noise values (B) for spectra collected from the UF sampling probe. High values are an indication of an air-bubble in the optical cell .................................................................................................................................................. 100

### IV-5. Pure component spectra of glucose, yeast extract, yeast nitrogen base w/o amino acid and soy peptone (A) and time profile for the coefficient ratios for the four components within the fermentation medium (B) ........................................ 102

### IV-6. Spectral residuals determined over the period of the fermentation after a four component regression analysis (A) and pure component spectra of glucose and lactate for byproduct identification (B) ........................................................................................................ 103

### IV-7. Time profile (A) and concentration correlation plot (B) for PLS calibration model for measuring the concentration of glucose during the fermentation ................................................................................................................................................................. 105

### IV-8. Time profile (A) and concentration correlation plot (B) for NAS calibration model for measuring the concentration of glucose during the fermentation ......................................................................................................................................................................... 106

### IV-9. Comparison of calibration vectors for PLS and NAS models ........ 110

### IV-10. Magnification of the last 1.5 hr of the glucose concentration profiles for the NAS and PLS models along with the YSI readings. ................................................ 111

### IV-11. Lactate spectra collected under different pH values ........................................ 113

### IV-12. Original sample spectra collected during a fermentation (A) and glucose-subtracted spectra used for the modified-HLA measurement of lactate (B) ......................................................................................................................................................... 116

### IV-13. Lactate concentration profiles as determined by the standard NAS method (A) and the modified-HLA method (B) models over the full spectral range (4900-4200 cm$^{-1}$) and NAS (C) and modified HLA(D) models over the pH insensitive spectral range (4500-4200 cm$^{-1}$) .... 117

### IV-14. Comparison of the calibration vectors for PLS (blue) and modified-HLA (red) models for lactate in the UF sampled fermentation broth .... 120

### IV-15. Glucose concentration profiles for fermentation Trial 2 as determined from NAS (A) and PLS (B) calibration models generated from data collected during fermentation Trial 1. ........................................................................................................................................................................... 121
IV-16. Glucose concentration profiles for fermentation Trial 3 as determined from NAS (A) and PLS (B) calibration models generated from data collected during fermentation Trial 1. ................................................................. 122

IV-17. Glucose concentration profiles for fermentation Trial 4 as determined from NAS (A) and PLS (B) calibration models generated from data collected during fermentation Trial 1. ................................................................. 123

IV-18. Comparison of lactate concentrations from theoretical calculation and in situ measured with YSI analyzer through Trail 1................................. 127

IV-19. Lactate concentration profiles for fermentation trial 2 as determined from modified-HLA (A) and PLS (B) calibration models generated from data collected during fermentation trial 1. ................................................................. 131

IV-20. Lactate concentration profiles for fermentation trial 3 as determined from modified-HLA (A) and PLS (B) calibration models generated from data collected during fermentation trial 1........................................ 132

IV-21. Lactate concentration profiles for fermentation trial 3 as determined from modified-HLA (A) and PLS (B) calibration models generated from data collected during fermentation trial 1........................................ 133

V-1. Schematic diagram of experimental configuration used to collect NIR spectra of the whole blood samples.................................................. 137

V-2. Single-beam spectra collected for blood samples with different hematocrit levels and glucose concentrations. .................................................. 140

V-3. Distribution of RMS noise levels for 100% lines for all blood spectra (A) and correlation between RMS noise and hematocrit levels (B)........ 141

V-4. Calculated absorption spectra of all whole blood samples (A) and blood samples with different hematocrit levels and similar glucose concentration range (B)................................................................. 144

V-5. PCA loading vectors from an analysis of all whole blood spectra with different hematocrit levels. ................................................................. 145

V-6. Eigen value plot for the first 12 PC vectors indicating the fraction of spectral variance explained by each factor (A) and principal component score plot (B)...................................................................................... 146

V-7. Surface response plot of cv-SEP values of PLS models with 9 factor models for spectral range optimization.................................................. 148

V-8. Concentration correlation plot for the prediction of glucose concentration from a leave-one-sample-out PLS cross validation model using an optimized spectral range. ................................................................. 149

V-9. Concentration correlation plot of the predicted glucose concentrations of the PLS model with leave-one-sample-out cross validation and optimized spectral range based on single-beam spectra. ............................ 152
V-10. Single-beam spectra of whole blood samples after MSC pre-processing for PLS model. ................................................................. 154

V-11. Concentration correlation plot of the predicted glucose concentrations of the PLS model with leave-one-sample-out cross validation and optimized spectral range based on single-beam spectra with MSC pre-processing................................................. 155

V-12. Calibration vectors from the PLS models based on single-beam after MSC pre-processing and absorption spectra................................. 157
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS</td>
<td>absorbance</td>
</tr>
<tr>
<td>AU</td>
<td>absorbance unit</td>
</tr>
<tr>
<td>CGM</td>
<td>continuous glucose monitoring</td>
</tr>
<tr>
<td>CLS</td>
<td>classical least-squares</td>
</tr>
<tr>
<td>CV-SEP</td>
<td>cross-validation standard error of prediction</td>
</tr>
<tr>
<td>FT-IR</td>
<td>Fourier transform infrared</td>
</tr>
<tr>
<td>HLA</td>
<td>hybrid linear analysis</td>
</tr>
<tr>
<td>ILS</td>
<td>inverse least-squares</td>
</tr>
<tr>
<td>ISF</td>
<td>interstitial fluid</td>
</tr>
<tr>
<td>LOD</td>
<td>limit of detection</td>
</tr>
<tr>
<td>MPE</td>
<td>mean percent error</td>
</tr>
<tr>
<td>MSC</td>
<td>multiplicative scatter correction</td>
</tr>
<tr>
<td>NAS</td>
<td>net analyte signal</td>
</tr>
<tr>
<td>NIR</td>
<td>near infrared</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis</td>
</tr>
<tr>
<td>PLS</td>
<td>partial least square</td>
</tr>
<tr>
<td>RMS</td>
<td>root-mean square</td>
</tr>
<tr>
<td>S/N</td>
<td>signal to noise ratio</td>
</tr>
<tr>
<td>SEC</td>
<td>standard error of calibration</td>
</tr>
<tr>
<td>SEP</td>
<td>standard error of prediction</td>
</tr>
<tr>
<td>SNV</td>
<td>standard normal variate</td>
</tr>
<tr>
<td>UF</td>
<td>ultrafiltrate</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Diabetes

Diabetes is a chronic disease associated with the abnormal metabolism of glucose. Glucose produced from the breakdown of food circulates throughout the body in the blood and serves as a major source of energy. For healthy people, the concentration of glucose is highly regulated through the actions of various hormones, most notably insulin and glucagon. People with diabetes have an inability to either produce or utilize insulin properly, which results in poor regulation of blood glucose levels and produces hyper- and hypo-glycemia. Chronic hyper-glycemia is a major cause of the long-term complications associated with diabetes, including heart attack, stroke, blindness, kidney failure, and nerve damage. Hypo-glycemia can result in an unexpected diabetic coma, which can lead to brain damage or death. Worldwide, diabetes affects 346 million people of all ages. As reported in the 2011 national diabetes fact sheet, the American Diabetes Association estimates that 25.8 million people in the U.S. have diabetes, 27.1% are undiagnosed, and the incidence of diabetes is growing at an annual rate of 11.5%. Overall, diabetes is estimated to be the seventh leading cause of death in the U.S. with an annual financial burden estimated to be 174 billion USD.

There are three major types of diabetes: type I, type II, and gestational diabetes. Type I diabetes, also called juvenile diabetes, is often diagnosed in children, teenagers, or young adults and is manifested by an inability to produce sufficient amounts of insulin. Type II diabetes generally begins with insulin resistance where muscle, liver, and fat cells do not use insulin properly. Gestational diabetes develops in some women during pregnancy and is caused by a shortage of insulin and other hormones. Women that get gestational diabetes have a high probability of developing type II diabetes later in life. Although no cure is available for any of these types of diabetes, studies have shown that
tightly controlled glycemia can significantly delay the onset of its long-term complications. Many health and quality of life benefits have been documented from tight glycemic control.\textsuperscript{6-7}

Conventional home glucose measurements are performed by collecting a sample of capillary blood and applying this blood to a test strip composed of immobilized reagents needed for the quantification of glucose.\textsuperscript{7} The exposed test strip is inserted into a portable glucose meter that provides the corresponding glucose concentration. There are many major brands of portable glucose meters, such as Contour®, ACCU-CHEK®, FreeStyle®, and OneTouch®.\textsuperscript{8} Although different devices have different sensing mechanisms, all of them require blood sampling by lancing a finger, forearm, or thigh. This invasive technology is effective and provides the means for self blood glucose monitoring, which has greatly enhanced the treatment of diabetes. Still, this test-strip technology requires a sample of blood, which can be painful to collect and renders the user susceptible to possible infection. Test-strips are also expensive and provide a single point measurement from which it is impossible to know the rate or direction of a change in the glucose concentration at the time of the measurement.\textsuperscript{9} These limitations fuel the development of implantable glucose sensing technology that can provide continuous readings with a single device over a period of days.

**Continuous Glucose Monitoring (CGM)**

Continuous glucose sensors provide estimates of glucose levels along with direction, rate of change, and rate of acceleration in glucose concentration.\textsuperscript{10} Conceptually, a continuous glucose monitor can be used to create an artificial pancreas where insulin can be delivered in response to the measured concentration of glucose. Two types of continuous glucose monitoring systems are described in the literature. In one case, the glucose concentration is measured in the subcutaneous fluid and, in the other, the glucose concentration is determined directly in blood. Both measurement locations
suffer from issues related to biocompatibility and surface contamination by proteins and biological factors.\textsuperscript{11} Sensors designed for measurements directly in blood are also complicated by the possibly of clots forming on the sensing area. For this reason, all commercial CGM systems are based on an implantable glucose biosensor which is placed in the subcutaneous tissue and make measurements in interstitial fluid (ISF).

The concentration of glucose in the ISF is generally known to correlate to its concentration in blood. In practice, the CGM is calibrated by collecting a sample of blood by a conventional finger stick and relating the signal from the implanted biosensor (located in the ISF) to this concentration through a linear function. This calibration procedure must be repeated periodically to account for differences in the mass transport properties of glucose at the sensing tip of the biosensor.

The above calibration procedure demands that the correlation between glucose in ISF and blood remains constant between calibration points. However, the correlation between ISF and blood glucose is known to be dynamic and dependent on many factors. One model for this correlation is based on the so-called “push-pull phenomenon”, whereby glucose is “pushed” into the ISF in response to a concentration gradient and “pulled” out of the ISF by insulin-induced cellular uptake. The dynamic relationship of glucose in ISF, fat/muscle and blood can be illustrated by Figure I-1.

As depicted in Figure I-1, glucose is transferred through blood capillaries into the ISF by diffusion. This exchange of glucose can be influenced by both blood flow and capillary permeability.\textsuperscript{12-16} A slow glucose transport rate results in a physiological lag between a step-increase in blood and when the corresponding steady-state concentration of glucose is established in the ISF. In addition, cellular uptake of glucose by fat and muscle cells results in lower levels of glucose in the ISF. This cellular uptake is regulated by insulin and other hormones. The concentration of glucose in the subcutaneous ISF is impacted by limited transcapillary exchange of glucose as well as the metabolic rate and glucose uptake by cells.
In practice, a physiology lag on the order of 5-15 min is widely acknowledged and a large degree of variation has been reported for this physiological lag-time. This lag-time complicates the procedure for generating a functioning calibration model that relates sensor current to blood glucose concentrations when the sensor measures glucose in the ISF. Still, commercial implantable glucose biosensors are available and these devices can provide valuable information pertaining to \textit{in situ} glucose concentrations and glycemic trends. Despite some clinical success, these implantable systems suffer from biofouling and uncontrollable variations in the blood-ISF correlation. Table I-1 provides a listing of CGM systems that are available commercially.

Besides the biosensors listed in Table I-1, the GlucoWatch Biographer\textsuperscript{®} (Animas Corp, West Chester, PA, USA), coupled with iontophoresis to extract a sample of interstitial fluid, was developed and approved by the FDA in 2001. However, this device was retired from the market in 2007.

\textbf{Noninvasive Glucose Sensing}

Various methods have been proposed for measuring glucose concentrations noninvasively in people with diabetes. The concept is to pass a band of radiation through a vascular region of the body and determine the concentration of glucose from an analysis of the resulting spectrum. If realized, this noninvasive approach can provide the glucose concentration information in a painless manner and without the need for a fresh test-strip for each measurement. Moreover, a noninvasive approach can be used to track the direction and rate of change in the glucose concentration without complications associated with biofouling. Such technology could be coupled with an insulin delivery system to enable tight glycemic control and improved healthcare.

Approaches toward noninvasive sensing can be direct or indirect in nature. The basis of an indirect approach involves measuring a parameter that is impacted by the concentration of glucose. The most common example of an indirect approach is the
Figure I-1. Compartmental model for the blood-ISF glucose relationship.
Table I-1. Currently available continuous glucose monitoring devices.

<table>
<thead>
<tr>
<th>Device</th>
<th>Company</th>
<th>Available year</th>
<th>Technique</th>
<th>Associated with Insulin Pump</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGMS®</td>
<td>Medtronic Minimed</td>
<td>1999</td>
<td>Subcutaneous sensor</td>
<td>No</td>
</tr>
<tr>
<td>Guardian®</td>
<td>Medtronic Minimed</td>
<td>2004</td>
<td>Subcutaneous sensor</td>
<td>No</td>
</tr>
<tr>
<td>Guardian REAL-Time®</td>
<td>Medtronic Minimed</td>
<td>2005</td>
<td>Subcutaneous sensor</td>
<td>Yes</td>
</tr>
<tr>
<td>Paradigm REAL-Time®</td>
<td>Medtronic Minimed</td>
<td>2006</td>
<td>Subcutaneous sensor</td>
<td>Yes</td>
</tr>
<tr>
<td>STS®</td>
<td>DexCon Inc.</td>
<td>2006</td>
<td>Subcutaneous sensor</td>
<td>No</td>
</tr>
<tr>
<td>SEVEN®</td>
<td>DexCon Inc.</td>
<td>2007</td>
<td>Subcutaneous sensor</td>
<td>No</td>
</tr>
<tr>
<td>SEVEN PLUS®</td>
<td>DexCon Inc.</td>
<td>2009</td>
<td>Subcutaneous sensor</td>
<td>No</td>
</tr>
<tr>
<td>FreeStyle Navigator®</td>
<td>Abbott Inc.</td>
<td>2008</td>
<td>Subcutaneous sensor</td>
<td>No</td>
</tr>
<tr>
<td>Metronic iPro® 2</td>
<td>Medtronic Minimed</td>
<td>2011</td>
<td>Subcutaneous sensor</td>
<td>No</td>
</tr>
</tbody>
</table>
measurement of the scattering properties of skin tissue. Skin scatters light and the magnitude of scattering is impacted by the concentration of glucose in the ISF within the skin structure. Research has shown, however, that indirect methods suffer from a lack of selectivity which greatly limits the utility of indirect approaches.

Direct approaches are based on measurements directly associated with the glucose molecule. Spectroscopic approaches are direct when the probed electromagnetic radiation is selectively absorbed by the glucose molecule. Although certain wavelengths of the radiation are absorbed by components with the tissue matrix, such as water, protein and fat, a portion is absorbed by glucose and the challenge is to find the selective spectroscopic signature for glucose relative to the background tissue matrix. The resulting spectrum can be collected by either a transmission, diffuse reflectance, or transflectance measurement geometry. Given the complexity of the tissue matrix, the resulting spectrum is typically analyzed by a multivariate method in order to relate the tissue spectrum to the concentration of glucose. As noted below, several spectroscopic approaches have been investigated for noninvasive glucose sensing.

Near infrared (NIR) spectroscopy is one of the most widely investigated and promising techniques for noninvasive glucose sensing. NIR photons penetrate skin tissue at the depths of 1 to 100 millimeters, depending on the exact NIR wavelength. Absorption increases and scattering decreases as the wavelength becomes longer. Tissue spectroscopy over NIR wavelengths strongly depends on the spectroscopic properties of water. Water absorbs throughout the NIR spectrum with major absorption peaks centered at (6900, 5200 and 3800 cm\(^{-1}\)). The regions of absorption minima located between these peak absorptions leaves NIR windows into the human body. Past studies compare these regions for analytical measurements and these comparisons demonstrate the analytical utility of the combination region of the NIR spectrum (4000-5000 cm\(^{-1}\) or 2.0-2.5 microns). Figure I-2 provides examples of NIR absorption spectra for glucose and other components found in biological matrices. As illustrated, these NIR absorption
features tend to be low in magnitude and highly overlapping. Still, the shapes of these spectral features are unique for the different molecules and this spectral uniqueness can be the basis for measurement selectivity.

Raman spectroscopy is another promising technology for noninvasive glucose monitoring. *In vitro* experiments indicate good correlation between Raman signals and glucose concentrations, and recent reports demonstrate the potential of using high quality fingerprint Raman spectra for *in vivo* glucose measurements.\textsuperscript{28} Based on the advantage of distinct and unique Raman spectral bands for different molecules, Raman spectroscopy has been applied to glucose measurements in various sample matrices, including water, whole blood and human tissue.\textsuperscript{19,28-35} However, obstacles such as low signal-to-noise, instability of the laser intensity, laser power limit on human subject and presence of a strong background fluorescence limit practical applications of this approach for the management of diabetes.

Optical polarimetry,\textsuperscript{36-38} and photoacoustic spectroscopy,\textsuperscript{39-41} have also been proposed for noninvasive glucose measurements. Optical polarimetry involves rotation of polarized light transmitted through a sample containing an optically active species.\textsuperscript{36} Optical polarimetry demands a non-scattering fluid, which can be obtained by the aqueous humor of a person’s eye. Investigations illustrate that glucose is the predominant optically active element within the aqueous humor pool and the corresponding rotation of polarized incident radiation can be used to measure glucose.\textsuperscript{38} Still, the low analytical signals, complicated instrumentation, and potential of interferences from endogenous chemicals represent barriers to clinical applications of this approach.

Photoacoustic spectroscopy is based on the use of a laser light for excitation of the sample and producing an acoustic pressure wave. Reasonable correlation has been reported between the photo-acoustic signal and glucose concentrations in samples.\textsuperscript{40} As
Figure I-2. NIR spectra of glucose (black) and other major biochemical analytes (See legend for details) collected over the combination region of the NIR spectrum.
compared to traditional spectroscopy, photoacoustic spectroscopy can provide greater
sensitivity and selectivity for glucose measurements. Furthermore, photoacoustic
technology is more robust in comparison to Raman spectroscopy and optical polarimetry.
This technique is sensitive to chemical and physical interferences present within the
measurement sites. At this time, instrumentation is too expensive for home applications.

The research presented in this dissertation explores the analytical utility of NIR
spectroscopy. For this reason, the following sections provide more detail on NIR
spectroscopy and the chemometric methods used to correlate spectral features to analyte
ccentration.

**Near Infrared Spectroscopy**

In the 1800s, William Herschel discovered radiation beyond the visible red light,
however, it was not considered useful for spectroscopic applications until the Second
World War. Infrared (IR) light is electromagnetic radiation with longer wavelengths
than those of visible light and primarily corresponds to vibrational and rotational
transitions associated with chemical bonds within molecules. The IR spectral region of
the electromagnetic spectrum extends from 13300 to 10 cm\(^{-1}\) while the near-infrared
(NIR) covers the spectral range from 13300 to 4000 cm\(^{-1}\). Throughout the NIR spectrum,
the spectral absorptions are associated with the combinations and overtones of the
fundamental vibrations of C-H, O-H and N-H functional groups which are the principal
components of most biological molecules. Compared to the fundamental vibration
modes observed over mid-infrared frequencies, the NIR absorption bands are typically
one to three orders of magnitude weaker as well as broader and highly overlapping with
each other.

Quantitation is based on the concentration dependent absorption of NIR radiation
according to the well-known Beer-Lambert law, as described in Equation I-1:
\[ A = -\log \left( \frac{I}{I_0} \right) = \varepsilon \cdot b \cdot c \]  

Equation I-1

where \( A \) is the absorbance of a sample, \( I_0 \) and \( I \) are the intensities of the incident and transmitted light, respectively, \( \varepsilon \) is the molar absorptivity at a certain wavelength in \( \text{(mol/L)}^{-1}\text{mm}^{-1} \), \( b \) is the path length of a sample in mm, and \( c \) is molar concentration of the absorbing species in mol/L. This equation is specific for transmission measurements through non-turbid solutions, however it provides the general basis for all absorption phenomena, including those involved in both transmission through a scattering medium (human tissue) or diffuse reflectance measurements. Figure I-3 shows schematic diagrams for transmission and diffuse reflectance type measurements, which are the most commonly used methods for noninvasive glucose measurements with NIR spectroscopy. Although the Beer-Lambert relationship does not consider the combination of scattering and absorption, more complex systems and algorithms are available to model such measurements.\(^{44}\)

NIR spectroscopy offers several major features for quantitative measurements. First, electromagnetic radiation over NIR frequencies is nonionizing, which is particularly important for clinical diagnostics or bioprocess monitoring where damage to the sample must be avoided. Second, the selectivity afforded over NIR frequencies demands minimal or no sample preparation, so measurements can be performed without removing a sample for analysis. Third, absorptivities over NIR frequencies permit measurements at sample thicknesses over several millimeters, depending on the exact wavelengths involved in the measurement. In comparison, measurements over mid-IR frequencies are limited to 10’s to 100’s microns, which can be impractical for noninvasive clinical measurements. The development of NIR spectroscopy for a wide variety of analytical applications for the petroleum industry, agriculture industry, food analysis, clinical diagnostics,
Figure I-3. Schematic diagrams of transmission and diffuse reflectance NIR measurements.
chemical monitoring of bioprocesses, pharmaceutical industry and environmental chemical monitoring.\textsuperscript{45-80}

**Fourier Transform Infrared Spectrometers**

Fourier transform (FT) spectrometers were first developed in the early 1950’s and the first commercial units appeared on the market in the 1960’s. The advantages of FT spectroscopy resulted in the displacement of dispersive spectrometers by the 1980’s. The central feature of an FT system is the interferometer which permits high throughput measurements over a band of optical frequencies. Figure I-4 presents a schematic diagram of the Michelson interferometer which was invented in the 19\textsuperscript{th} century and is the most widely used interferometer for spectroscopic purposes.\textsuperscript{81} In this device, a beam splitter splits a collimated beam of incident radiation into two nearly equal paths toward a fixed mirror and a movable mirror. After reflection at these mirrors, the light is recombined at the beam splitter and then travels to the detection optics. The position of the moving mirror determines the optical path length difference between the two light paths. As the moving mirror is scanned, constructive and destructive interference is observed at the detector as this path difference passes multiples of the wavelengths involved. Once a full scan is completed, a Fast Fourier Transform (FFT) algorithm is used to convert the time-domain interferogram into a single-beam frequency-domain spectrum.

Features of this FT approach are well documented and include multiplex and throughput advantages. The multiplex advantage, also referred to as Fellgett’s advantage, recognizes that all frequencies of the incident radiation are measured simultaneously, thereby reducing the acquisition time in comparison to a dispersive spectrometer. The throughput advantage, referred to as Jaquinet’s gain, stems from the absence of a slit in the optical path, thereby permitting high radiant power to reach the detector and providing high signal-to-noise ratios.
Figure I-4. Schematic diagram of the Michelson interferometer.
Lastly, a HeNe laser is used as an internal wavelength reference for frequency precision as low as 0.01 cm\(^{-1}\), which permits effective signal averaging and enhanced signal-to-noise ratios.

**Multivariate Calibration Analysis**

The overlapping nature of NIR spectra for condensed phase samples and the sensitivity of these spectra to physical parameters demand the use of multivariate methods of analysis to achieve selectivity and robustness. Simple univariate methods based on absorbance values at single wavelengths are impractical for measurement of clinically relevant species in complex biological matrices. Powerful multivariate methods are available to enhance selectivity by basing measurements on an analysis of the full spectrum.

The objective of a multivariate calibration method is to establish a mathematical model that relates a chemical or physical property of interest to the spectral information encoded across multiple wavelengths. To achieve this aim, multivariate calibration methods follow a two-step procedure: first, spectroscopic measurements are collected from a set of standard samples. For these standard samples, the physical or chemical property of interest is pre-determined by an established independent reference assay. Second, this group of sample spectra and reference measurements is treated as the calibration data set for the purpose of establishing a mathematical model that effectively correlates the targeted property to features within the spectral data set. Once the mathematical model is constructed, it can be used to predict the modeled property for subsequent unknown samples.

A variety of multivariate calibration methods are available for the analysis of NIR spectral data. The simplest methods include classical least-squares (CLS) and inverse least-squares (ILS) regressions which are often referred as K and P matrix methods, respectively.\(^82\)\(^83\) More complicated and robust algorithms include multiple linear
regression (MLR), principal component analysis and regression (PCA and PCR), partial least squares (PLS) and net analyte signal (NAS). Each of these methods is described below.

Classical Least Squares

Classical least squares (CLS) regression, also known as direct least squares, models the spectral variances as a function of analyte concentration to the classical expression of the Beer-Lambert Law (See Equation I-1) where the instrument response is expressed as a function of the constituent’s concentration and the corresponding spectra as shown in Equation I-2:\(^{83-84}\)

\[ R = CS^T + E \quad \text{Equation I – 2} \]

where \( R \) is the vector of sample spectra, and \( C \) is the corresponding concentration matrix of analytes of interest, \( S \) contains the spectra of each individual constituent in \( R \) and is determined by regression, and \( E \) is the error matrix associated with the calibration. According to Equation I-3, the constituent spectra \( S \) can be estimated simply with an ordinary least squares pseudo-inversion of \( C \):

\[ \hat{S}^T = C^TR \quad \text{Equation I-3} \]

Equation I-4 is used to predict the concentration of analytes in a new sample from collected spectra and the pseudo-inversion of estimated constituent spectra \( S \) as defined above:

\[ \hat{C} = \hat{S}^R_{\text{unknown}} \quad \text{Equation I-4} \]
The concept of manipulating spectra and concentrations with CLS is straightforward and the constituent spectra are estimated through the regression. A major complication of this approach, however, is the need to have a component spectrum for each element of the sample that impacts the composite spectrum. Advanced CLS models such as augmented CLS (ACLS) could be utilized to compensate the disadvantage and reach a robust modeling within a complex matrix.

Inverse Least Squares

The ILS calibration model assumes that the concentration of sample could be predicted quantitatively from sample spectra. The basic model for ILS calibration with multiple components in C can be expressed in Equation I-5:

\[ C = RP + E \quad \text{Equation I-5} \]

where \( C \) is the matrix of reference concentrations and \( R \) is the matrix of spectral data. The vector \( P \) is the correlation between the spectral matrix \( R \) and the concentration matrix \( C \) while \( E \) is the matrix of concentration errors that are not included by the model. The vector \( P \) is also called the regression vector, which is designed to be selective to the analyte for modeling and orthogonal to all other components of spectral variation in the spectra matrix \( R \). The regression vector \( P \) can be estimated as Equation I-6:

\[ P = R^C \quad \text{Equation I-6} \]

where \( R^C \) is the pseudo-inverse of the spectral matrix \( R \). The modeling process for ILS fits the data to the model in the least squares sense by minimizing the total sum of squared concentration residuals. One major advantage of ILS calibration is its pure mathematical construction and no need of information for the constituent spectra. ILS
has an additional advantage over CLS in that it can build a calibration model efficiently over few frequencies that obey the Beer-Lambert Law and the nonlinearities existed in the spectra have less influence on the calibration model. Due to these advantages, many modern and advanced multivariate calibration techniques such as MLR, PCR, and PLS are developed on the basis of the ILS approach.

**Multiple Linear Regression**

Multiple linear regression (MLR), also called linear least square regression, is a statistic method used to model a linear relationship between one dependent variable and one or more independent variables. In this method, an ordinary least squares approach algorithm is utilized to compute the pseudo-inverse of spectra as shown in Equation I-7 below:

\[ R^+ = (R^T R)^{-1} R^T \]  

Equation I-7

This technique is very sensitive to the conditioning of the spectral matrix and suffers from its instinctive shortcomings since the pseudo-inverse utilized the inversion of \( R^T R \). First, collinearity existing in the spectral matrix \( R \) makes estimation of the true inversion of \( R^T R \) difficult and inaccurate. As mentioned before, all absorption bands are relatively broad in NIR spectra, spectral features are usually highly overlapped between components. The resulting collinearity in the spectral matrix \( R \) diminishes performance of the MLR approach, especially in samples composed of complex components. In addition, fewer independent spectra in the spectral matrix \( R \) than the number of spectral variables will result in inaccurate estimation of the regression vector. So it is required that the sample size should be larger than the number of variables in \( R \). In order to compensate for these limitations, latent variable algorithms, such as PCA and PLS, have been developed.
Principal Components Analysis

Principal component analysis (PCA) is a statistical method that characterizes covariance structure of variances in the spectral data set. Sample spectra are linear combinations of the original variations and the PCA provides an interpretation and a better understanding of the sources of these variations. Spectral variances are characterized as a set of orthogonal vectors typically referred to as principal components, eigenvectors, spectral loadings or loading vectors. PCA is a powerful method for reducing the dimensionality of the data matrix and eliminating noise. For these reasons, PCA is widely used for preliminary analysis of high-dimensional data prior to classification, cluster analysis or other multivariate calibration models.

In PCA, the spectral matrix can be decomposed as in Equation I-8:

\[
X = TP + E
\]

Equation I-8

where \(X\) is an \(m \times n\) matrix of spectral data which contain \(m\) spectra and each spectrum has \(n\) data points over the spectral range. \(T\) is an \(m \times p\) matrix containing all score values for the spectra, \(P\) is a \(p \times n\) matrix of the principal components and \(E\) is the noise matrix which cannot be explained by the optimal principal components. The principal components in the \(P\) matrix are independent and orthogonal to each other and sorted by the contribution of each principal component to the whole set of spectral variance. The first principal component represents the most significant spectral variance and the second one represents the second most significant variance. With an increase in sample complexity, more PCs are required to characterize the data set. Principal components can contain spectral variances associated with the sample components as well as physical variations, such as changes in temperature, and instrumental variations. Generally, noise is characterized by the higher order principal components.
The singular value decomposition (SVD) algorithm is commonly used to extract principal components from the spectral matrix $X$, as in Equation I-9:

$$X = USV$$

Equation I -9

where $U$ is an $n \times n$ orthogonal matrix containing eigenvectors of $XX^T$ and $V$ is a $p \times p$ orthogonal matrix which contains eigenvectors of $X^TX$. Correspondingly, $S$ is an $n \times p$ diagonal matrix which is called singular values of the matrix. The diagonal elements of the matrix $S$ are equal to the square root of the eigenvalues of $XX^T$. The eigenvalues quantify the contribution for each principal component, so the columns of $U$ and $V$, also called loadings, are usually ordered according to decreasing magnitude of the eigenvalues.

In building the PCA calibration model, usually the first $k$ principal components are selected as the number of latent variables or factors. The spectral loading matrix would be a $p \times k$ matrix as $V_k$ and the corresponding $n \times k$ score matrix $T_k$ could be calculated as indicated in Equation I-10:

$$T_k = U_kS_k$$

Equation I-10

PCA provides a powerful tool for characterizing the spectral information within a set of spectral data. Spectral features in each principal component can be utilized to locate the source of the chemical components in the matrix, even the variance from the physical and environmental variances as temperature could be estimated. The noise level in the principal components could be utilized to determine if all chemical information have been estimated. With a properly chosen of number of factors, the necessary information for concentration modeling can be included while principal components of interferences and noise can be excluded.
Partial Least Squares Regression

Partial least squares (PLS) was first developed by H. Wold in the 1960s.\cite{90,91} The goal of PLS is to estimate the regression coefficients in a linear regression model between spectral variance and concentration variance.\cite{90,91} To accomplish this, the PLS method simultaneously decomposes the spectral matrix $R$ and the concentration matrix $C$ into loadings and scores, and then establishes the correlation of covariance as presented in Equation I-11 and I-12. The latent variables from spectral matrix $R$ in PLS are developed as a linear combination of the spectral information to ensure maximum correlation with the loadings from the concentration matrix $C$ in order to enhance a subspace in the spectral matrix $R$ that is better to model concentration. The relevant equations are:

$$R = TP + E \quad \text{Equation I-11}$$
$$C = UQ + F \quad \text{Equation I-12}$$

where $T$ and $U$ are score matrices of $R$ and $C$ and, $P$ and $R$ are the respective loading matrices. $E$ and $F$ are the errors associated with the corresponding $R$ and $C$ matrices.

The connection between matrices $R$ and $C$ is constructed by a matrix of loading weight vectors $W$ to maximize correlation with the concentration matrix $C$.

Firstly, the normalized first weight vector $w_1$ is computed as follows:

$$w_1 = \frac{R^TC}{\|R^TC\|} \quad \text{Equation I-13}$$

Consequently, the regression vector related to the first factor can be estimated as:

$$\beta_1 = \frac{C^T(R^Tw_1)}{\|(Rw_1)^TRw_1\|} \quad \text{Equation I-14}$$
Subsequently, the algorithm repeats for additional factors by calculating $w_2$ and further regression vectors with the concentration and spectra residuals based on the first factor. For concentration predictions, the calibration vector matrix is used according to Equation I-15:

$$\hat{c} = XB$$  \hspace{1cm} \text{Equation I-15}

The standard error of calibration (SEC) and the standard error of prediction (SEP) are commonly used to evaluate performance of a PLS model. These parameters are obtained according to the following equations:

$$\text{SEC} = \sqrt{\frac{\sum (c_m - \hat{c}_m)^2}{n_m - k - 1}}$$  \hspace{1cm} \text{Equation I-16}

$$\text{SEP} = \sqrt{\frac{\sum (c_n - \hat{c}_n)^2}{n_n}}$$  \hspace{1cm} \text{Equation I-17}

where $c_m$ and $c_n$ are reference concentrations for the samples used in the calibration and prediction sets and $\hat{c}_m$ and $\hat{c}_n$ represent the concentrations predicted by the PLS model for these calibration and prediction samples. Also, $n_m$ and $n_n$ are the numbers of spectra in calibration and prediction sets and $k$ is the number of factors used in the optimized PLS model. In Equation I-16, the degree of freedom is $n_m-k-1$ because one degree of freedom is lost for mean centering procedure and $k$ degrees of freedom are lost for the $k$ latent variables.

Because the PLS model extracts the latent variables by building covariance between the spectral and concentration matrices, the PLS method efficiently minimizes the impact of interferences. For this reason, PLS methods represent popular tools for multivariate measurements in complex samples.
Net Analyte Signal

The concept and properties of the net analyte signal (NAS) have been introduced by A. Lorber. Compared to the empirical methods of PCA and PLS, the NAS approach focuses on information associated specifically with the targeted analyte. The NAS is defined as that portion of an analyte’s spectrum that is orthogonal to the background spectra, where the background spectrum contains all sample components with spectral features over the spectral range of interest. The NAS can be obtained by a vector projection procedure as illustrated schematically in Figure I-5. In this method, a spectrum of the pure analyte is decomposed into two parts by projecting this analyte pure component spectrum onto a spectral space defined by the background spectral matrix. One part is co-linear with the background matrix vector and the other is the component of the analyte’s pure component spectrum that is orthogonal to the background matrix. This orthogonal component of the analyte spectrum is the net analyte signal and represents the component of the analyte spectrum that is unique or selective relative to the background matrix. Mathematically, the NAS is independent of the spectral space of the co-constituents and provides a multivariate calibration method specifically derived from the analyte molecule of interest.

Equation I-18 presents the calculation of the orthogonal part of the pure component spectrum of the kth analyte:

\[
\text{NAS}_k = (I - A_k A_k^+) A_k
\]

Equation I-18

where NAS\(_k\) is the NAS of the kth analyte in the mixture sample and A\(_k\) represents the pure component spectrum of the kth analyte. A\(_k\) represents the matrix of spectra with all components in the mixture except the kth component while A\(_k^+\) is the Moore-Penrose pseudo inverse of A\(_k\). In the equation, I is the identity matrix that has the same dimension
Figure I-5. Illustration of NAS projection.
as $A_k A_k^*$. With the net analyte signal from spectrum projection procedure, the predicted concentration of the analyte with selected factor $k$ can be calculated as in Equation I-19:

$$\hat{c} = NAS_k X$$  \hspace{1cm} \text{Equation I-19}

Figures of merit can be used to characterize NAS calibration methods with respect to sensitivity, selectivity, and limit of detection (LOD) for the multivariate calibration models.\textsuperscript{93}

Hybrid Linear Analysis

Proper development of a NAS calibration demands a pure component spectrum for the analyte in question and a representative set of background spectra that incorporate spectral variations associated with all non-analyte components in the sample. In some real-world situations, it is not possible to collect blank background spectra. HLA was first introduced by A.J. Berger to overcome this limitation.\textsuperscript{94} In the HLA method, the spectral contribution of the analyte is subtracted from the sample spectrum. This subtraction requires knowledge of the analyte concentration associated with each sample spectrum. Once the spectral contribution of the analyte has been subtracted, the resulting spectrum can be used as a background spectrum for the purpose of computing the NAS. In other words, the analyte subtracted spectra can be used to characterize the background spectral matrix when blank background spectra are not available. The analyte subtracted background spectra are computed according to Equation I-20:

$$A_b = A - ka$$  \hspace{1cm} \text{Equation I-20}

where $A_b$ is the matrix of calculated background spectra, $A$ is the spectra matrix in which the concentration of analyte is known as in $a$, a row vector of analyte concentration and $k$
is the pure component spectra of unit analyte.

With the calculated background matrix and the pure component spectra of analyte, the normal NAS algorithm is used to develop the calibration vector for predicting analyte concentrations from sample spectra. The HLA algorithm uses the pure component spectra allowing separating the spectral contributions of the analyte and reserving the background information for modeling. Because the HLA algorithm constructs the calibration vector through a projection procedure instead of regression, it can be more intuitive and robust. The performance of HLA models depends on the accuracy of the analyte concentrations assigned for the samples used to generate the background spectra.

Digital Fourier Filtering

Digital filtering techniques are signal-processing tools that can be effective in removing confounding spectral features from spectral data sets. Both high and low frequency spectral variations can be removed by proper selection of the filter bandpass function. The mechanism of removing spectral variations by digital filtering is based on the assumption that spectral information of analyte can be separated from others based on differences in their underlying frequencies. For example, baseline variations are generally low frequency structures within the spectra while noise is typically characterized by high frequencies. Digital filtering can be used to remove both baseline variation and noise coupled in the NIR spectra.

The digital filtering technique used in this work is Fourier filtering where the FFT algorithm is used to decompose a spectrum into its underlying frequency components. These frequencies are distinguished from optical frequencies by referring to them as digital frequencies and recognizing that they correspond to the shapes of the spectral features within the sample spectrum. The resulting digital frequency spectrum can be multiplied by an optimized Gaussian shape function that passes selected digital
frequencies while discriminating against high frequency noise and low frequency baseline variations. An inverse Fourier transform is used to return to the original data domain and the resulting filtered spectrum is available for further processing. With the digital Fourier filter, the signal-to-noise ratio of the spectra can be enhanced and the baseline variations, from physical interference such as temperature, can be reduced in magnitude. This pre-processing tool can improve the performance of the multivariate calibration model and make a more robust prediction of analyte concentration.

Overview of Thesis

In this dissertation, NIR spectroscopy is developed for the measurement of glucose in various matrices, from simple buffer solutions to whole blood and fermentation broth. During the research project, the short term and long term performance of PLS and NAS calibration models were evaluated and the means to improve robustness over time determined.

In Chapter II, digital Fourier filter is used in combination with PLS to achieve a temperature insensitive glucose measurement in bovine blood ultrafiltrate. Overall baseline variations caused by changes in sample temperature are minimized through digital Fourier filtering with an optimized Gaussian shape filter. With filtering, prediction errors for PLS calibration models are similar with and without variations in the sample temperature.

Chapter III presents a method to simulate variations in the concentrations of glucose in ISF. A stable and accurate platform is described as a means to mimic changes in the concentration of glucose within ISF for an in vitro evaluation of a future microsensor technology. Both PLS and NAS models are used to evaluate this configuration and the resulting SEP values demonstrate excellent prediction ability for in vitro continuous measurements for short term and long term experiments. Different
glucose concentration profiles demonstrate no significant influence on the performance of either model.

The ability to track both glucose and lactate during fermentation is illustrated in Chapter IV. NAS calibration models are used for each analyte. For glucose, the NAS model shows similar performance as a PLS model for short term glucose concentration predictions and NAS model achieved less off-set or bias and more resistance to systematic error than PLS models for long term performance. For lactate measurements, the spectral information for glucose was linearly subtracted from spectra to achieve effective background spectra for the NAS model development. This HLA approach provided promising SEP values for future industrial application.

Chapter V demonstrates the feasibility of measuring the concentration of glucose in whole blood. Measurements are collected at different hematocrit levels and calibration models are generated from both single-beam and absorption spectra. SEP values of 0.8-0.9 mM were achieved with PLS models based on absorbance spectra with optimized spectral range. For single-beam spectra, a multiplicative scatter correction pre-processing step eliminates the impact of scattering created by different hematocrit levels and improves analytical performance of the PLS model. This study demonstrates the potential of accurate glucose concentration measurements in a whole blood matrix.

The final chapter provides an assessment of these projects and offers future directions. Major efforts will focus on further optimizing and developing both hardware set-up and chemometrics algorithms for mature glucose monitoring systems.
CHAPTER II
TEMPERATURE-INSENSITIVE NEAR-INFRARED
SPECTROSCOPIC MEASUREMENT OF GLUCOSE IN BOVINE
BLOOD ULTRAFILTRATE

Introduction

Water is both a major absorber of NIR radiation and a major component within biological samples. For these reasons, it is critical to have a firm understanding of the NIR spectroscopy of water as a precursor to noninvasive measurements in biological samples, such as the measurement of glucose in blood, ISF or fermentation broth. As has been detailed elsewhere, the major absorption bands of water are centered at approximately 6900, 5200 and 3800 cm\(^{-1}\) and they divide the NIR spectrum into three regions: the combination region (4800-4200 cm\(^{-1}\)), the first overtone region (6700-5400 cm\(^{-1}\)), and the short wavelength NIR region (12000-7100 cm\(^{-1}\)). Hydrogen bonding within the structure of water renders the exact position and width of these absorption bands sensitive to the temperature of the sample. Absorptivities for water have been characterized at different temperatures. 

Two-dimensional (2D) correlation spectroscopy has been used to understand the temperature sensitivity of the NIR spectrum for water. V.H. Segthan et al. found that more than 99% of the spectral variation for water occurs around the wavelengths at 7082 and 6707 cm\(^{-1}\) which correspond to the major weaker and stronger hydrogen bonds of water, respectively. Absorption at 6954 cm\(^{-1}\) was found to be constant and provided an isosbestic point across temperatures. Similar findings of this isosbestic point are reported by others when using conventional NIR and Raman spectroscopic methods. S.H. Chung took advantage of this strong temperature sensitivity of the water NIR spectrum to develop a non-invasive method for measuring the absolute temperature of thick samples of tissue. Chung’s method used this isosbestic point as a standard point in order to
account for non-temperature variations and improve accuracy of the temperature measurement.  

For biological samples, the temperature sensitivity of the underlying water spectrum demands that the temperature of the solution be tightly controlled for repeatable spectroscopic results. For NIR analysis, baseline variations in absorbance spectra caused by temperature changes can be several orders of magnitude larger than the intensities of the analyte absorption bands. Others have shown that a 1 °C change in the sample temperature creates changes in the sample’s absorbance spectrum that is 2 to 8 times larger than spectral changes caused by a glucose concentration of 100 mg/dL.

One solution is to thermostat the sample, although the current trend to miniaturize clinical instrumentation can make this a challenge. An alternative strategy is to develop a spectral processing method to remove the spectral variance created by changes in sample temperature.

A number of methods have been reported for creating temperature-insensitive glucose calibration models. In one approach, a PLS model is generated from an array of spectra composed from samples where temperatures and glucose concentrations were varied widely and, most importantly, independently. Predictions from the resulting PLS model were insensitive to temperature, which means the calibration vector for glucose concentration predictions was orthogonal to the spectral variance created by changes in temperature. X. Zhang et al. used PLS2 to eliminate the influence of temperature on non-invasive blood glucose measurements with the GlucoStats System. In this case, the error of prediction was significantly lower through temperature variance when finger temperature was considered in the modeling compare to the result of PLS1 models without this temperature information. Furthermore, K. Hazen et al. applied digital Fourier filtering to eliminate the effect of temperature variance on NIR spectra in buffer solutions with glucose. This work focused on spectra collected over the first overtone region of the NIR spectrum and dealt with a simple matrix of glucose in buffer.
More complex temperature effects are anticipated for complex biological samples, such as ISF, blood, and fermentation broth, depending on the temperature sensitivity of all matrix components. In this chapter, digital Fourier filtering is coupled with PLS modeling of NIR spectra collected over the combination spectral range to provide a temperature-insensitive glucose measurement in the matrix of bovine blood ultrafiltrate. Over a temperature range of 25 to 40 °C, this strategy generates a calibration model that gives essentially identical standard errors for spectra with no temperature variation and spectra with temperature variation. The validity of this approach is further verified by applying it to a series of NIR spectra collected with an instrument resolution ranging from 2 to 32 cm$^{-1}$.

**Experimental Section**

**Sample Preparation**

Bovine blood was collected from a single animal from a local abattoir and mixed with 2 g/L EDTA as an anti-coagulant. After centrifugation, the bovine plasma was removed and treated with an ultrafiltration unit contained a MiniKros hollow fiber tangential flow module. The molecular-weight cutoff (MWCO) was 50 kDa (Spectrum Laboratories, Inc, Rancho Dominguez, CA) and a Materflex peristaltic pump (Cole-Parmer Instrument Co., Vernon Hills, IL) was used to circulate the plasma through the ultrafiltration cartridge. Bovine ultrafiltrate is a clear mixture without scattering bodies, such as red blood cells and no large proteins. The ultrafiltrate samples can serve as a model for ISF, which is also plasma filtered through the capillary bed. After sterilizing the filtered plasma with a Stericup filter unit (Millipore Co, Billerica, MA), 0.483 g/L 5-fluorouracil was added as a preservative. The endogenous glucose, urea and triacetin concentrations in bovine ultrafiltrate were analyzed with an YSI model 2300 STA plus (Yellow Springs Instrument Co., Yellow Springs, OH) and ACE analyzer (Alfa Wassemann, Inc. West Caldwell, NJ). Overall, 80 samples were created with random
concentrations of glucose, urea, and triacetin ranging between 3 and 30 mM. Care was taken to avoid correlations between the concentrations of these analytes and the resulting correlation coefficients ($R^2$) were 0.0031, 0.0476 and 0.0086, (See Equation II-1) for the glucose-urea, glucose-triacetin, and urea-triacetin correlations, respectively. The disodium salt of EDTA, 5-fluorouracil, glucose, urea, and triacetin were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO). The reference glucose concentration for each sample was calculated as the sum of the original glucose concentration measured from the bovine ultrafiltrate plus the glucose added for each sample. The final concentrations were verified by analysis with the YSI analyzer just before spectra were collected.

$$R = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

Equation II -1

where $n$ is the number of data pairs, $X_i$ and $Y_i$ are the values for components x and y, $\bar{X}$ and $\bar{Y}$ are the mean values for components x and y, respectively.

**Instrumentation and Apparatus**

Triplicate NIR single-beam spectra were collected with a Nicolet 670 Nexus Fourier transform (FT) spectrometer (Nicolet Analytical Instruments, Madison, WI). This spectrometer was equipped with a 20 Watt tungsten-halogen light source, with a calcium fluoride beam splitter and a cryogenically cooled indium antimonide (InSb) detector. All NIR spectra were restricted to 5000-4000 cm$^{-1}$ by a K-band multilayer optical interference filter (Barr and Associates, Westford, MA) which was positioned before the sample cells. The aperture and instrument gain were maintained at 70 and 8, respectively, for the collection of spectra from filtrate and buffer samples. Samples were placed in a borosilicate capillary tube (Warner Instrument Co., Hamden, CT) with a path length of
0.94 mm which was coupled to a thermoelectric cooler (TEC) as shown in Figure II-1. Sample temperatures were controlled and monitored with a LFI3751 Digital Temperature Control Instrument (Wavelength Electronics Inc., Bozeman, MT) which can reach a temperature stability of ± 0.001°C.

Spectral Collection

All samples were injected and maintained in the sample cell during the collection of three consecutive NIR single-beam spectra. For each sample, spectra were collected at temperatures of 25, 30, 35, 37, and 40 °C (± 0.2 °C) in a random order. Single-beam spectra were collected as 32 co-added, 16384 point double-sided interferograms. The OMNIC software (Version 5.1B, Thermo-Nicolet) was utilized to apply triangular apodization, Mertz phase correction and one level of zero-filling. This same software package was used to convert the interferograms into a single-beam spectrum. Single-beam spectra of buffer were collected after every fourth sample as reference spectra for absorbance calculations. Absorbance spectra were calculated by taking the negative logarithm of the ratio of each single-beam spectrum to the most recently collected corresponding background spectrum. All single-beam spectra were collected with a resolution of 2 cm⁻¹.

Data Processing

The quality of the raw single-beam spectra was evaluated to ensure the performance of the multivariate calibration models. The root mean square (RMS) noise of 100% lines was calculated to characterize spectral quality. Each 100% line was obtained by calculating the negative logarithm of the ratio of two single beam spectra collected from the same sample. The ratioed values were fitted to a second-order polynomial over the 4500-4600 cm⁻¹ spectral range. The RMS over this fitted range was obtained from Equation II-2. Overall, the average RMS noise across all samples was 16.0 µAU (micro absorbance units).
Figure II-1. Sampling system coupled with a thermoelectric cooler (A) and temperature controller (B).
\[ \text{RMS} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \]  

Equation II - 2

where \( n \) is the number of wavenumbers in the fitted region, \( y_i \) is the absorbance value of the 100 % line measured at the wavenumber \( i \), and \( \hat{y}_i \) is the fitted absorbance value from the second-order polynomial at wavenumber \( i \).

Calibration models for glucose were established from NIR spectra collected from the different mixtures prepared in the bovine plasma ultrafiltrate matrix. All models were developed from the partial PLS algorithm based on subsets of calibration and prediction data. In total, 240 spectra from 80 samples were randomly selected into calibration (80%, 192 spectra from 64 samples) and prediction (20%, 48 spectra from 16 samples) sets. All models were prepared from absorbance spectra unless states otherwise. Two types of absorbance spectra were created. One was based on the ratio of the sample single-beam spectrum and the buffer single-beam spectrum collected closest in time and at the same temperature. The other type of absorbance spectra consisted of all sample single-beam spectra ratioed to the mean buffer spectrum obtained at 35 °C. In the first case, the temperature of the buffer spectra matched the temperature of the sample spectra. In the second case, the same buffer spectrum was used throughout and did not match in terms of solution temperature when the sample temperature was not 35 °C. For each case described above (temperature matched and temperature mismatched), models were generated at each temperature individually and combining spectra collected at all temperatures. Data at all temperatures were analyzed with and without digital Fourier filtering prior to the PLS algorithm. All calculations were performed in MATLAB R2007a (The MathWorks, Inc., Natick, MA). SEC and SEP values were calculated as noted in Chapter 1 (See Equations I-16 and I-17).

To investigate the effect of spectral resolution, the reprocess function of OMNIC was applied to reprocess each spectrum to a poorer resolution and the corresponding
point spacing. One level of zero filling was maintained in all data processing. Single-beam spectra were created with resolutions of 4, 8, 16, and 32 cm\(^{-1}\) and point spacings of 1, 2, 4 and 8 cm\(^{-1}\). Under each resolution, absorbance spectra were calculated and PLS models were created with and without digital Fourier filtering.

**Results and Discussion**

**Temperature and Absorption Band**

Figure II-2A shows water spectra collected over the combination region of the NIR spectrum at temperatures of 20 to 40 °C. The combination region corresponds to the water absorption minimum between the absorption bands centered at 5200 and 3600 cm\(^{-1}\). As has been reported before, a blue shift is observed in the water minimum spectrum with an increase in temperature.\(^{100-101}\) Figure II-2B presents the corresponding synchronous map from a 2D correlation analysis of these absorbance spectra. One auto-peak at 4972 cm\(^{-1}\) is evident and this peak reveals that the spectral features at this position vary in phase with each other as temperature varies. The two circle-edges at the opposite corners are negative cross-peaks and their signs indicate that the changes are correlated in opposite trends. These negative cross-peaks were not complete and the other autopeak is not shown in Figure II-2B because the signal was attenuated at 4000 cm\(^{-1}\) with the K-band filter. This autopeak is positioned around 3900 cm\(^{-1}\) can be easily observed in reference.\(^{117}\)

Comparing the magnitude of the spectral shifts presented in Figure II-2B and the magnitude of the absorption bands presented in Figure I-2 for 100 mM glucose, it is apparent that the spectral variations caused by temperature are significant compared to absorptions for physiological concentrations of glucose (3-10 mM). It is also noteworthy that the positions of the glucose absorption bands over the combination spectral range (\(i.e.,\) those centered at 4750, 4400, and 4300 cm\(^{-1}\)) are temperature-insensitive over the range of 10.5 to 40.4 °C.\(^{118}\)
Figure II-2. Absorption spectra of distilled water at temperature of 20, 25, 30, 35, 40 °C (A); Synchronous variable-variable two-dimensional correlation spectra (B), generated from the absorption spectra shown in (A).
The impact of a temperature mismatch can be seen from the data plotted in Figure II-3A. In this plot, absorbance spectra are presented for solutions at all temperatures with the reference spectrum set at 35 °C. These spectra can be easily distinguished into five groups, one for each of the sample temperatures: 25, 30, 35, 37, and 40 °C. When the temperature of the sample is greater than 35 °C, absorbances are negative at the higher frequencies and positive at the lower frequencies. The opposite trend is observed when the temperature of the sample is below 35 °C. Such baseline deviations are consistent with the water absorbance bands shifting to higher frequencies at higher temperature. The magnitude of the variation is greatest at the higher frequencies, in accord with the greater sensitivity to temperature displayed in Figure II-3A.

**Digital Fourier Filtering**

Digital Fourier filtering is a method for removing spectral variance based on the assumption that spectral features derived from analyte and nonanalyte sources can be distinguished by differences in shapes of these spectral features. In the digital frequency domain, broadly varying components, such as the baseline curvature displayed in Figure II-3A, correspond to low digital frequencies and noise is located at high digital frequencies.

The key is to find the bandpass function that effectively blocks both the low frequency baseline variations and high frequency noise, while passing digital frequencies that correspond to the intermediate shapes of the analyte absorption features.

In the procedure followed here, the raw absorbance spectra were transformed to the digital frequency domain by applying a FFT to the spectrum. The digital frequency axis was scaled to a linear standard from 0.0 to 0.5 \( f \). Figure II-4A shows the resulting digital frequency display for the spectra plotted in Figure II-3A. In this digital frequency domain, each spectrum has a large signal at low digital frequencies, corresponding to baseline variations. Figure II-4A also shows a Gaussian-shape bandpass function used to
Figure II-3. Temperature mismatch absorbance spectra calculated with reference spectra collected at 35 °C (A); Concentration correlation plot of the predicted glucose concentrations of PLS model (B), based on the temperature mismatch absorbance spectra shown in (A).
reject both low and high digital frequency information. The complex product is returned to the original spectral domain with an inverse FFT. Ideally, assuming different spectral features have different digital frequency components, the filter can selectively pass glucose information and block unwanted spectral features. The Gaussian function of the bandpass filter is described by its mean position and standard deviation width along the digital frequency axis. Optimization of the mean position and the standard deviation width of the Gaussian function was performed by a standard grid search. Values for the mean position and standard deviation width were varied systematically. For each combination of mean and standard deviation, a PLS model was generated for glucose and the performance of this model was judged. Model performance was evaluated by splitting the sample spectra into five temperature groups with 48 spectra each. Five different PLS models were developed by sequentially leaving out one group at a time in a cross validation manner. The average SEC and SEP values obtained for the five models were used to judge model performance. These average values are denoted as the cross-validation SEC (cv-SEC) and cross validation SEP (cv-SEP). The number of latent variables used for each model was selected on the basis of the cv-SEP values of different latent variables with an F-test. Initially, calibration models were built with latent variables from 1 to a certain number (depends on the complexity of the matrix). With the corresponding cv-SEC and cv-SEP obtained, if the minimal number of latent variables for which the cv-SEP value is not significantly higher then the next minimal cv-SEP value (F-test) is selected.

Results for the Gaussian filter optimization are presented in Figure II-4B as a surface response plot based on 1/cv-SEP as a function of the mean and standard deviation of the Gaussian function. Figure II-4B represents a typical surface map and this plot corresponds to a PLS model based on the use of six latent variables with temperature mismatched spectra. An optimum peak is evident in this surface plot, corresponding to a minimum cv-SEP and the best model performance.
Figure II-4. Absorbance spectra in frequency domain (A), and optimized Gaussian shape digital filter (solid blue line). Response surface map for optimization of mean and standard deviation of digital filter based on 1/SEP (B).
In this case, the top PLS model was achieved with a Gaussian filter defined by a mean position of 0.0097 \( f \) and a standard deviation width of 0.0037 \( f \) and this is the Gaussian curve plotted in Figure II-4A.

**PLS Calibration Models**

PLS calibration models were created for the measurement of glucose in the bovine plasma ultrafiltrate matrix. Models were generated from two types of spectra, one for the temperature matched spectra and one for the temperature mismatched spectra. For each case, temperature matched and temperature mismatched spectra, models were created with and without digital Fourier filtering prior to running the PLS algorithm. In all cases, sample temperatures varied from 25 – 40 °C, the spectral range was held constant at 4800–4200 cm\(^{-1}\), and the cross validation method described above (leaving 20% out) was used. The optimum number of latent variables, or factors, was determined for each model by comparing cv-SEP values for a series of factors and selecting the number of factors that gives the lowest significantly different cv-SEP using the F-test to compare sequential cv-SEP values at the 95% confidence level.

The impact of temperature mismatch is illustrated in Figure II-5 which plots SEP values from PLS models based on mismatched spectra. Two calculations were performed in this experiment. First, five independent PLS calibration models were constructed for glucose from temperature mismatched spectra of five different temperature groups and then the SEP values were calculated for the left-out set of temperature mismatched spectra within the same temperature group. The corresponding SEP values are plotted as light orange bars and the SEP values are consistent across the temperatures and average 0.39 mM. Second, the PLS model prepared from the temperature matched data group at 35 °C, was used to predict the concentrations of glucose from the temperature mismatched spectra for the left-out temperature groups. Again, the observed SEP values are plotted in Figure II-5 for comparison purposes.
These dark orange bars illustrate that SEP values are much higher for the temperature mismatched spectra and this mismatch increases for temperatures further from 35 °C. As expected, the lowest SEP corresponds to the 35 °C data set where the temperatures are matched.

In a separate experiment, two pairs of PLS calibration models were created by using all spectra collected at 35 °C. The first pair focused solely on temperature matched spectra and PLS calibration models were generated with and without digital Fourier filtering prior to the PLS calculation. In this first case, the calibration process used temperature matched spectra collected at 35 °C and the predictions were made on temperature matched spectra collected at the remaining temperatures. For the second pair of calibration models, all calculations focuses on temperature mismatched spectra and the PLS models were again generated with and without digital Fourier filtering prior to the PLS calculation. Each of these 35°C PLS calibration models, predictions were performed on unfiltered or filtered spectra consistent with the use of digital filtering to build the calibration model. In other words, if the digital Fourier filter was used as a preprocessing step in generating the calibration model, then this same digital filtering step was used for the predictions. Likewise, if no filtering was used as a preprocessing step for the calibration model, then no filtering was used for predictions.

Table II-1 summarizes the results obtained by using 35 °C spectra to build calibration models and then predicting glucose concentrations from spectra collected at different temperatures. As noted in the table, the optimum mean position and standard deviation width of the Gaussian filter are slightly different when predicting glucose concentrations from temperature matched versus temperature mismatched data. In addition, two fewer latent variables (factors) are required in both cases where digital filtering is used prior to the PLS calculation. The need for fewer factors is expected when using the filtered spectra because the filtering step removes spectral variance that must be modeled in the unfiltered spectra.
For the two calibration models without digital filtering, the SEC values are similar and are slightly lower compared to the SEC values for models when digital filtering is used as a preprocessing step. When the digital filtering excludes the low frequency information related to temperature variance, it can also eliminate spectral variance incorporated in calibration models. Thus, the SEC values of PLS models based on spectra after digital filtering increased but SEC values have similar values as those without digital filtering.

The most interesting point, however, is the comparison of SEC and SEP values with and without digital filtering. Regardless if temperature matched or temperature mismatched spectra are used, the SEP is much higher than the SEC for models without digital filtering. The SEP/SEC ratio is 3.4 and 6.2 for the temperature matched and mismatched spectra, respectively. On the other hand, the SEP/SEC ratio drops to only 1.2 and 1.1 for the matched and mismatched spectra when digital filtering is used. Digital filtering provides a 3 to 6 fold improvement in analytical performance.

Inspection of the concentration correlation plots for the temperature mismatched spectra is informative. Figure II-3A illustrates results for the 35 °C model with mismatched temperature spectra and no digital filtering. The prediction of glucose concentrations is offset according to temperature. In comparison, Figure II-6B shows results for the same spectra but with digital filtering. The offset behavior is removed as expected by filtering out the baseline variations created by sample temperature. A comparison of the filtered spectra in Figure II-6A and unfiltered spectra in Figure II-3A illustrates the effectiveness of the filtering step in removing temperature-dependent systematic spectral variance.

**PCA Analysis**

A PCA was performed on the temperature mismatched absorbance spectra before and after application of a digital filter defined by a mean position of 0.0097 $f$ and
Figure II-5. Errors in predictions from PLS models generated with single temperature spectra and combined temperature without digital Fourier filtering.
Table II-1. Error summary from the PLS models with and without digital Fourier filter.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Temperature matched</th>
<th>Temperature mismatched</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W/o digital filter</td>
<td>With digital filter</td>
</tr>
<tr>
<td></td>
<td>W/o digital filter</td>
<td>With digital filter</td>
</tr>
<tr>
<td>Optimized Mean</td>
<td>-</td>
<td>0.0097</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0.0091</td>
</tr>
<tr>
<td>Optimized Standard Deviation</td>
<td>-</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0.0019</td>
</tr>
<tr>
<td>Spectral Range (cm⁻¹)</td>
<td>4800-4200</td>
<td>4800-4200</td>
</tr>
<tr>
<td></td>
<td>4800-4200</td>
<td>4800-4200</td>
</tr>
<tr>
<td>Number of Factors</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>cv-SEC (mM)</td>
<td>0.23</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>0.28</td>
<td>0.36</td>
</tr>
<tr>
<td>cv-SEP (mM)</td>
<td>0.78</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>1.75</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Figure II-6. Temperature mismatch absorbance spectra after digital Fourier filtering (A); Concentration correlation plot of the predicted glucose concentrations of PLS model (B), based on absorbance spectra shown in (A).
standard deviation width of 0.0037 \( f \). Score plots are presented in Figure II-7 for these spectra without filtering (A) and with filtering (B). Without filtering, the principal group is by temperature, as is evident by the clustering of the data by temperature along the first principal component axis. With filtering, on the other hand, the magnitude of variance along the first principal component axis is much reduced and the clustering by temperature is completely removed.

Inspection of the shape of the first principal component vector (loading vector) also confirms the effectiveness of this digital filtering step. Figure II-8 presents the first loading vectors for the PCA of the temperature mismatched spectral data without and with digital Fourier filtering. Without filtering, the shape of this factor resembles the broad sloping baseline observed in Figure II-3A. A much different spectral shape is observed for the filtered spectral data. Such derivative-looking features are common for filtered spectra that use a Gaussian shaped filtering function and are associated with combination spectral absorptions associated with C-H, O-H, and N-H bonds.

Effect of Spectral Resolution

Spectral resolution of an FT-NIR spectrum impacts the digital filtering process by altering both location and shape of the FFT processed spectrum in the digital-frequency domain. Spectral resolution is also a critical design parameter in developing instrumentation configured for spectroscopic sensing and noninvasive analyses.\(^{121}\) In general, lower spectral resolution designs are preferred for \textit{in situ} sensing applications where operational durability is a primary concern. For this reason, it is important to understand the effect of spectral resolution on the utility of digital Fourier filtering in reducing the impact of sample temperature.

Table II-2 summarizes results from a series of PLS calibration models generated with temperature mismatched absorbance spectra at different spectral resolutions.
Figure II-7. Score plots from PCA models without (A) and with (B) digital Fourier filtering defined by a mean position of 0.0097 $f$ and standard deviation width of 0.0037 $f$. 
Figure II-8. First loading of PCA models with (red) and without (blue) digital filtering.
Results are presented for models generated without and with filtering. These results indicate that the optimized number of factors is the same regardless of resolution. In addition, the SEC and SEP values are similar across resolutions (2-32 cm\(^{-1}\)) for data without and for data with Fourier filtering. Although the optimized filter parameters differ for each resolution setting, the filtering process is effective in reducing the adverse impact of changes in solution temperature.

The optimized mean position and standard deviation width of the Gaussian function vary in a systematic manner for different resolutions. For spectra with higher resolution, the digital-frequency domain spectra would have broader frequency components. However, after scaling the frequency axis from 0 to 0.5 \(f\), the frequency components appear more compressed. When the resolution of the original (pre-filtered) spectrum decreases from 2 to 4 cm\(^{-1}\), as is the case for the plots in Figures II-9A and B, it appears that the digital frequency axis is doubled. For the optimized Gaussian shape filter, the optimized parameters must double to eliminate similar information as baseline variance and remaining similar analyte information for multivariate calibration models. Indeed, the plots in Figure II-10 illustrate the linear relationship between the optimum Gaussian parameters and spectral resolution. Basically, magnitude of the optimized parameter doubles for every power of two increases in the resolution setting.

Conclusions

A strategy of coupling digital Fourier filtering with PLS is realized and found to be effective in producing a multivariate calibration model that is insensitive to sample temperature for the quantification of glucose in bovine blood ultrafiltrate. The adverse effect of changes in the temperature of the sample can be reduced by using a Gaussian-shaped digital Fourier filtering step prior to the PLS calculation. The resulting filter
Table II-2. Results from the PLS models based on temperature mismatch absorbance spectra.

<table>
<thead>
<tr>
<th>Digital Filter</th>
<th>Resolution (cm(^{-1}))</th>
<th>Optimized Mean</th>
<th>Optimized Standard Deviation</th>
<th>Factor No.</th>
<th>SEC (mM)</th>
<th>SEP (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without</td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>0.28</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.28</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>0.25</td>
<td>27.8</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>0.26</td>
<td>26.8</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>0.27</td>
<td>24.8</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>0.27</td>
<td>22.6</td>
</tr>
<tr>
<td>With</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>0.36</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0091</td>
<td>0.0019</td>
<td></td>
<td>0.36</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0197</td>
<td>0.0039</td>
<td>6</td>
<td>0.38</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.0385</td>
<td>0.0083</td>
<td>6</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.0799</td>
<td>0.0159</td>
<td>6</td>
<td>0.38</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>0.1561</td>
<td>0.0302</td>
<td>6</td>
<td>0.38</td>
<td>0.43</td>
</tr>
</tbody>
</table>
effectively passes glucose spectral information while attenuating spectral variance associated with sample temperature. Calibration models from temperature-dependent spectra are developed to achieve SECs and SEPs in the same range as temperature-independent models. This methodology supports the feasibility of measuring glucose and others under clinical conditions when temperature control of the sample is not practical.

PLS models with digital filtering of absorbance spectra with different resolutions from 2 to 32 cm\(^{-1}\) are used to illustrate the effectiveness of the digital Fourier filtering step at different spectral resolutions. PLS models generated from spectra with different resolutions achieve the same level of performance when the optimized mean position and standard deviation of the Gaussian shape digital filter are used.
Figure II-9. Frequency domain spectra (left y-axis) and optimized Gaussian shape digital filter (right y-axis) under resolution 2 cm$^{-1}$ (A) and 4 cm$^{-1}$ (B).
Figure II-10. Correlation plots of resolution and optimized parameters for digital filtering for PLS models with absorbance spectra of temperature match (A) and mismatch (B).
CHAPTER III
CREATION AND MEASUREMENT OF IN VITRO GLUCOSE TRANSIENTS FOR MICROSENSOR DEVELOPMENT

Introduction

The ability to monitor glucose continuously is critical for the development of an artificial pancreas. Ideally, a continuous glucose monitor could feed information to an insulin delivery system to realize euglycemia and provide a mechanical cure for type I diabetes. Currently, commercially available implantable biosensors can be used to track glucose concentrations, however these FDA approved devices are only permitted to follow trends and are specifically not approved to guide treatment for either hypo- or hyper-glycemia. Problems with biofouling and poor calibration stability limit sensor accuracy and demand frequent in situ calibration. Efforts are underway to improve the analytical robustness of these implantable devices with the ultimate aim to achieve high sensitivity, selectivity and measurement accuracy over the entire physiological concentration range for glucose.

An alternative technology, still in the early stages of development, involves implantation of a miniaturized NIR spectrometer that is capable of measuring glucose in the ISF located in the subcutaneous tissue. The concept is to flow ISF through a measurement chamber where a NIR spectrum is collected and used to determine the concentration of glucose. By continuously pumping fresh ISF through the measurement chamber, real-time glucose readings would be possible and these reading could be utilized to drive a feedback-controlled insulin delivery system. An attractive feature of this approach is the relative simplicity of the ISF matrix compared to blood or whole tissue. Early prototypes of a miniaturized NIR spectrometer system (or microsensor) and the corresponding optoelectronics have been described elsewhere.
A clinically acceptable microsensor must be small for implantation and rugged enough to operate within the human body for at least a year while providing accurate information. A lack of chemical reactions during its operation and the fact that no chemical reagents must be stored for operation are major advantages of the NIR spectroscopic approach. A schematic diagram and photograph of an early microsensor prototype are presented in Figure III-1. As indicated in this figure, a thermoelectrically cooled GaInAsSb LED supplies incident radiation over the 4600-4200 cm\(^{-1}\) spectral range (2.18-2.38 \(\mu\)m). This light passes through a 1 mm thick sample of ISF and the transmitted light is focused through a linear variable filter onto a 32 element GaInAsSb detector array. Multivariate methods can be used to provide concentration information from the resulting spectrum.

Ultrafiltration (UF) can be utilized to collect a sample of the ISF and guide it past the sensing region of the microsensor unit. UF is a membrane-based sampling technique where a sealed loop of a semi-permeable membrane is connected to a negative pressure. As illustrated schematically in Figure III-2, pores within the semi-permeable membrane allow water, salts and low molecular weight molecules to pass into the inner volume of the probe. Proteins and other large molecules do not pass through the membrane and are excluded from the sampled fluid.

Characterization of any microsensor prototype demands an experimental setup that can be used to create transitory changes in the concentration of glucose within a matrix similar to ISF. The ability to generate transients is critical in order to evaluate the sensor’s ability to follow such transients accurately. Such an experimental system is designed and characterized in this chapter. The features of this system are evaluated by using a temperature controlled capillary flow-through cell located within the sample compartment of an FT-NIR spectrometer. Besides following glucose transients, the experiment setup is used to explore long-term and short-term stability of PLS and NAS calibration models for glucose.
Figure III-3 shows a schematic representation of the automated sample generation system that was used to create and measure glucose transients. As indicated, a computer was used to control a series of four peristaltic pumps, where each pump was used to pump a specific volume of a standard solution into a mixing chamber. By calibrating the flow rate for each pump, the precise volume of each standard can be known, thereby enabling calculation of the final concentrations of each analyte in the mixture. As indicated in this figure, pumps were used to mix programmed volumes of blank buffer, glucose, lactate and urea. The final solution was flowed past an UF probe before going to waste. A syringe pump was used to create a negative pressure within the UF probe and this negative pressure extracted a portion of the sample through the UF membrane and into the inner volume of the UF probe. Upon collection, the UF-sample flew through a temperature controlled flow-through capillary located within the sample compartment of the FT-NIR spectrometer. As documented below, different transient profiles could be generated by adjusting the computer controller.

**Experimental Section**

**Sample Preparation**

Standard pH 6.86 phosphate buffer solution was prepared by dissolving 3.40 g of potassium dihydrogen phosphate and 3.55 g of disodium hydrogen phosphate into 1 L of deionized water. Glucose, urea and lactate stock solutions were prepared with concentrations of 60 mM. 5-fluorouracil was added to each standard solution at a concentration of 0.2 g/L as an antimicrobial agent. Potassium dihydrogen phosphate and disodium hydrogen phosphate for buffer solution were purchased from Thermo Fisher Scientific Inc. (Fairlawn, NJ) and glucose, urea, lactate, and 5-fluorouracil were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO). Deionized water for all aqueous solutions was purified with a Milli-Q reagent water purification system (Millipore, Bedford, MA).
Figure III-1. Schematic (A) and assembled view (B) of a prototype microsensor designed for *in vivo* glucose sensing.
Figure III-2. Schematic of the ultrafiltration process where molecular flow is driven by a negative pressure across a semi-permeable membrane.
Instrumentation and Apparatus

The three major subsystems that make the system displayed in Figure III-3 include the peristaltic pump system, ultrafiltration probe and FT-NIR spectrometer. The peristaltic pump system was provided by ASL Analytical, Inc., Coralville, IA and consisted of five independent peristaltic pumps whose rates can be controlled through the controller and PC software. Three of the pumps control the volumes of glucose, urea and lactate and a fourth pump provided blank buffer in order to adjust the final concentration of the analytes.

An Acculab ALC-320.3 milligram balance (Precision Weighing Balances, Inc., Bradford, MA) was use to calibrate each pump system. Solutions from the pumps were fully mixed with a 1 cm diameter mini-mixer equipped with a magnetic bar. Before the mixture went to waste, the mixed solution flowed past an ultrafiltration probe, model UF-3-12 purchased from Bioanalytical Systems, Inc., West Lafayette, IN. This probe consisted of three 12 cm long loops composed of a membrane with a 30 kDa molecular weight cut-off. A Pump 22 multiple syringe pump (Harvard Apparatus, Inc., Holliston, MA) was connected to the inner volume of the UF loops and created the negative pressure required to draw sample across the membrane and into the spectrometer for measurement.

Near infrared spectra were collected with a Nicolet 670 FT spectrometer (Nicolet Analytical Instruments, Madison, WI). This spectrometer was equipped with a 20 Watt tungsten-halogen lamp, calcium fluoride beam splitter and cryogenically cooled InSb detector. A K-band interference filter (Barr and Associates, Westford, MA) was positioned before the sample cell to restrict the collected spectra to the combination region (5000-4000 cm$^{-1}$). The sample path length was set at 0.94 mm with a borosilicate capillary tubing (Warner Instrument Co., Hamden, CT). The sample temperature was controlled and monitored with a LFI3751 Digital Temperature Control Instrument (Wavelength Electronics Inc., Bozeman, MT).
Figure III-3. Schematic layout of the experimental design for generating and measuring glucose transients.
Spectral Collection

Over a 10 day period, 8 separate sets of spectral data were collected by using one of the concentration profiles presented in Figure III-4. Prior to the start of the programmed profile, blank buffer was passed through the system for 10 minutes followed by the described concentration profiles for glucose, urea and lactate. Custom designed software controlled the pumps to produce the profiles.

Spectra were collected as 32 co-added double sided interferograms with 2048 points at every zero-crossing of the HeNe reference laser. After applying triangular apodization, Mertz phase correction and one level zero-filling to each interferogram, single-beam spectra were created by the spectral processing software available on the spectrometer (OMNIC software, Version 5.1B, Thermo-Nicolet). In order to match the spectral resolution anticipated for the microsensor, the spectra were collected with a resolution of 16 cm\(^{-1}\). The resulting spectra were saved as CSV files for future processing.

Data Processing

Air bubbles are commonly present in the fluid collected from an UF probe. For spectroscopic analyses, air bubbles adversely impact the spectra and should be avoided. For the data collected here, bubble formation averaged one per hour and each lasted one minute in the optical path. As air passed through the capillary optical cell, the collected spectra were easy to distinguish from actual sample spectra and were discarded from the analysis. Two methods were used to distinguish spectra for which air was present in the optical path. The first was to check the intensity of the transmitted light. The strong absorptivity of water compared to the lower absorptivity of air resulted in high intensities in the presence of air. RMS noise levels of 100% lines over the 4600-4500 cm\(^{-1}\) spectral range provided a second method to distinguish spectra with air. Increases in intensity and RMS noise were simple indications of air bubbles in the samples.
Figure III-4. Four designed concentration profiles (A, B, C, and D) with different varying trends and concentration levels for each analyte.
PLS and NAS calibration models were developed by using MATLAB R2007a (Math Works, Inc., Natick, MA). Spectra for PLS calibration models were taken from the time profiles. Spectra for the NAS calibration models corresponded to pure solutions of buffer alone, glucose, urea, and lactate through the time profiles of some trials.

Results and Discussion

Peristaltic Pump System

The rate of solution flow through the peristaltic pumps was controlled by varying the frequency of rotation of the pump’s rotors. The following equation was used to relate solution flow rate to pump frequency:

\[
\text{Flow rate (ml/min)} = \beta_1 \times \text{Frequency (Hz)} + \beta_0 \quad \text{Equation III-1}
\]

where \( \beta_1 \) and \( \beta_0 \) are the slope and intercept of the linear function, respectively.

The concentration of each analyte was determined from the relative flow rates for the individual pumps used to flow the standard solutions into the mixing chamber and then through the spectrometer. Equation III-2 gives an example of the calculation used for determining the concentration of glucose at any point in time during the experiment. Flow rates were for each pump by using the computer to set the frequency of the pump rotors and the glucose concentration in the final solution after mixing is given by the following equation:

\[
[Glc]_t = \frac{Q_1 \times C_{\text{Glc}}}{Q_1 + Q_2 + Q_3 + Q_4} \quad \text{Equation III-2}
\]
where $Q_1, Q_2, Q_3$ and $Q_4$ are the assigned flow rates for four pumps, $Q_1$ is the pump for the standard glucose solution, and $C_{Glc}$ is the concentration of the stock glucose solution. $[Glc]_t$ is the calculated glucose concentration at time $t$.

Accuracy of the solution flow rates is critical because analyte concentrations in the final solutions will be used to generate the PLS calibration models and to evaluate analytical performance of the PLS and NAS calibration models. For this reason, each pump must be calibrated and the stability of these calibration functions must be assessed. Pumps were calibrated individually by setting the rotor frequency and measuring the mass of water displaced as a function of time. This calibration process was automated with a computer for pump frequency settings of 100, 300, 500, 700, 900 and 1000 Hz. Water displacement rate was measured by recording the mass of water accumulated on the tray of an analytical balance. Mass readings were recorded every 2 seconds over a period of 240 minutes. The flow rate was determined assuming the density of the water is 0.997 g/ml at 25 °C. A plot of pump frequency versus solution flow rate is linear and the slope provides the necessary information to adjust the pump for different solution flow rates to achieve the targeted concentrations. Figure III-5 demonstrates the linearity of such a pump calibration plot. In this typical calibration result of one pump, the slope of the linear fitting between the flow rate and the pump frequency is $0.0029\pm9.3\times10^{-6}$ and the intercept is $0.010\pm0.0062$. Pumps were calibrated at the end of each day of usage.

Stability of pump calibrations is documented by the data in Figure III-6. This plot shows the magnitude of the calibration slope over the 22 separate calibration procedures performed over the 30 day period of this investigation. Data are provided for the five available pumps, although only four of these pumps were used to create transients for the three analytes. For the most part, pump calibration slopes were constant over the time course of this experiment and concerns over possible changes in the calibration function owing to wear of the pump tubing were unsubstantiated by these findings. Table III-1 summarizes results from a simple statistical analysis of these calibration slope values. No
Figure III-5. Linear relationship between flow rate and pump frequency for calibration of Pump 1.
Figure III-6. Calibration results of 22 trials over a period of 30 days for the peristaltic pump system.
Table III-1. Pump calibration slopes determined for implementation of 22 calibrations for each of five pumps over a 30-day period.

<table>
<thead>
<tr>
<th>Pump No.</th>
<th>Mean of Slope (ml/(min*Hz))</th>
<th>Standard Deviation (ml/(min*Hz))</th>
<th>Relative Standard Deviation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.907e-3</td>
<td>2.9e-5</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>2.668e-3</td>
<td>4.7e-5</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>2.655e-3</td>
<td>5.0e-5</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>2.808e-3</td>
<td>5.7e-5</td>
<td>2.0</td>
</tr>
<tr>
<td>5</td>
<td>2.604e-3</td>
<td>1.5e-5</td>
<td>0.58</td>
</tr>
</tbody>
</table>
major trends in the slope term are noted in Figure III-6 and the relative standard deviation is less than 2.0%. Based on the calibration uncertainties for the pumps in Table III-1, error propagation techniques suggest a corresponding relative uncertainty of 3.0% for analyte concentrations. Of course, this level of uncertainty can be reduced by determining the calibration function on each day of operation.

System Delay

System delay corresponds to the time between when a new concentration is started by a change in the pump frequency and when this new concentration reaches the optical cell for detection. Two primary sources of delay are evident in the experimental design illustrated in Figure III-3. First, once the pump frequency is changed to create a new concentration, time is required for the resulting sample to reach the UF probe. Second, time is required for the new sample to be sampled by the UF probe and be transported to the optical cell within the spectrometer. The slower of these two processes is the UF sampling, which is governed by a syringe pump set to provide a solution flow rate of 5 μl/min within the UF tubing. Given the dimensions of the optical cell, length and diameter of the UF tubing, and solution flow rate, the estimated time for the sample to transverse from the UF probe to the spectrometer cell is 10 minutes. On the other hand, only 3 minutes is estimated as the time from the pumps to the UF probe. In total, the estimated system delay is around 13 minutes. System delay times were measured by following a bubble purposed placed in the system tubing and delay times of 10-13 minutes were determined by this method.

Impact of system delay on PLS measurements was examined. In this case, PLS calibration models were constructed from spectra for which analyte concentrations were assigned with different assumed system delays. The corresponding SEP values were determined and the results are presented in Figure III-7. The minimum SEP is observed at the delay time that best matches the system delay, thereby provided the most accurate
Figure III-7. Identification of delay time based on cv-SEP values from cross-validation PLS models while systematically varying delay time.
Table III-2. Delay time measurements based on cv-PLS models.

<table>
<thead>
<tr>
<th>Experiment Day</th>
<th>Concentration Profile&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Glucose</th>
<th>Urea</th>
<th>Lactate</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>11.6</td>
<td>12.0</td>
<td>11.4</td>
<td>11.7</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>12.8</td>
<td>12.4</td>
<td>12.8</td>
<td>12.7</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>12.6</td>
<td>12.8</td>
<td>12.6</td>
<td>12.7</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>12.8</td>
<td>12.6</td>
<td>12.8</td>
<td>12.7</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>10.4</td>
<td>10.4</td>
<td>10.2</td>
<td>10.3</td>
</tr>
<tr>
<td>8</td>
<td>D</td>
<td>10.0</td>
<td>10.6</td>
<td>10.4</td>
<td>10.3</td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>10.2</td>
<td>10.4</td>
<td>10.6</td>
<td>10.4</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>10.4</td>
<td>10.6</td>
<td>10.8</td>
<td>10.6</td>
</tr>
<tr>
<td>Average ± Standard Deviation</td>
<td></td>
<td>11.4±1.2</td>
<td>11.5±1.1</td>
<td>11.5±1.1</td>
<td>11.4±1.1</td>
</tr>
</tbody>
</table>

<sup>1</sup>See Figure III-4 for details.
concentration assignments for each spectrum within both the calibration and prediction calculations. SEP values were estimated by using a leave-10%-out cross validation calculation coupled with 7 latent variables in the PLS models and 12 second spectral acquisition times. The minimum cv-SEP was observed at a delay time of 10.4 minutes, which is within the range measured by timing air bubbles.

Table III-2 summarizes the identified delay time for each experiment based on cv-PLS models for three major components. Similar delay times are evident for each analyte and are consistent with visual observations.

**PLS Calibration Models – Predictions within the Calibration**

Pure component spectra are presented in Figure III-8A for glucose, urea and lactate. Over this spectral range, the spectrum for glucose is characterized by a broad band located approximately at 4700 cm\(^{-1}\) as well as two narrower bands centered at 4400 and 4300 cm\(^{-1}\). For urea, two vibrational bands are centered around 4575 and 4650 cm\(^{-1}\), respectively. There are three absorbance bands for lactate located around 4700, 4430 and 4350 cm\(^{-1}\) and these bands overlap considerably with those of glucose. Figure III-8B shows all the spectra collected during a single day of data collection. Features within these sample spectra could be easily distinguished due to the component variance within the concentration profile.

PLS models were built on the basis of a splitting of spectra into calibration and prediction data sets with 80% of the spectra used for calibration and 20% for prediction. The first 600 spectra collected (50 minutes) were not used in the calibration or prediction process. These spectra were collected during a period when the flow rate was unstable for the syringe pump used to collect fluid from the UF probe. The instability of this pump was common at the beginning of each day and required less than 50 minutes to stabilize.
Figure III-8. Pure component spectra for 1 mM solutions of glucose, urea and lactate (A). Examples of sample spectra collected during a single day experiment (B).
The number of factors was selected for each PLS calibration model on the basis of cv-SEP values. As noted above, the cv-SEP was calculated for each model with a different number of factors ranging from 1 to 15. Figure III-9 provides a typical plot of cv-SEP for such a model. As indicated in this figure, the cv-SEP drops as more factors are used. The optimum number of factors was determined by applying the $F$-test at the 95% confidence level to assess the statistical difference between sequential cv-SEP values.

The fewest number of factors for which the cv-SEP was not significantly smaller than the lowest cv-SEP was taken as optimal. As a check, the loading vector for the final factor was visually inspected to verify the absence of random noise. An optimum of seven factors was found for each PLS model, as noted below.

Table III-3 presents results for the PLS generated for glucose measurements during the eight experiments. In each case, seven factors were deemed optimal. SEC and SEP values range from 0.30 to 0.44 mM for glucose and the corresponding mean percentage errors (MPEs) ranged from 1.5% to 2.0%.

Figure III-10A shows the calibration and prediction results for one of the eight trial. This trial used the concentration profiles displayed in Figure III-4C where glucose had two separated transients. In Figure III-10A, time profiles are presented for the reference points (what the concentration should be) and the glucose concentrations provided by the PLS calibration model for both the calibration and prediction data sets. The PLS predictions follow the programmed transient. After the transient was completed, pure components of each analyte were passed through the system sequentially and the glucose PLS model responds to the 60 mM concentration of glucose in the standard solution. Figure II-10B shows the corresponding concentration correlation plot superimposed on the ideal unity line. Again, accuracy of the calibration and predict concentrations are noteworthy as all the points fall on the unity line.
Figure III-9. Trend of SEP values versus number of factors used in PLS modeling.
Table III-3. Summary of the within calibration performance of PLS calibration models for glucose.

<table>
<thead>
<tr>
<th>Experiment Day</th>
<th>Number of Latent Variables</th>
<th>SEC (mM)</th>
<th>SEP (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>0.30</td>
<td>0.32</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0.40</td>
<td>0.37</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.38</td>
<td>0.39</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>0.42</td>
<td>0.44</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>0.34</td>
<td>0.32</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>0.37</td>
<td>0.38</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>0.38</td>
<td>0.36</td>
</tr>
</tbody>
</table>
Figure III-10. Glucose concentration time profile as defined by the reference (blue), calibration (green) and prediction (red) measurements (A). Concentration correlation plot for glucose PLS model (B).
PLS Calibration Models – Predictions Outside the Calibration

The ability to predict glucose concentrations outside the calibration model was explored by building the PLS calibration model using data collected on the first day and then using the PLS model to predict the concentration of glucose over the next seven trials of the experiment. Figure III-11 shows the time profile and concentration correlation plot for the initial calibration period (Day 1) as well as the subsequent predictions for trials 2-8. This PLS calibration model used spectra over the 4800-4200 cm\(^{-1}\) spectral range and 7 latent variables. This model was applied to the 20175 spectra collected in subsequent seven trials that spanned nine days. In both plots, the predicted concentrations match the targeted concentrations all the way to the Day 10 data. The 60 mM concentrations of glucose are evident for the pure component spectra collected at the end of Days 6, 9, and 10. Predicted glucose concentrations fall slightly below zero (< -1 mM) for several of the Days. Attempts to avoid these negative concentrations by using various common preprocessing techniques, such as derivatives, baseline correction, single-day mean center, auto scale, standard normal variance (SNV) and multiplicative scatter correction (MSC) were unsuccessful. The concentration correlation plot in Figure III-11B illustrates measurement accuracy over this time period. No bias is indicated, SEC and SEP values are 0.30 and 0.71 mM, respectively, and the MPE is 3.3% which is similar to the calculated concentration errors for the peristaltic pump system. Overall, the PLS model provides accurate glucose concentration values over a period of 10 days.

Table III-4 summarized the PLS model SEP values for the different days for predictions within the calibration and outside the calibration. For the most part, SEP values are approximately a factor of two higher for predictions outside the calibration. However, there is no trend in terms of an increase in SEP values as a function of time. In addition, there is no evidence that SEP values correlate to the type of concentration profile.
Figure III-11. Glucose concentration transient profiles for reference (blue), PLS calibration (green) and PLS predicted (red) concentrations within and outside the calibration process (A); and concentration correlation plot for PLS glucose concentration predicts within and outside the calibration process (B).
NAS Calibration Models – Predictions within Calibration

Construction of a NAS calibration model requires a set of pure component spectra for each of the constituents within the sample matrix. Such pure component spectra were only collected for three of the transient trials (Days 6, 9, and 10). The NAS calibration vectors were determined for these days by projecting the pure glucose spectrum onto the pure component spectra collected for buffer solution, urea and lactate. In these experiments, 60 pure component spectra were collected for each analyte and the NAS was calculated by projecting the pure component spectra to the background matrix which includes the spectral information of other components except the targeted analyte.

NAS model results are summarized in Table III-5 for the three trials for which pure component spectra were collected. The spectral range and delay time were the same as those used for the PLS models. As indicated in the table, 6 or 7 PCA factors were used to characterize the replicate pure component spectra before constructing the NAS model. SEP values across the entire transient profile for Days 6, 9, and 10, respectively are 0.70, 0.56, and 0.52 mM. SEP values for the NAS models are approximately 1.5-fold higher than those for the corresponding within calibration PLS prediction errors. One possible explanation is that the PLS algorithm effectively excludes noise from the calibration vector, unlike the NAS method.

Figure III-12 shows both the profile and concentration correlation plot for the NAS prediction results for the Day 6 data. Predictions follow the reference concentrations throughout the entire transient and no unpredictable offset or bias is evident as all predicted concentrations fall along the unity line. It is important to remember that the concentrations of urea and lactate are varying according to the profiles presented in Figure III-4C throughout the predictions plotted in Figure III-12A. The NAS and PLS calibration vectors are compared in Figure III-13 for the Day 6 data. A similarity is evident between these vector shapes, although more variation is evident in the NAS vector, which is consistent with the higher SEP values.
Table III-4. Comparison of PLS Glucose Concentration Prediction Errors Outside and Within the Calibration Process.

<table>
<thead>
<tr>
<th>Experiment Day</th>
<th>Concentration Profile(^1)</th>
<th>SEP (mM) Outside Calibration</th>
<th>SEP (mM) within Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0.89</td>
<td>0.42</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>0.81</td>
<td>0.41</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>0.75</td>
<td>0.39</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>0.86</td>
<td>0.48</td>
</tr>
<tr>
<td>8</td>
<td>D</td>
<td>0.72</td>
<td>0.39</td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>0.45</td>
<td>0.41</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>0.79</td>
<td>0.39</td>
</tr>
</tbody>
</table>

\(^1\)See Figure III-4 for details.
Table III-5. Summary of NAS Model Predictions within Calibration Days and Comparison to within Calibration Day Prediction Errors for PLS Models.

<table>
<thead>
<tr>
<th>Experiment Day</th>
<th>Concentration Profile$^1$</th>
<th>Factor Number</th>
<th>NAS SEP (mM)</th>
<th>PLS SEP (mM)$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>C</td>
<td>6</td>
<td>0.70</td>
<td>0.44</td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>7</td>
<td>0.56</td>
<td>0.38</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>7</td>
<td>0.52</td>
<td>0.36</td>
</tr>
</tbody>
</table>

$^1$See Figure III-4 for details;

$^2$Same results as in Table III-3.
NAS Calibration Models – Predictions Outside the Calibration

The NAS calibration model generated for glucose from the pure component spectra collected on Day 10 was applied to all the spectroscopic data collected in all eight days. In this case, all glucose concentrations were taken from Equation III-3:

\[ \hat{c}_i = NAS \cdot X_i \quad \text{Equation III-3} \]

where \( \hat{c}_i \) is the predicted concentration, NAS is the calibration vector and \( X_i \) is the collected spectrum.

Figure III-14A shows the NAS predicted concentration profiles superimposed on the corresponding reference glucose concentrations for each day in the study. Similar to the PLS results, the NAS model is able to follow the glucose transients in all cases and to respond properly to the 60 mM glucose pure component spectra collected in Days 6, 9, and 10. The concentration correlation plot in Figure III-14B indicates accurate concentration predictions with no bias across days. Across all measurements and across all days, the SEP is 0.80 mM, which is slightly higher than the value of 0.70 mM from the PLS model. The measured MPE value is 3.7\%, which is similar to the estimated pump error of 3.0\%.

Prediction errors outside the calibration days for both PLS and NAS calibration models are summarized in Table III-6. These values indicate a similar ability to predict glucose concentrations across days.

Conclusions

The results presented in the chapter demonstrate the functionality of a novel experimental apparatus for generating independent analyte concentration transients. The apparatus provides a stable and accurate platform for \textit{in vitro} evaluation of NIR spectroscopic sensors, including an envisioned microsensor. Two types of multivariate
Figure III-12. Predicted glucose concentrations versus time, varying follow reference concentrations (A). Concentration correlation plot for glucose NAS model during same day as calibration spectra (B).
Figure III-13. Comparison of calibration vectors for PLS and NAS models.
Figure III-14 Glucose concentration transient profiles for reference (blue), and NAS predicted (red) concentrations within and outside the calibration process (A); and concentration correlation plot for NAS glucose concentration predicts within and outside the calibration process (B).
Table III-6. Comparison of NAS and PLS Glucose Concentration Prediction Errors Outside the Calibration Process.

<table>
<thead>
<tr>
<th>Experiment Day</th>
<th>Concentration Profile&lt;sup&gt;1&lt;/sup&gt;</th>
<th>SEP (mM) in PLS Model</th>
<th>SEP (mM) in NAS Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>0.32</td>
<td>0.72</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0.89</td>
<td>0.77</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>0.81</td>
<td>0.84</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>0.86</td>
<td>0.94</td>
</tr>
<tr>
<td>8</td>
<td>D</td>
<td>0.72</td>
<td>0.94</td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>0.45</td>
<td>0.94</td>
</tr>
<tr>
<td>10</td>
<td>A</td>
<td>0.79</td>
<td>0.53</td>
</tr>
</tbody>
</table>

<sup>1</sup>See Figure III-4 for details.
calibration methods are tested and both the PLS and NAS models provided excellent glucose concentration predictions during programmed transient profiles of glucose, urea, and lactate. Glucose concentration predictions are compared for predictions made within and outside the calibration process. Results indicate that SEP values are slightly higher for both the PLS and NAS concentration predictions outside the calibration process. Comparing prediction errors for glucose concentrations outside the prediction process, the PLS method provides slightly lower errors, although no evidence of bias or offset are present from either method. Overall, the MPE’s across day for predictions outside the calibration process are on the order of the estimated concentration errors caused by the transient generation pumping hardware (3.0%).
CHAPTER IV
SIMULTANEOUS MONITORING OF GLUCOSE AND LACTATE DURING FERMENTATIONS

Introduction

Fermentation processes use microorganisms, such as yeasts, bacteria and fungi, to convert basic raw materials into valuable products. Fermentations are widely used in the production of food, pharmaceutical therapeutics, enzymes and bulk chemicals. Analytical methods are desired to follow the concentrations of key microbial metabolites, thereby providing the means to increase the productivity, efficiency and reproducibility of the upstream bioprocess. Such new analytical methods can be coupled with existing bioprocess monitors for pH, temperature and dissolved oxygen to optimize conditions to maximize cell growth and product generation.\textsuperscript{147-149}

The Food and Drug Administration published their Process Analytical Technology (PAT) initiative in 2004 with the objective of encouraging the development of innovative analytical methods to monitor bioprocesses used in the generation of therapeutics.\textsuperscript{150} As a result, many PAT methods have been reported where targeted analytes are measured by optical,\textsuperscript{151-153} electrochemical,\textsuperscript{154-155} chromatographic,\textsuperscript{156-159} and mass spectrometric methods.\textsuperscript{160-162} The inherent advantages of optical sensors for bioprocess monitoring are well recognized, and include the potential for non-destructive measurements that avoid possible contamination of the reactor medium.\textsuperscript{151-152,163} Optical methods also offer the ability to perform continuous measurements of several analytes simultaneously. Such features are attractive for process monitoring and control.

NIR spectroscopy is a particularly appealing option for bioprocessing monitoring because of its fast, low cost and non-destructive nature.\textsuperscript{163-169} NIR spectroscopy is based on combination and overtone vibrational modes associated with C-H, N-H, and O-H bonds\textsuperscript{170} and, as such, is well suited for direct measurements with no or minimal sample
preparation. On the other hand, NIR absorption bands are relatively weak, broad, highly overlapped and often suffer low sensitivity and specificity. NIR spectroscopy can be difficult to reproduce owing to the sensitivity of NIR spectra to a host of physical phenomena, including the scattering of light through different interfaces. Chemometrics methods are typically used to compensate for these limitations and achieve accurate and robust analytical measurements. Modern NIR spectroscopic methods rely heavily on chemometrics modeling, especially for industrial applications.

NIR spectroscopy has been used as a tool for monitoring fermentation processes since the early 1990s. Vaccari et al. proposed NIR spectroscopy for online monitoring of glucose during the production of lactic acid with Lacto-bacillus casei. Despite large relative errors in these early experiment, this work initiated efforts to explore the potential of NIR spectroscopy for bioprocess monitoring. The general NIR methods and approaches for bioprocess monitoring with NIR spectroscopy are outlined in Figure IV-1.

Martin Rhiel et al. investigated the use of NIR spectroscopy over the combination region of the NIR spectrum (5000-4000 cm\(^{-1}\)) for off-line monitoring of important nutrients and byproducts (glucose, glutamine, lactate and ammonia) in a serum-based mammalian cell culture medium. In this work, the impact of spectral range was investigated and findings show that similar analytical performance could be achieved with fewer factors by using an optimized spectral range as opposed to the full range of spectral frequencies. It was concluded that the selecting a spectral range that distinguishes the absorption features for each analyte could enhance the performance of PLS calibration models. Results for glucose, lactate, glutamine and ammonia corresponded to SEP values of 0.82, 0.94, 0.55, and 0.76 mM, respectively.

NIR spectroscopy was also demonstrated as a tool for in-situ monitoring of glucose, lactate, glutamine and ammonia which are critical for mammalian cell cultures,
Figure IV-1. Diagrammatic representation of the different sample presentation modes and measurement approaches used for bioprocess monitoring with NIR spectroscopy.
including the cultivation of CHO-K1 animal cells. A steam sterilizable in-situ fiber optic probe was coupled with NIR spectroscopy to collect spectral of fermentation medium during cultivation. These spectra were processed by a 2nd derivative and SNV technique prior to submission to individual PLS calibration models; one for each analyte. SEP values as low as 0.072 g/L, 0.0144 g/L, 0.308 mM and 0.036 mM were reported for glucose, lactate, glutamine and ammonia, respectively.

Payal Roychoudhury et al. proposed a multiplexed calibration technique which allows calibration models to be built from multiple bioreactors. This approach provides a transferable calibration function for monitoring bioreactors for multiple vessels and facilitates. Such calibration models were developed for glucose and lactate in CHO cell lines with pre-defined analyte specific information. With optical differences between probes and other contributory sources of variance from the different bioreactors, more factors were necessary to explain the additional variance compared to single reactor models. Despite operation across reactors, only slight model degradation were reported compared to those of single reactor models. This work proved the potential of real-time multi-analyte monitoring from multiple bioreactors.

Because of the powerful correlation ability of multivariate regression techniques, PLS calibration models are the most widely used methods for bioprocess monitoring. In this chapter, PLS, NAS and a modified HLA calibration technique were used to quantify glucose and lactate in a microbial fermentation process. An UF probe is used for ex-situ sampling as a means to produce medium samples free of cells. Cells must be eliminated to reduce scattering effects and maximize radiant powers for the spectroscopic measurements. The resulting NAS and modified HLA models are shown to provide similar prediction errors as conventional PLS models while being more robust.

**Experimental Section**

**Sample Preparation**
The fermentation broth was prepared from standard materials including yeast extract (Sensient Bio-ingredients Co., Indianapolis, IN); yeast nitrogen base w/o amino acid (Beckson and Dickinson Co., Franklin Lakes, NJ) and soy peptone (Kerry Bio-Science Co., Brantford, Ontario, Canada). Buffer components for the fermentation broth were KH$_2$PO$_4$ and K$_2$HPO$_4$ which were purchased from Thermo Fisher Scientific Inc. (Fairlawn, NJ). Glucose was added to the fermentation medium as the primary carbon source and was purchased from Sigma Chemical Co. (St. Louis, MO). Distilled-deionized water was used for the preparation of all solutions and was prepared by purifying house-distilled water with a Milli-Q reagent water system (Millipore, Bedford, MA).

Instrumentation and Apparatus

Figure IV-2 shows the experimental arrangement of the bioreactor, sampling system and FT-NIR spectrometer. Fermentation proceeded in a narrow mouth bottle equipped with a magnetic stir bar and plate to provide agitation. A pH electrode coupled with an Orion 320 PerpHecT Basic Benchtop pH Meter (Thermo Scientific Inc., Beverly, MA) was used to follow pH of the fermentation broth in real-time. For ex situ sampling, an UF-3-12 UF probe was purchased from Bioanalytical Systems, Inc. (West Lafayette, IN). Each UF probe contained three loops of 12 cm long membrane with a molecular weight cut-off of 30 kDa. A common peristaltic pump was used to circulate the fermentation medium through a section of Tygon tubing into which the UF probe was positioned. A Pump 22 multiple syringe pump (Harvard Apparatus, Inc., Holliston, MA) was used to apply a negative pressure to the inner chamber of the UF probe, thereby drawing clear sample from the Tygon tubing loop filled with circulating fermentation broth.

A Nicolet 670 FT spectrometer (Nicolet Analytical Instruments, Madison, WI) equipped with a 20 watt tungsten-halogen lamp, calcium fluoride beam splitter and
Figure IV-2. Schematic layout of experimental design for fermentation process monitoring.
cryogenically cooled InSb detector was used to collect near infrared spectra of the ex situ sampled solution. To restrict the collected spectra to the combination region (5000-4000 cm⁻¹), a K-band interference filter (Barr and Associates, Westford, MA) was positioned before the sample cell. The measurement cell consisted of a borosilicate capillary with an inner diameter of 0.94 mm (Warner Instrument Co., Hamden, CT). This inner diameter of the capillary established the optical path length for the measurement.

Experimental Procedures and Spectral Collection

For each experiment, one liter of fermentation broth was prepared by mixing 10 g/L yeast extract, 15 g/L yeast nitrogen base w/o amino acid, 20 g/L soy peptone, 12 g/L KH₂PO₄ and 2 g/L K₂HPO₄ with deionized water. A 25 ml aliquot of the broth was left out for the collection of the background spectra needed to construct the NAS calibration model. Then 40-50 g of glucose and 1 ml of cell ‘seed’ were added into the fermentation broth. The cell ‘seed’ corresponds to an aliquot of fermentation broth with a high cell density retained from a previous fermentation experiment. Once all components were added, the fermentation mixture was left overnight at room temperature for an initial growth period. During this overnight period, the medium become noticeably turbid and the pH dropped from about 5.7 to 4.4 because of cellular metabolism.

Initially, NIR spectra were collected of water and blank background fermentation medium (no cells or glucose). After the fermentation was initiated by adding glucose and cells to the medium, the fermentation broth would be left alone through the night and spectra of the UF sampled medium commenced and Spectral collection were repeated continuously during the fermentation process until the glucose was totally consumed. During the fermentation, the pH value was monitored and periodically adjusted to 5.7 - 5.9 by adding 6 M NaOH dropwise. These pH adjustments were needed to maintain cell growth and continued cellular metabolism. After the glucose was consumed and the
fermentation was completed, spectra of distilled water and the blank background medium were passed through the sample cell and their NIR spectra collected.

Through the experiments, all spectra were collected as 32 co-added double sided interferograms with 2048 points at every zero-crossing of the HeNe reference laser. Triangular apodization, Mertz phase correction and one level of zero-filling were applied to each interferogram. The Fourier processed interferograms were processed to produce single-beam spectra with a spectral resolution of 16 cm\(^{-1}\). All spectra processing to produce single beam spectra were performed with the OMNIC software (Version 5.1B, Thermo-Nicolet).

Reference concentrations of glucose and lactate were analyzed every 20 minutes with an YSI model 2300 STAT plus (Yellow Springs Instrument Co., Yellow Springs, OH). Meanwhile, cell density was checked by measuring turbidity at 600 nm with a Perkin-Elmer Lambda 1 UV/Vis spectrophotometer. A typical cell growth curve is shown as Figure IV-3.

Four trials of the fermentation were performed over a period of 10 days and Trial 1,2,3 and 4 were executed on day 1, 4, 6 and 10, respectively.

**Data Processing**

NIR spectra of water before the fermentation process were used as reference spectra for absorbance calculations. Since the reference concentration of glucose and lactate was measured with a YSI analyzer every 20 minutes and the spectral acquisition time was approximately 12 seconds, the concentrations of glucose and lactate assigned to each spectrum for multivariate calibration purposes were calculated by fitting a plot of concentration versus time to a third order polynomial. Assigned concentrations were interpolated from this fit for each spectrum.

Air bubbles were observed infrequently in the UF sample fluid. On average, an air bubble would appear once an hour and last approximately one minute in the optical
Figure IV-3. Growth curve of the cells by measuring OD$_{600}$ of the fermentation broth sample during the experiment.
cell. As noted in Chapter 3, air bubbles were easy to detect on the basis of intensity of the single-beam spectra and RMS noise on 100% lines. Figure IV-4 shows an example of bubbles detected by both methods. The dramatic difference of the results from the spectra with air bubbles could be easily distinguished. By setting a threshold for separating and isolating, all spectra suspected of containing air in the optical path were discarded and not used for analysis.

Multivariate calibration models were generated for glucose and lactate by using the PLS and NAS methods, as described in earlier chapters. PLS models were built based on a randomly splitting of the spectral data where 80% of the spectra were used for calibration and 20% for prediction. For NAS models, pure component spectra were collected from individual standard 60 mM solutions of glucose and lactate at pH of 5.80. Background spectra for the NAS model corresponded to spectra collected from fermentation broth without cells and glucose as well as spectra collected at the end of the fermentation after the glucose had been depleted from the medium. These calibration models were used to predict the concentrations of glucose and lactate during fermentations on other days. All model calculations were performed using MATLAB R2007a (Math Works, Inc., Natick, MA).

The cell-free UF samples provided scatter free measurements that resulted in high optical throughputs and high signal-to-noise ratios for the measurements. A drawback of using an UF probe, however, is the delay between when the sample passed across the UF membrane and when it reached the optical cell where the spectrum can be collected. As discussed in Chapter 3, this delay time must be considered in order to assign accurate concentrations for each spectrum. The cv-SEP method described in Chapter 3 was used here to determine the delay time on the basis of a minimal cv-SEP. This method resulted in a delay time of 8.6 minutes, which has been applied to each of the four fermentations. This delay time matches the time for a bubble to transverse from the UF probe to the optical cell within the spectrometer.
Figure IV-4. Single-beam spectral intensities (A) and RMS noise values (B) for spectra collected from the UF sampling probe. High values are an indication of an air-bubble in the optical cell.
Results and Discussion

Four Component Decomposition Analysis

The fermentations performed in this experiment were performed without a specific cell line. Basically, the fermentation broth was exposed to the ambient atmosphere and a non-select set of endogenous microbes grew in the broth. After an initial pilot, an aliquot of the ending high cell density broth was used to seed the next fermentation. This seeding procedure resulted in fast fermentations.

Without specifying the cell line in the fermenter, it was not possible to know the metabolite pathway. The metabolic reaction was characterized in the following manner. First, a linear regression analysis was performed on the four known pure component spectra to fit a sample spectrum to the initial four components of glucose, yeast extract, yeast nitrogen base w/o amino acid and soy peptone. Figure IV-5A shows the spectrum for each component. This corresponding regression analysis is described by the following equation:

$$A_s = \beta_1^*A_{\text{Glc}} + \beta_2^*A_{\text{YE}} + \beta_3^*A_{\text{YNB}} + \beta_4^*A_{\text{Pep}} + \beta_5^*A_{\text{Constant}} + \beta_6^*A_{\text{Slope}} + \epsilon \quad \text{Equation IV-1}$$

where $A_s$ is the absorption spectrum of fermentation broth ultrafiltrate, $A_{\text{Glc}}$, $A_{\text{YE}}$, $A_{\text{YNB}}$ and $A_{\text{Pep}}$ represent the pure component absorption spectra of glucose, yeast extract, yeast nitrogen base and soy peptone and $A_{\text{Constant}}$ and $A_{\text{Slope}}$ have shapes of a horizontal line and a slope, respectively. The constant term is included to compensate for scattering effects, as well as a slope term to account for the combination of scattering and temperature fluctuations. $\beta_i$ represents the corresponding regression coefficients, and $\epsilon$ is the spectral residual.

Relative changes in these coefficients were examined to determine what components change concentration during the fermentation. The coefficient ratio was
Figure IV-5. Pure component spectra of glucose, yeast extract, yeast nitrogen base w/o amino acid and soy peptone (A) and time profile for the coefficient ratios for the four components within the fermentation medium (B).
Figure IV-6. Spectral residuals determined over the period of the fermentation after a four component regression analysis (A) and pure component spectra of glucose and lactate for byproduct identification (B).
calculated for each time point as the magnitude of the coefficient at time \( t \) divided by the initial value for this coefficient. Figure IV-5B presents how these coefficient ratios vary with time. This plot reveals that the coefficients for the major components such as glucose, peptone, yeast nitrogen base and yeast extract decrease during the fermentation. Besides, the coefficient of constant decreases while that for slope remains constant. The byproduct of the fermentation process was identified by examining the spectral residual from the regression treatment described in Equation IV-1. Figure IV-6A overlays spectral residuals collected over time during the fermentation. The two peaks centered at 4350 and 4430 cm\(^{-1}\) grow as the fermentation proceeds.

The spectral features of the accumulating byproduct resemble the spectrum of lactate. The pure component spectrum of lactate is provided in Figure IV-6B for comparison. On the basis of spectral similarity, the byproduct was identified as lactate, which is consistent with microbial metabolic processes.\(^{180}\)

**Glucose Measurements within a Single Fermentation**

For comparison purposes, both PLS and NAS calibration models were generated for the measurement of glucose during the fermentation. In both cases, the spectral range was set at 4900 - 4200 cm\(^{-1}\) and a system delay time of 8.6 minutes was used to relate concentrations to spectra. SEC and SEP values of 1.15 and 1.13 mM were obtained from the PLS model optimized with five latent variables. These PLS results are presented in Figure VI-7A as a time profile with the reference YSI measurements superimposed on the results from the PLS model. These results are also presented in the form of a concentration correlation plot in Figure VI-7B with the points superimposed on the unity line for the ideal correlation. In both cases, the PLS results accurately track the concentration of glucose during the fermentation. The MPE is 1.2% across the full time profile.
Figure IV-7. Time profile (A) and concentration correlation plot (B) for PLS calibration model for measuring the concentration of glucose during the fermentation.
Figure IV-8. Time profile (A) and concentration correlation plot (B) for NAS calibration model for measuring the concentration of glucose during the fermentation.
Table IV-1. NAS and PLS calibration model results for tracking the concentration of glucose during a single fermentation.

<table>
<thead>
<tr>
<th>Model</th>
<th>NAS (cm⁻¹)</th>
<th>PLS (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral Range</td>
<td>4900-4200</td>
<td>4500-4200</td>
</tr>
<tr>
<td>Factor No.</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>SEC (mM)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SEP (mM)</td>
<td>2.13</td>
<td>1.62</td>
</tr>
<tr>
<td>MPE (%)</td>
<td>2.0</td>
<td>1.5</td>
</tr>
</tbody>
</table>
The corresponding NAS calibration model was established by using a pure component spectrum for a standard glucose solution prepared in a blank buffer matrix coupled with a set of background spectra collected from the set-aside blank fermentation medium with no glucose or cells and the spectra collected at the final stage of the experiment when the glucose is supposed to be totally consumed and on other components left. Six PCA factors were used to characterize the spectral variance within the repeated spectra collected with blank medium. As noted above, the spectral range was 4900-4200 cm\(^{-1}\) and the delay time was 8.6 minutes. Results from this NAS model are presented in Figure IV-8. The good correlation between predicted and reference concentrations illustrates the analytical potential for the major fermentation components in fermentation process. Again, results for the NAS model demonstrate accuracy throughout the fermentation with no signs of bias or systematic errors.

Results from the PLS and NAS calibration models are summarized in Table IV-1. When the spectral range is set at 4900-4200 cm\(^{-1}\), the SEP for the NAS model is higher than that for the PLS model. The ability of the PLS model to provide a lower SEP is based on its ability to discriminate noise from the calibration by focusing on latent variables with higher signal-to-noise ratios. No such discrimination is possible with the NAS approach, thereby leading to higher prediction errors. As noted in Table IV-1, the SEP for the NAS model drops when the spectral range is reduced to 4500-4200 cm\(^{-1}\), which effectively removes the noisy spectral range at the higher frequencies. Furthermore, the narrower spectral range is prone to avoid the pH sensitive region. Spectral feature centered around 4700 cm\(^{-1}\) correspond to combinations of O-H bending and C-H stretching and give raise to same pH sensitivity, the detail could be found in the session of Lactate Measurements during a Single Fermentation. Still, the NAS model with an SEP of 2 mM is functional, as is evident by the results in Figure IV-6. Moreover, the MPE of 2.0% for the SEP is similar to the concentration uncertainty of the YSI analyzer, further demonstrating the practical utility of the NAS approach.
Calibration vectors for these PLS and NAS models are superimposed in Figure IV-9. This plot emphasizes the similarity in the vectors, particularly at the low frequency portion of the spectral range. As was noted in Chapter 3, the NAS model contains more variance which is a result of an inability to discriminate between analyte-dependent and analyte-independent sources of spectral variance, in contrast to the PLS method. On the other hand, PLS modeling is prone to chance correlation effects when non-analyte spectral variances are correlated to the concentration variance. In contrast, NAS calibration vectors are derived directly from the spectral features associated with the analyte of interest, rendering it less susceptible to chance correlation effects and making it unique relative to spectral features associated with matrix components, as a major advantage defined by Lober. Similarity between the shapes of the PLS and NAS calibration vectors is an indication that the PLS model is based on analyte specific information but no chance correlation effects.

It is interesting that the NAS model demonstrated significant resistance to systematic errors toward the end of the fermentation when the concentration of glucose is lowest. By design of the experiment, the concentration of glucose should be totally depleted at the end of the fermentation. Figure IV-10 shows a close up of the final 90 minutes of the glucose concentration profiles. These are the same data presented in Figures IV-7A and IV-8A but with a concentration axis scaled to highlight inaccuracies at the final stages of the fermentation. Ideally, the reference method, as well as the PLS and NAS calibration models, should measure zero for the concentration of glucose at the end of the process. At the end of the fermentation, the PLS measurements leveled-off at 3 mM, the YSI measurements reached 2.4 mM and predictions from the NAS model leveled-off at 0 mM. Although interferences are known for the YSI measurement of glucose, no attempt was made to identify if such interferences were present. Instead, the fermentation conditions were maintained and the pH value is set to 5.8 for an activating environment for the cells for an additional 72 hours with no additional change
Figure IV-9. Comparison of calibration vectors for PLS and NAS models.
Figure IV-10 Magnification of the last 1.5 hr of the glucose concentration profiles for the NAS and PLS models along with the YSI readings.
in the concentration of glucose measured by the YSI, which is an indication of either an interference present in the medium or the presence of a metabolic inhibitor that completely shuts down the fermentation reactions. Although the exact conditions of the final fermentation broth are not known, the NAS model is the only one that predicts zero for the concentration of glucose.

Lactate Measurements during a Single Fermentation

With lactate identified as the principal byproduct of the fermentation process, attempts were made to monitor its concentration as a function of time. Two critical aspects of this measurement are pH and suitable background spectra for building an NAS model.

The production of lactic acid during the fermentation process has the impact of lowering the pH of the fermentation broth. If left unattended, the pH will quickly drop below a value necessary to support cellular metabolism and the fermentation process stops. For this reason, continuous pH adjustments were necessary to maintain metabolism and the fermentation. The impact of solution pH on the normalized spectrum of lactate is shown in Figure IV-11. In this figure, changes are notable over the pH range of 4.0 to 6.0 which encompasses the anticipated pH variations during the fermentation process. Spectral feature centered around 4700 cm\(^{-1}\) correspond to combinations of O-H bending and C-H stretching and give raise to same pH sensitivity, as is evident in Figure IV-11. Absorption bands centered at 4350 and 4430 cm\(^{-1}\) are relatively insensitive to solution pH while the band at 4700 cm\(^{-1}\) shows pH sensitivity within the pH values of 4.6 to 6.0. For pH value lower than 4.6, the center positions of peaks around 4350 and 4430 cm\(^{-1}\) start to shift, however, this shift will not be considered in this experiment since the pH value of the fermentation broth would be maintained in range of 5.70 to 5.90. For this reason, two spectral ranges were examined for the lactate calibration models, the full range (4900-4200 cm\(^{-1}\)) and a narrower pH insensitive range (4500-4200 cm\(^{-1}\)).
Figure IV-11. Lactate spectra collected under different pH values
Background spectra are required for building NAS calibration models. Such spectra must contain all matrix components except for the analyte in question. For lactate, background spectra are difficult to obtain because the lactate is produced in situ and some type of separation would be required to generate lactate free blank samples once the fermentation reactions begin. As noted in Chapter 1, the HLA method introduced by Berger and Feld can be used to create background spectra in this situation. Briefly, the HLA method subtracts the contribution of the analyte spectrum on the basis of its known absorptivity and concentration in a particular sample.

The method assumes the sample spectrum is the linear sum of the individual spectra of the sample components. Spectra available for the measurement of lactate consist of: 1) spectra of blank medium (no lactate, glucose, or cells) and 2) spectra collected during the fermentation (containing both glucose and lactate). Since the in situ spectra collected contain glucose information and the background spectra do not contain glucose, a modified HLA method was developed and tested. In this method, the concentration of glucose obtained from the YSI reference method was used to subtract the contribution of glucose from each in situ spectrum. The same procedure described above for assigning glucose concentrations to spectra for PLS modeling was used here, where a third ordered polynomial function was used to interpolate glucose concentrations between YSI measurements. The subtraction of glucose contribution was accomplished in the same way the analyte spectrum is normally removed from sample spectra in the conventional HLA method (e.g., subtracting the product of the glucose absorptivity spectrum and the measured glucose concentration from each in situ sample spectra). The NAS calibration model for lactate was generated from the original background spectra (containing no glucose or lactate) and the pure component spectrum of lactate. For the measurement of lactate from each spectrum collected during the fermentation, the contribution of glucose was removed by subtraction, the concentration of lactate was determined by from the NAS calibration model.
This modified-HLS approach uses the HLA concept of subtracting a concentration adjusted pure component spectral from each sample spectrum. In this case, the contribution of an interfering species is subtracted, as opposed to the analyte, as is done in the conventional HLA method.

Figure IV-12 shows the impact of removing the glucose contribution from the sample spectral. Figure IV-12A shows the raw spectra from the UF probe fluid and Figure IV-12B shows these spectra once the glucose contribution has been removed. Subtraction of the glucose noticeably simplifies the sample spectra, as expected.

Figure IV-13 shows the results for lactate concentration predictions for calibration models based on a standard NAS calibration method and the proposed modified-HLA method with superimposed reference measurements. Results from each method are compared for the two spectral ranges noted above. Inspection of lactate concentration predictions for the standard NAS and modified-HLS methods over the 4900-4200 cm$^{-1}$ spectral range (Figures IV-13A and B, respectively) indicate a negative bias in the predicted concentrations. This bias is reduced for the NAS model when restricting the spectral range to 4500-4200 cm$^{-1}$ (Figure IV-13C) and is essentially eliminated by restricting the spectral range with the modified-HLA method (Figure IV-13D).

In the standard NAS model (generated from blank spectra and the lactate absorptivity spectrum), the impact of glucose in the sample spectrum is not accounted for, which results in systematic errors in the measurements. These systematic errors are manifested as off-sets in predictions and the magnitude of offsets depends on the concentration of glucose. For the standard NAS model over the 4500-4200 cm$^{-1}$ spectral range, the difference between reference and predicted lactate concentrations was largest at the beginning of the fermentation when glucose is at its highest concentration and decreases as glucose is consumed. When glucose is depleted in the in situ spectra, lactate predictions from the standard NAS model match the reference values. These findings suggest glucose is creating a negative bias in these lactate measurements. The modified-
Figure IV-12. Original sample spectra collected during a fermentation (A) and glucose-subtracted spectra used for the modified-HLA measurement of lactate (B).
Figure IV-13. Lactate concentration profiles as determined by the standard NAS method (A) and the modified-HLA method (B) models over the full spectral range (4900-4200 cm\(^{-1}\)) and NAS (C) and modified HLA (D) models over the pH insensitive spectral range (4500-4200 cm\(^{-1}\)).
HLA model results presented in Figure IV-13D provide the best results and minimal prediction bias by eliminating the glucose information from sample spectra which matches constituents of the background spectra.

Table IV-2 summarizes results from the standard NAS, modified-HLA and PLS calibration models generated for lactate. For these measurements, the PLS models were constructed by using four latent variables over the two spectral ranges discussed above. Overall, the PLS models give the lowest prediction errors and is relatively insensitive to the spectral range, although the lowest prediction error were obtained with PLS and the narrower spectral range. In contrast, the inaccuracies of the NAS model are evident by huge SEP values, as expected with the systematic offsets. The modified-HLA result is strongly influenced by the spectral range with prediction errors dropping from 51.19 mM to 2.64 mM and the MPE dropping from 23.3% to 1.2% when moving from the wide to narrow range. Because the reference error of the YSI analyzer is 2.0%, both the optimized PLS and modified-HLA models give comparable measurement capabilities as the reference analyzer with MPE’s below 2.0%. In addition, Figure IV-14 shows the calibration vectors for the modified HLA and PLS models and the similarity of these vectors is noted.

Glucose Measurements across Multiple Fermentations

The ability of the NAS and PLS calibration models to predict the concentration of glucose across fermentations was tested. In this case, the calibration models generated from the first fermentation (Trial 1) were used to predict the concentrations of glucose during the subsequent three trials. Results are present as concentration profiles in Figures IV-13, 14, and 15, for Trials 2, 3, and 4, respectively for both the NAS and PLS models. Overall, the NAS glucose concentration predictions match the reference YSI points well throughout each of these subsequent fermentations. In comparison, poorer matches are noted for the PLS profiles, where a significant off-set is indicated for fermentation Trial 2.
Table IV-2. Comparison of NAS, modified-HLA and PLS calibration models for lactate concentration measurements within a single fermentation.

<table>
<thead>
<tr>
<th>Model</th>
<th>NAS</th>
<th>Modified-HLA</th>
<th>PLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral Range (cm⁻¹)</td>
<td>4900-4200</td>
<td>4500-4200</td>
<td>4900-4200</td>
</tr>
<tr>
<td>Factor No.</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>SEC (mM)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SEP (mM)</td>
<td>54.22</td>
<td>49.73</td>
<td>51.19</td>
</tr>
<tr>
<td>MPE (%)</td>
<td>40.8</td>
<td>118.3</td>
<td>62.2</td>
</tr>
</tbody>
</table>
Figure IV-14. Comparison of the calibration vectors for PLS (blue) and modified-HLA (red) models for lactate in the UF sampled fermentation broth.
Figure IV-15. Glucose concentration profiles for fermentation Trial 2 as determined from NAS (A) and PLS (B) calibration models generated from data collected during fermentation Trial 1.
Figure IV-16. Glucose concentration profiles for fermentation Trial 3 as determined from NAS (A) and PLS (B) calibration models generated from data collected during fermentation Trial 1.
Figure IV-17. Glucose concentration profiles for fermentation Trial 4 as determined from NAS (A) and PLS (B) calibration models generated from data collected during fermentation Trial 1.
and systematic differences are evident at high (negative bias) and low (positive bias) glucose concentrations for Trials 3 and 4. In these latter two trials, a negative bias is observed at high glucose concentrations and a positive bias is evident at low glucose concentrations.

Results are summarized in Table IV-3 for the NAS and PLS calibration models for the measurement of glucose during fermentation Trials 2-4. Small SEP values are reported for the NAS calibration models, which is consistent with the observations from Figures IV-15, 16, and 17. SEP values for the NAS model range of 2.5 mM to 3.3 mM and the MPE values are 3.1 to 3.5 %, which is slightly higher than that for the reference YSI analyzer but is acceptable for industrial bioprocess monitoring. In contrast, higher SEP values are recorded for the PLS models, which is an indication of inaccuracy in the model predictions. PLS prediction errors range from 5.45 to 28.2 mM. These higher SEP values demonstrate the sensitivity of PLS models to the exact physical and chemical composition of the sample, variations in the ambient conditions and changes in the instrumentation. The NAS calibrations demonstrate lower sensitivity to such parameters, which is an attribute across these fermentations.

Lactate Measurements across Multiple Fermentations

The concentrations of lactate in the broth during fermentation Trials 2-4 were determined on the basis of the relationship between glucose and lactate in cellular metabolism. In assessing the concentration of lactate, the following reaction was assumed:

\[ \text{C}_6\text{H}_{12}\text{O}_6 \xrightarrow{\text{Cell}} 2\text{CH}_3\text{CHOCHOH} \]  

Equation IV-2

where two moles of lactic acid are produced for each mole of glucose consumed. In addition, the production of lactic acid will lower the solution pH and require a specific
Table IV-3. NAS and PLS calibration model results for glucose concentration prediction across fermentations.¹

<table>
<thead>
<tr>
<th>Trial</th>
<th>Model</th>
<th>Factor No.</th>
<th>SEP (mM)</th>
<th>MPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 (day 4)</td>
<td>NAS</td>
<td>6</td>
<td>2.73</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>PLS</td>
<td>5</td>
<td>28.25</td>
<td>192.0</td>
</tr>
<tr>
<td>3 (day 6)</td>
<td>NAS</td>
<td>6</td>
<td>3.28</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>PLS</td>
<td>5</td>
<td>5.45</td>
<td>10.0</td>
</tr>
<tr>
<td>4 (day 10)</td>
<td>NAS</td>
<td>6</td>
<td>2.52</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>PLS</td>
<td>5</td>
<td>6.59</td>
<td>28.4</td>
</tr>
</tbody>
</table>

¹Calibration models generated from information collected during fermentation Trial 1.
volume of 6 M sodium hydroxide to maintain the pH at the targeted value (5.8) and forming sodium lactate. The addition of sodium hydroxide will dilute the medium and lower the molar concentrations of the medium constituents, including glucose and lactate.

Considering both metabolite production of lactate and dilution with the addition of sodium hydroxide, the concentration of lactate was calculated by Equation IV-3:

$$[L]_{t_2} = \frac{([L]_{t_1} + 2 \times ([G]_{t_1} - [G]_{t_2}))}{(1 + 2 \times ([G]_{t_1} - [G]_{t_2})/([NaOH]))}$$

Equation IV-3

where $[L]_{t_1}$ and $[G]_{t_1}$ are the concentrations of lactate and glucose, respectively, at time $t_1$, $[L]_{t_2}$ and $[G]_{t_2}$ are the concentrations of lactate and glucose, respectively, at time $t_2$ during the fermentation, and $[NaOH]$ is the concentration of sodium hydroxide used for making for pH adjustments (6 M in this work). As the $[L]_{t_0}$ and $[G]_{t_0}$ at start time $t_0$ are known and the $[G]_{t_i}$ can be obtained at time $t_i$ from the NAS calibration model described above, the concentration of lactate at time $t_i$ $[L]_{t_i}$ can be calculated with Equation IV-3. Such calculated lactate concentrations were compared to the lactate concentrations predicted from the PLS and modified-HLA models described above.

Validity of Equation IV-3 and the process described above was tested by calculating the concentration of lactate for each spectrum in the first fermentation (Trial 1) for which YSI concentrations were measurements. Figure IV-18 compares in situ lactate concentrations determined from the YSI measurements and these calculated values. The excellent agreement denoted in this figure confirms the validity of this approach for fermentation Trials 2-4.

As was done for glucose concentration predictions, the ability to predict the concentration of lactate during fermentation Trials 2-4 was assessed when the calibration was performed with data collected during the initial fermentation (Trial 1). Table IV-4
Figure IV-18. Comparison of lactate concentrations from theoretical calculation and *in situ* measured with YSI analyzer through Trail 1.
Table IV-4. Modified-HLA and PLS calibration model results for lactate concentration prediction across fermentations.\(^1\)

<table>
<thead>
<tr>
<th>Trial</th>
<th>Model</th>
<th>Factor No.</th>
<th>SEP (mM)</th>
<th>MPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 (day 4)</td>
<td>Modified-HLA</td>
<td>4</td>
<td>3.02</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>PLS</td>
<td>4</td>
<td>11.26</td>
<td>6.6</td>
</tr>
<tr>
<td>3 (day 6)</td>
<td>Modified-HLA</td>
<td>4</td>
<td>4.70</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>PLS</td>
<td>4</td>
<td>13.19</td>
<td>4.3</td>
</tr>
<tr>
<td>4 (day 10)</td>
<td>Modified-HLA</td>
<td>4</td>
<td>4.81</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>PLS</td>
<td>4</td>
<td>13.00</td>
<td>4.1</td>
</tr>
</tbody>
</table>

\(^1\)Calibration models generated from information collected during fermentation Trial 1.
summarizes the findings and gives an indication of the stability of the modified-HLA and PLS models for lactate. As listed in Table IV-4, prediction errors for the modified-HLA model range from 3.0 to 4.8 mM and the corresponding MPE values are 1.6% to 1.9%. These percent errors are similar to those determined for the NAS models for glucose, as presented above.

As for glucose, the PLS models exhibit off sets that are evident by higher SEP values. As noted in Table IV-4, prediction errors are 11.3 to 13.2 mM for PLS models, which are significantly higher than those for the modified-HLA lactate model. The concentration profiles in Figures IV-19, 20 and 21 illustrate the off sets for the PLS model and the improved prediction accuracy for the modified-HLA model.

Conclusions

The experimental results presented in this chapter demonstrate the potential of simultaneous monitoring glucose and lactate during microbial fermentations. Calibration models were constructed for the measurement of glucose by using the PLS and NAS methods and both provide low measurement errors when applied to the fermentation used to build the calibration model. For concentration predictions outside this calibration process, the NAS model provided better prediction ability as indicated by lower SEP and MPE values for three subsequent fermentations. The superior performance of the NAS model is attributed to the fact that the NAS calibration vector is derived directly for the spectral features of the analyte and, therefore, is less susceptible to variations in instrumental and environmental parameters.

The best method for lactate measurements uses the modified-HLA method where the impact of glucose is removed by spectral subtraction. The spectral range for this modified-HLS model was optimized to reduce sensitivities to pH variance and to lower the impact of spectral noise. Based on glucose-subtracted sample spectra, prediction of the concentration of lactate from modified-HLA model accurately matched with reference
concentrations for fermentations outside the initial fermentation trial used to generate the calibration model. SEP and MPE values for the modified-HLA model were 3-5 mM and below 2.0% which is lower than those of the PLS model.

MPE values for measurements across fermentation trials were in the range of 1.6~3.5% for both glucose and lactate when using the NAS and HLA methods of analysis. These levels of measurement accuracy are sufficient for many industrially relevant fermentation processes. Overall, the NAS and HLA model methods proved to be simple and robust for the simultaneous measurements of glucose and lactate during microbial fermentations.
Figure IV-19. Lactate concentration profiles for fermentation Trial 2 as determined from modified-HLA (A) and PLS (B) calibration models generated from data collected during fermentation Trial 1.
Figure IV-20. Lactate concentration profiles for fermentation Trial 3 as determined from modified-HLA (A) and PLS (B) calibration models generated from data collected during fermentation Trial 1.
Figure IV-21. Lactate concentration profiles for fermentation Trial 4 as determined from modified-HLA (A) and PLS (B) calibration models generated from data collected during fermentation Trial 1.
CHAPTER V

IN VITRO GLUCOSE MEASUREMENT IN WHOLE BLOOD WITH
NEAR-INFRARED SPECTROSCOPY

Introduction

The measurement of blood glucose concentrations in the neonatal intensive care unit (NICU) is critical to monitor and control hyperglycemia and hypoglycemia. Although hyperglycemia should be avoided, neonatal hypoglycemia can cause permanent brain damage and death.\textsuperscript{183-184} Although conventional test-strip glucose meters are commonly used in the hospital setting,\textsuperscript{185} these devices are not designed for NICU blood glucose measurements and, therefore, are not ideal for this application. The primary concern is measurement accuracy, especially for the preterm infants in the NICU. Accuracy can be impacted by known differences in adult and neonatal blood.\textsuperscript{186-189} For example, higher blood oxygen tensions for newborn can produce negative effects on blood glucose meter results. High hematocrit levels can also lead to systematically low glucose results in the NICU. In addition, the physical act of collecting a representative blood sample by squeezing an arterial catheter in a newborn without adequately flushing can cause misleading glucose readings. Finally, thresholds for the diagnosis of hyperglycemia and hypoglycemia are at the limits of accuracy for most of the blood glucose meters. These complications encourage the development of new technology for this critical clinical application.

A special situation for all NICU patients, but especially for very low birth weight neonates, is the need to minimize phlebotomy blood loss.\textsuperscript{190-193} This patient population is small and the total volume of blood is limited. The need to draw blood samples to monitor care and improve clinical outcomes is often counterbalanced by the goal of avoiding red blood cell (RBC) transfusions, which are known to lead to adverse health
effects. Ideally, the blood chemistry information could be obtained without sacrificing blood volume.

Near infrared spectroscopy offers an innovative method for measuring blood glucose in the NICU patient population. One possible measurement configuration would be to draw a sample of blood out of the body through an indwelling arterial catheter, collect a NIR spectrum by passing the NIR radiation through the blood while it is located within the catheter tubing and then return the blood to the patient after the spectrum has been collected. Although this approach would permit collection of blood glucose information without loss of blood volume, it requires the ability to measure glucose concentrations in whole blood in a non-destructive manner. The objective of the research described in this chapter is to establish the feasibility of glucose measurements in whole blood and to characterize the impact of hematocrit levels on such measurements.

In the work presented here, measurements are made with whole bovine blood placed in a research grade FT-NIR spectrometer equipped with a high intensity source and thermoelectrically (TE) cooled InGaAs detector, designed for high signal-to-noise measurements. Spectra are collected with the blood flowing through a 1 mm path length thermostatted optical cell composed of sapphire windows, as opposed to the catheter tubing described above. PCA is used to characterize the scattering and glucose information in spectra collected from whole blood samples prepared with different hematocrit levels. PLS regression coupled with cross-validation and optimized spectral range is used to establish the feasibility of measuring glucose in a whole blood matrix.

Experimental Section

Sample Preparation

Bovine blood was obtained from a single animal and mixed with 2 g/L EDTA as the anti-coagulant. Glucose metabolism was halted by adding 3 g/L sodium fluoride (NaF). Measurements with a YSI analyzer indicated that the concentration of
glucose in the treated whole blood reached a stable level within two hours after adding NaF. Half of the treated blood was centrifuged in order to separate the plasma and RBC’s. These components, plasma and RBC’s, were used to prepare blood samples with different levels of hematocrit. Four different hematocrit levels were prepared by adding either plasma or RBC’s to the non-centrifuged blood. The concentration of glucose in each final sample was determined with a YSI model 2300 STAT plus analyzer (Yellow Springs Instrument Co., Yellow Springs, OH). Hematocrit levels were measured with micro-hematocrit capillary tubes.

Glucose was purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) and added as dried solid to the blood samples as needed to vary its concentration. 20 samples were prepared with the non-centrifuged blood to which solid glucose was added to give concentrations from 1 to 40 mM. An additional 10 samples were prepared over the same concentration range with different levels of hematocrit by adding either plasma or RCB’s.

Instrumentation and Apparatus

Figure V-1 shows the experimental configuration used in this study. Basically, blood samples were pumped through the flow through cell while spectra were collected with a Nicolet 6700 FTIR spectrometer (Thermo Fisher Scientific Inc., Waltham, MA). A 50 W tungsten-halogen light bulb with an integrated gold coated reflection mirror (Gilway Technical Lamp Inc., Peabody, MA) was used as the light source. An independent E3633A 200 W DC power supply (Agilent Technology Inc., Van Nuys, CA) supplied the lamp with 49.2 Watts of power (12.00 V and 4.10A). Light from the sources was collimated, passed through the interferometer of the 6700 spectrometer and then focused into a 3 mm diameter sapphire rod. This rod directed the light to a flow cell composed of sapphire windows with a 1-mm optical path. Light transmitted through the blood sample was captured from a second 3-mm sapphire rod that directed the collected transmitted light to the detection optics.
Figure V-1. Schematic diagram of experimental configuration used to collect NIR spectra of the whole blood samples.
A K-band interference filter (Barr and Associates, Westford, MA) was positioned before
the sapphire rod interface to restrict the collected spectra to the combination region
(5000-4000 cm\(^{-1}\)). A four-stage TE cooled extended-wavelength InGaAs detector
(Judson Technologies Inc., Montgomeryville, PA) was used as for detection. The
detector element had a surface diameter of 2 mm.

Experimental Procedure and Spectral Collection

Blood samples were grouped with four different hematocrit levels and spectra
were collected for each group in a random order. For samples within the same hematocrit
group, spectra were collected randomly with respect to the concentrations of glucose. A
peristaltic pump was used to flow each sample through the flow-through cell. The flow
arrangement was set so the blood entered the bottom of the cell and exited the top. This
flow path was used to reduce the impact of cells settling within the flow cell. In total, 15
sequential 32-coadded, 2048 points double-sides interferograms were collected for each
sample and the OMNIC software (Version 6.2, Thermo-Nicolet) was used to implement a
triangular apodization, Mertz phase correction and one level zero-filling to each
interferogram before conversion to a single-beam spectrum. Reference spectra of blood
void of glucose were collected before and after all samples with the same hematocrit
level for absorbance calculations. The absorbance spectra were calculated by taking the
negative logarithm of the ratio of the single-beam spectra for a sample relative to the
most recently collected reference spectrum. All single-beam spectra were processed by
the OMNIC software to provide a resolution of 16 cm\(^{-1}\).

Data Processing

Matlab R2007a (The MathWorks, Inc., Natick, MA) was utilized for all spectral
data analysis. PCA models were established and evaluated based on absorbance spectra
and PLS models were built with optimized spectral range and factor numbers for glucose
concentration measurements.
Results and Discussion

Spectral Quality

Figure V-2 combines the single-beam spectra collected on all blood samples. These single-beam spectra are easily grouped by hematocrit level where spectra for samples with lower hematocrit levels having higher optical throughput owing to less scattering. Many parameters of the scattering RBC’s, such as size, shape, aggregation and volume, are known to affect the light scattering properties of blood. The spectra in Figure V-2 clearly illustrates the major impact of scattering on the attenuation of the NIR radiation through these samples. Experimentally, it was determined that the rate of blood flow through the optical cell affected the settling rate for the RBCs’. Changes in the blood flow rate resulted in changes in the intensity of the measured single-beam spectra. For this reason, care was taken to ensure a constant flow rate across all experiments.

Spectral quality was characterized as the RMS noise of 100% lines. The RMS noise level was calculated over the 4600-4500 cm\(^{-1}\) spectral range. Figure V-3A shows the distribution of the measured RMS noise levels. No systematic difference is noted between the individual spectral groupings for each sample (first/second, second/third, and first/third). Different RMS noise values were obtained for each of the four groups of hematocrit levels. These different noise levels can be seen in Figure V-3A where blood with hematocrit levels of 27.0, 16.7, 11.8, and 7.9 are grouped by sample sequence as 1-100, 101-150, 151-200, and 201-250, respectively. Table V-1 lists the average RMS noise for each group and Figure V-3B shows a plot of these noise levels as a function of hematocrit level. An increase in the hematocrit level results in more scattering, which results in a reduction in the radiant power at the detection, a lowering of the signal-to-noise ratios, and higher levels of 100% line noise. On the basis of this analysis, a higher density of RBC’s will result in a degradation in performance of the analytical spectroscopic measurements.
Figure V-2. Single-beam spectra collected for blood samples with different hematocrit levels and glucose concentrations.
Figure V-3. Distribution of RMS noise levels for 100% lines for all blood spectra (A) and correlation between RMS noise and hematocrit levels (B).
Table V-1. Average RMS noise levels for whole blood samples composed of different hematocrit levels.

<table>
<thead>
<tr>
<th>Hematocrit Level (%)</th>
<th>RMS noise level (µ AU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>27.0</td>
<td>14.08</td>
</tr>
<tr>
<td>16.7</td>
<td>8.16</td>
</tr>
<tr>
<td>11.8</td>
<td>5.89</td>
</tr>
<tr>
<td>7.9</td>
<td>4.60</td>
</tr>
</tbody>
</table>
Absorbance spectra are presented in Figure V-4A for each of the blood samples. The overall variation in these spectra is based on the combination of scattering and absorption properties of the blood samples. Because the reference spectrum is a sample of blood with the same hematocrit, the absorption properties of water are removed. Others have related NIR glucose measurements on the impact of glucose on the scattering properties of the sample on the basis of changes in the refractive index of the solution surrounding the scattering body. Such effects might be contributing to the variations observed in Figure V-4A. Overall, it is difficult to judge how much the hematocrit level influences these absorbance spectra. Figure V-4B compares absorbance spectra for samples with different hematocrit levels but similar glucose concentrations. The impact is evident from these spectra, but the impact appears nonlinear.

PCA of Blood Absorbance Spectra

PCA was performed on the full set of absorbance spectra collected for the 50 blood samples. Principal component (PC) factors were calculated by singular value decomposition and the factors were ordered by degree of spectral variance explained. The first 10 PC loading vectors are presented in Figure V-5. The first PC loading vector incorporates 99.7% of the overall variance including scattering and its shape resembles the broad intensity profile as would be expected for scattering. Additional PC vectors explain the remaining spectral variance and the shapes of these vectors become nosier for the higher factors. Figure V-6 (A) shows the eigenvalues calculated for the first 13 PC vector. Clearly, the first PC vector dominates by explaining 99.7% of the spectral variance.

The PC score plot in Figure V-6B shows the magnitude of each spectrum projected onto the first and second PC vectors, presented in Figures V-5. Again, hematocrit level is a major source of spectral variance, particularly comparing samples
Figure V-4. Calculated absorption spectra of all whole blood samples (A) and blood samples with different hematocrit levels and similar glucose concentration range (B).
Figure V-5. PCA loading vectors from an analysis of all whole blood spectra with different hematocrit levels.
Figure V-6. Eigen value plot for the first 12 PC vectors indicating the fraction of spectral variance explained by each factor (A) and principal component score plot (B).
with hematocrit levels above 27%. This result matches the absorbance spectra with different hematocrit levels presented in Figure V-4B.

PLS Analysis for Glucose Measurements

Spectral range was optimized by a grid-search where the size and position of the spectral window was varied systematically across the frequency range. For each start and stop wavenumber pair, PLS calibration models were generated by using a leave-one-sample-out cross validation. In this case, all 15 spectra for each blood sample were separated for monitoring purposes, and the remaining data set was used to build the PLS calibration model. In total, 50 independent PLS models were established by this method and all predicted glucose concentrations for the left-out samples were combined to give a single cv-SEP for that spectral window for optimization. The spectral range with the minimum cv-SEP and a reasonable number of factors was selected as optimal. Figure V-7 shows the surface response plot of cv-SEP for 9 factor models as a function of the start and stop window frequencies. The minimal cv-SEP is 0.95 mM for this data set and was obtained with a spectral range of 4280 – 4740 cm\(^{-1}\). Based on the knowledge of the glucose absorbance in combination region, all major NIR features centered at 4750, 4400, and 4300 cm\(^{-1}\) are included in this optimized spectral range for the PLS modeling.

Figure V-8 presents the concentration correlation plot for the prediction of glucose concentrations in these blood samples. This PLS model was based on the input absorbance spectra with no pre-processing and computation of the PLS regression vector over the optimized spectral range. Hematocrit levels for these samples are indicated by color in Figure V-8. Table V-2 lists the SEP values calculated for each hematocrit group. No systematic variation in prediction error could be evident as a function of hematocrit level.
Figure V-7. Surface response plot of cv-SEP values of PLS models with 9 factor models for spectral range optimization.
Figure V-8. Concentration correlation plot for the prediction of glucose concentration from a leave-one-sample-out PLS cross validation model using an optimized spectral range.
Table V-2. cv-SEP values from the PLS glucose calibration model based on absorbance spectra for bovine blood samples with different hematocrit levels.

<table>
<thead>
<tr>
<th>Hematocrit Level (%)</th>
<th>27.0</th>
<th>16.7</th>
<th>11.8</th>
<th>7.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEP (mM)</td>
<td>1.22</td>
<td>0.77</td>
<td>0.67</td>
<td>1.31</td>
</tr>
</tbody>
</table>
PLS Calibration Model for Glucose with Single-beam Spectra

Figure V-2 shows single-beam spectra for blood samples with different hematocrit levels. Water is the dominant spectral absorber in these spectra and the high concentration of water overshadows the absorption features of all other sample constituents. The scattering properties from varying hematocrit levels also have a major impact on the intensity of these single-beam spectra, which can significantly impact results from the PLS model for the quantification of glucose. The same procedure described above was used to generate the PLS calibration model from these single-beam spectra. The optimized spectral range was determined to be 4450-4900 cm\(^{-1}\), and the cv-SEP for this model was found to be 2.58 mM, which is significantly higher than that for the PLS model based on absorbance spectra. This optimized spectral range does not include the two major absorption bands of glucose centered at 4400 and 4300 cm\(^{-1}\), which is a strong indication that this model is not based on spectral feature of glucose, but a secondary effect of glucose on either the spectroscopic measurement or the sample composition. Figure V-9 presents the concentration correlation plot for this single-beam PLS calibration model for glucose. Although the predicted glucose concentrations follow the unity line, these points are highly scattered and this degree of scatter is consistent with the large cv-SEP value.

To improve the non-analytical nature of the PLS model described above for single-beam spectra, multiplicative scatter correction (MSC) was attempted to reduce the influence of RBC scattering. MSC is a common pre-processing method for scattering samples.\(^{204-207}\) In this method, the objective is to identify a region of the spectrum that does not contain chemical information and is only influenced by light scattering. In practice, the full spectral range for all sample spectra is used to estimate the average spectrum for the raw spectral data through a linear regression. The slopes and offsets from this regression analysis are used to adjust the original raw spectra to the same scatter level.
Figure V-9. Concentration correlation plot of the predicted glucose concentrations of the PLS model with leave-one-sample-out cross validation and optimized spectral range based on single-beam spectra.
In this analysis, the MSC calculation was applied using Equations V-1 and V-2:

\[ x_j = \beta_0 + \beta_1 \cdot x_j + \epsilon_j \quad \text{Equation V-1} \]

\[ x_{j,\text{MSC}} = \frac{x_{j,\text{raw}} - \beta_0}{\beta_1} \quad \text{Equation V-2} \]

where \( X_j \) is the NIR spectrum and \( \bar{X}_j \) symbolizes the mean spectrum of the calibration set. For each spectrum, \( a \) and \( b \) are estimated by ordinary least squares regression of spectrum \( X_j \) and \( \bar{X}_j \) over the full spectral range. The MSC corrected spectrum is then calculated with coefficients \( \beta_0 \) and \( \beta_1 \) according to Equation V-2.

Figure V-10 shows the MSC corrected NIR spectral for the blood samples with different hematocrit levels. The variance in the intensity is greatly reduced in comparison to the spectra presented in Figure V-2. The PLS model based on these MSC corrected spectra, used an optimized spectral range of 4200-4650 cm\(^{-1}\) and 9 factors. Unlike the PLS model based on raw single-beam spectra, this spectral range includes the spectral features associated with glucose. The cv-SEP for this MSC-based PLS model is 0.83 mM, which marks a significant improvement compared to the raw single-beam spectra. In fact, this cv-SEP is lower than the value obtained for the PLS model based on absorbance spectra. The corresponding concentration correlation plot is presented in Figure V-11. This plot shows excellent agreement with the unity line across the full concentration range with no obvious correlation with hematocrit levels. The cv-SEP values are listed in Table V-3.

Figure V-12 compares calibration vectors from the PLS models based on single-beam and absorbance spectra. Both calibration vectors are based on 9 factors and show major features centered at 4400 and 4300 cm\(^{-1}\) as would be expected for the measurement of glucose.
Figure V-10. Single-beam spectra of whole blood samples after MSC pre-processing for PLS model.
Figure V-11. Concentration correlation plot of the predicted glucose concentrations of the PLS model with leave-one-sample-out cross validation and optimized spectral range based on single-beam spectra with MSC pre-processing.
Table V-3. cv-SEP values from PLS model based on single-beam spectra after MSC pre-processing for different hematocrit levels.

<table>
<thead>
<tr>
<th>Hematocrit Level (%)</th>
<th>27.0</th>
<th>16.7</th>
<th>11.8</th>
<th>7.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEP (mM)</td>
<td>1.30</td>
<td>0.76</td>
<td>0.57</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Figure V-12. Calibration vectors from the PLS models based on single-beam after MSC pre-processing and absorption spectra.
Conclusions

The findings presented in this chapter demonstrate the potential to quantify glucose in whole blood with varying hematocrit levels with NIR spectroscopy. Effective calibration models are illustrated for both single-beam and absorbance NIR spectra. For each type of spectra, high calibration accuracy is demonstrated by excellent linearity between predicted concentrations and reference values. The cv-SEP values for both PLS models are in the range of 0.8-0.9 mM and the optimized spectral ranges include glucose absorption features. For the single-beam spectra, the MSC calculation greatly reduces the spectral variance associated with RBC scattering, thereby lowering sensitivity to hematocrit levels.

These results are encouraging that both types of spectral data can be used for glucose concentration measurements in blood samples with different hematocrit levels. Of particular note, the MSC calculation results in a similar level of performance between PLS models based on single-beam and absorbance spectra. By eliminating the scattering effects and building proper calibration models, this study supports the further application of NIR spectroscopy for the quantification of glucose in blood sample from newborn infants and NICU patients.
CHAPTER VI

FUTURE WORK

Throughout this dissertation, new advances are presented for glucose monitoring by NIR spectroscopy. Advances range from a temperature insensitive method for measurements in bovine blood ultrafiltrate to glucose measurements in whole bovine blood with different densities of RBC’s. Future research will focus on applying, expanding and adapting these methodologies to real-world applications. In addition, efforts will focus on further optimizing and developing both hardware set-up and chemometrics algorithms for mature glucose monitoring systems.

Explore Utility of Temperature Insensitive Measurements in Complex Systems

Feasibility of the digital Fourier filtering based temperature insensitive algorithm should be expanded further by investigating data set collected from complex matrices with more spectral variances. The complex matrix could be whole blood, fermentation broth or even human skin. Source of variances could include variations associated with the source power, sample path length and detector responsivity.

For complex matrices, the major effect is more inconsistent and complicated sample components where more spectral overlap exists between glucose and other matrix components. The sensitivity and selectivity of the digital filtering method will be challenged under these conditions. The optimized parameters for digital Fourier filter could be improved in order to filter interfering spectral features which could originated from absorption or scattering sources.

As the ultimate goal, additional experiments must be directed toward miniaturizing the NIR spectrometer. As described in Chapter 2, the lower resolutions commonly associated with miniaturized spectroscopic instrumentation has no negative influence on the ability of the digital filtering method to eliminate the impact of
temperature variance. Other parameters of miniaturized spectrometers might degrade analytical performance, such as higher noise levels, varying detector responsivity, narrower spectral range and wavelength shifts. These parameters could have significant adverse effects on the analytical measurements. Therefore, the feasibility of the digital filtering as a robust way for eliminating temperature variance demands further investigation.

**Simulations with More Variances and Corresponding Calibration Modeling**

The challenge of accurate glucose monitoring outside the calibration process for noninvasive glucose sensing across human skin is mostly from two sources: first, too much interfering from other constituents such as water, fat, proteins and other scattering components. The concentrations of these matrix components are typically higher than that of glucose and their concentrations of these matrix components can vary with time. Second, the effective optical path length through human skin tissue can vary considerably, particularly between different days and different people. Such variations in the sample can lead to large errors in the predicted glucose concentrations.

As a simplified research tool to mimic the components varying in the skin and further establish calibration monitoring for noninvasive glucose sensing, more components such as proteins and fat must be added to simulations to reach optimal experimental conditions. Meanwhile, the path length of the spectral measurement could vary within different experimental conditions which would represent the varying skin thickness over different days of a measurement.

Based on spectral data collected with more components and optical path lengths, a more complicated and robust multivariate calibration model could be established. If necessary, pre-processing for scattering correction, algorithms for temperature variance elimination and path length correction, determination of internal or external standards
could be applied to develop a robust calibration method with verification of any chemical and physical correlations that impact measurement accuracy.

**Direct Measurements in Fermentation Broth**

Improvements in simultaneous measurements of multiple components during fermentations can be envisioned by using a fiber-optic probe to collect NIR spectra directly in the fermentation broth. Such an optical probe would eliminate the time delays created by the UF probes used here. The ability to collect NIR spectra with high signal-to-noise ratios, to control the optical path length at the probe tip and to prevent biofouling would be required for successful implementation. A NIR probe of this nature would clearly be valuable for numerous industrial applications and fit within the spirit of the FDA's Process Analytical Technology Initiative.

The UF probe used in this dissertation for fermentation monitoring provides a sample that is void of the metabolizing cells located in the fermentation broth. If this UF probe is eliminated and measurements are performed directly in the broth, issues of light scattering by the cells will be critical, and perhaps limiting. The impact of cells within the optical path will be especially important when the cell density reaches relatively high levels. As demonstrated in this dissertation, the negative impact of scattering can be minimized by using certain pre-processing methods, such as standard normal variate (SNV) and multiplicative scatter correction (MSC). Such methodologies have the potential to enable direct NIR analytical measurements in complex and scattering samples, such as fermentation broth.
REFERENCES


(31) Ren, M., Comparison of Near Infrared and Raman Spectroscopies for Noninvasive Clinical Measurements. *Doctoral Thesis*, The University of Iowa, **2007**.


(53) Fernandez Pierna, J.A.; Vermeulen, P.; Amand, O.; Tossens, A.; Dardenne, P.; Baeten, V., NIR Hyperspectral Imaging Spectroscopy and Chemometrics for Detection of Undesirable Substances in Food and Feed, Chemometrics and Intelligent Laboratory Systems, 2012,


(79) Moros, J.; Barciela-Alonso, M.C.; Pazos-Capeans, P.; Bermejo-Barrera, P.; Pena-Vazquez, E.; Garrigues, S.; Guardia, M., Characterization of Estuarine Sediments...


(129) Pfeiffer, E.F.; Thum, C.; Clemens, A.H., the Artificial Beta Cell - A Continuous Control of Blood Sugar by External Regulation of Insulin Infusion (Glucose Controlled Insulin Infusion System), *Hormone and Metabolic Research*, 1974, 6, 339-342.


(177) Rhiel, M.; Cohen, M.B.; Murhammer, D.W.; Arnold, M.A., Nondestructive Near-infrared Spectroscopic Measurement of Multiple Analytes in Undiluted Samples of...


