Spring 2014

Nocturnal hypoglycemic alarm based on near-infrared spectroscopy

Sanjeewa Rasika K. Ranasinghe Pathirajage
University of Iowa

NOCTURNAL HYPOGLYCEMIC ALARM BASED ON NEAR-INFRARED SPECTROSCOPY

by

Sanjeewa Rasika K Ranasinghe Pathirajage

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

May 2014

Thesis Supervisor: Professor Gary W. Small

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D THESIS

This is to certify that the Ph.D. thesis of

Sanjeewa Rasika K Ranasinghe Pathirajage

has been approved by the Examining Committee
for the thesis requirement for the Doctor of
Philosophy degree in Chemistry
at the May 2014 graduation.

Thesis Committee:

Gary W. Small, Thesis Supervisor

Mark A. Arnold

M. Lei Geng

Alexei V. Tivanski

Markus Wohlgenannt

To my parents,
who taught me the importance of
being educated

ACKNOWLEDGMENTS

First and foremost, I must thank my advisor, Professor Gary Small, for all of the opportunities and the learning experiences to become the scientist that I am now. I am truly happy about the decision I made five years ago to join the Small research group. Much of my success has come from your direction and your challenges that pushed me to a level that I would be able to accomplish. You inspired me to be a good scientist a chemist and a chemometrician.

I am also grateful to my committee members, Prof. Mark A. Arnold, Prof. Lei Geng, Prof. Alexei V Tivanski. and Prof. Markus Wohlgenannt. They offered invaluable discussion and suggestions on my work. Many thanks for the important experiences and kind help from former group members, Dr. Qiaohan Guo, Dr. Wei Wang, Dr. Chamatca Priyanwada as well as current group members, Hua Yu, Brian Dess and Ziqi Fan.

I would also like to thank our collaborators Prof. Mark Arnold for allowing me to use some of the rat animal data collected in their research laboratory. Special thanks go to Dr. Chuannan Bai, Dr. Jue Qian and Terry L. Graham for their great work in the animal studies.

The foundation of my life in chemistry came from my undergraduate university, University of Kelaniya. I am very thankful to all the faculty members of the Department of Chemistry, University of Kelaniya, Sri Lanka, for all the knowledge and support and your encouragement inspired me to pursue a Ph.D. in Chemistry and become the chemist that I am now.

I wish to thank my parents, who inspired me for further education and for being with me in every decisions I made. They were the first teachers in my life. I would like to thank my wife, Sumana Nilanthi, for her support throughout and she was with me in my every hardships. Last but not least, I would like to remind my Sweet little angel Dinali Adhya whose smiling face takes all my stress away the moment I enter home.

ABSTRACT

Noninvasive glucose monitoring has been the subject of considerable research because of the high number of diabetes patients who must monitor their glucose levels daily by taking blood samples. Among methods being evaluated for possible use in this application, near-infrared (NIR) spectroscopy has received significant attention because of available glucose absorption bands that can be observed in the presence of the large aqueous background found in tissue spectra. The objective of the research presented here is to evaluate the potential for implementing a noninvasive nocturnal hypoglycemic alarm with NIR spectroscopy. Such an alarm would be used by a diabetic to detect potentially dangerous occurrences of hypoglycemia during sleep.

The approach used is to collect spectra continuously from the patient during the sleep period, followed by the application of pattern recognition methods to determine if a spectrum represents a blood glucose level that exceeds a hypoglycemic threshold. A reference spectrum is collected and a conventional finger-stick glucose concentration measurement is made at the start of the sleep period. The ratio is then taken of each subsequent spectrum to the collected reference, forming a differential spectrum corresponding to the signed difference in concentration relative to the reference. The identification of these differential spectra as “alarm” or “non-alarm” is performed with a classification model computed with piecewise linear discriminant analysis.

This methodology is initially tested with *in vitro* laboratory data that simulated the glucose excursions that occur during sleep. The performance of the hypoglycemic alarm methodology in the presence of varying levels of urea, glyceryl triacetate, and L-lactate

as potential spectral interferents is tested. The robustness of the methodology with respect to time is also evaluated.

The thesis further discusses an experimental procedure to prepare tissue phantoms composed of two main proteins that exist in human skin tissue, keratin and collagen. A new methodology is developed to produce varying-thickness films that allowed the simulation of changes in the content of skin tissue proteins present within the optical path of the NIR measurement. The prepared films are incorporated into *in vitro* laboratory measurements in which varying levels of glucose, urea, keratin, and collagen are introduced in order to provide a test of the hypoglycemic alarm algorithm that simulates the spectral properties of human tissue.

Finally, the hypoglycemic alarm algorithm is tested with *in vivo* data collected with rat animal models. Data are presented for single-day experiments performed with anesthetized rats, as well as for multiple-day experiments conducted with awake rats. The results obtained from both the *in vitro* and *in vivo* studies confirm that if high-quality spectral data are attainable, the alarm methodology can work effectively to identify hypoglycemic events while exhibiting a low rate of false detections.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

Diabetes mellitus is a chronic disease whereby a person is unable to metabolize glucose effectively. According to the International Diabetes Federation (IDF) around 382 million people worldwide suffered from diabetes in 2013 and this number is projected to increase to 592 million by the year 2030.[1] In the United States (US), 25.8 million children and adults have diabetes.[2] Type 1 diabetes, generally diagnosed in children and young adults, results from the body's failure to produce insulin, the hormone that allows glucose to be taken from the blood into the cells and stored or used as energy. This form of diabetes is generally termed juvenile diabetes. Only 5 % of the people with diabetes have this form of the disease and with the help of insulin therapy, individuals with Type 1 diabetes can live long lives.[2]

Type 2 diabetes arises from the body's inability to use the produced insulin properly. This is the most common form of diabetes.[3] This type is more common among adults but increasingly affects children as childhood obesity increases. This form of diabetes is manageable by eating well, exercising, and maintaining a healthy weight. Diabetes medications or insulin therapy are required if the disease is not manageable with these practices.

There are many medical complications associated with diabetes such as heart disease and stroke, high blood pressure, kidney disease and blindness, etc.[4] Diabetes is chronic but a patient can prevent acute complications and reduce the risk of long-term complications by continuous medical care and by good self-management. The treatments

for diabetes are extremely costly, however. The total costs of diagnosed diabetes in 2012 in the US were estimated to be $245 billion.[5]

Both extremely higher and lower blood glucose levels lead to the health problems associated with diabetes. Hyperglycemia occurs when an excess of glucose is circulating in the blood and is diagnosed when the fasting blood glucose level is ≥126 mg/dL or 7.0 mM after two consecutive blood tests.[6] Hypoglycemia involves abnormally diminished levels of glucose in the blood, generally defined as < 72 mg/dL or 4.0 mM. The symptoms of hypoglycemia usually start to appear when the blood glucose readings drops below 54 mg/dL or 3.0 mM.[7] Nocturnal hypoglycemia or nighttime hypos are common in diabetics who treat their disease with insulin. The symptoms related to nocturnal hypoglycemia are usually only realized once the patient has wakened, and this is particularly worrying for parents of children with diabetes.

The complications of hypoglycemia can range from mild to severe. These include mild and short term weakness, dizziness, or serious damage to brain power and ultimate death. Most Type 2 diabetics are encouraged to monitor their blood glucose levels at least once per day so that they can respond appropriately to hypo- or hyperglycemia. These responses may include diet adjustments, exercise, or insulin injection. More frequent blood glucose monitoring, at least three times per day, is required for patients with Type 1 diabetes and many with Type 2 in order for these patients to assess the effectiveness of their prior insulin dose and to help determine their next insulin dose.

Frequent monitoring of blood glucose levels is thus required for many diabetic patients to allow effective management of the disease. Self-monitoring means the patient measures their blood glucose at home or on the go with a glucose meter which consists of

a lancing device, lancet, test strip and a meter. In the common procedure, a small drop of capillary blood obtained from the fingerstick or fingerprick is put on a test strip and the glucose concentration is measured by inserting the test strip into a glucose meter.

Although self-testing is considered to be a key development in glucose monitoring, it has some major disadvantages, too. The required blood sampling is associated with pain and risk of infection which discourages patients from frequent monitoring. Also the cost of the enzymatic test strips (cost of a test strip is ~ $1) limits the number of tests per day for many patients. With only a few measurements per day, an accurate assessment of the patient's overall fluctuations of the blood glucose levels is not possible.

Because of these drawbacks in self-monitoring methodologies, a continuous monitoring approach is preferred which better reflects the overall fluctuations of the glucose levels. Most of the existing continuous glucose monitoring (CGM) devices are composed of three parts: (1) an electrochemical sensor that is placed subcutaneously in the abdomen or upper arm which measures glucose levels in the interstitial fluid (ISF), (2) a receiver that shows real-time blood glucose levels, and (3) a transmitter that is connected to the sensor and transfers the signals from the sensor to the receiver at intervals of ~5 min.[8] The sensor stays in place for several days to a week and then must be replaced.

The transformation of the electrochemical signal from the sensor into a glucose concentration reading is achieved by one-point or two-point *in-vivo* calibration procedures for which a reference glucose concentration is obtained by use of a conventional fingerstick method. The user must therefore check blood samples with a

glucose meter to program the CGM device. Because currently approved CGM devices are not as accurate and reliable as standard blood glucose meters, users are directed to confirm glucose levels with a conventional meter before making any changes in their insulin treatment. In addition to lower accuracy and reliability, these CGM systems have some limitations including discomfort and skin irritations. A more detailed description of self-monitoring of blood glucose and CGM is given in Chapter 2.

Truly noninvasive glucose sensing is the ultimate goal of much glucose monitoring research and is the goal of the work described in this dissertation. If realized, a noninvasive glucose measurement would eliminate the challenges, discomfort, and potential risk for infection associated with self-glucose monitoring and implantable devices. The successful development of a continuous noninvasive glucose sensor that can be used in human subjects has the potential to revolutionize the daily treatment and control of diabetes. Such a sensor would allow close monitoring of the glucose fluctuations in blood, thereby providing the necessary information required to maintain tight glycemic control by insulin intervention. Noninvasive methods are clearly preferable to invasive techniques, as long as they can meet the same clinical accuracy associated with invasive methods.

Most research directed to the development of noninvasive glucose sensing technology has focused on the use of spectroscopic methods. The concept in optical noninvasive glucose sensing is to pass a selected band of electromagnetic radiation through a vascular region of the body and then extract the glucose concentration from the resulting spectral information. In this approach, there is no direct contact between the glucose molecules and the transduction element.[9]

The work described in this dissertation focuses on the use of near-infrared (NIR) spectroscopy to develop noninvasive glucose sensing methodology. Among the potential optical approaches to noninvasive glucose sensing, NIR spectroscopy has received significant attention because of the unique NIR absorption spectrum of glucose and the significant penetration of NIR light into tissue.[10] A detailed description of the theory, applications, and instrumentation of NIR spectroscopy is provided in Chapter 2.

Spectral features in the NIR region are relatively weak and broad, leading to significant overlap of bands in complex samples. In addition, samples that contain water (e.g., any biological sample) have a broad and temperature-sensitive background absorbance that leads to significant variation in spectral baselines. These factors preclude the use of simple quantitative determinations based on measurements at a single wavelength. Multivariate calibration methods based on multiple wavelength measurements are thus required to extract weak analyte signals from the overlapping features of the spectral background.

The growth of applications of NIR spectroscopy has coincided with the parallel development of powerful chemometric techniques for use in analyzing spectral data. Chemometrics is a methodology used in analytical and measurement science that employs mathematical and statistical methods for signal processing, experimental design, and qualitative and quantitative model building. In NIR data analysis in a noninvasive glucose sensing application, chemometric techniques are used for several different purposes: to decrease the effect of the background information contained in glucose spectra (spectral pre-processing methods), to group sample components according to the features of the spectra (classification methods), or to extract a quantifiable property (e.g.,

glucose concentration) of the sample from the spectra collected (regression methods).[11] A detailed description of the data analysis methods used in the dissertation research is given in Chapter 3.

The specific objective of the research presented here was to evaluate the potential for implementing a noninvasive nocturnal hypoglycemic alarm with NIR spectroscopy. Such an alarm would be used by a diabetic patient to detect potentially dangerous occurrences of hypoglycemia during sleep. Because the patient is sleeping, a portable monitor is not required, and many engineering challenges related to the size and ruggedness of the instrumentation are eliminated. By virtue of this simplification, the dissertation work was envisioned as a key first step toward the development of practical and more generally applied noninvasive glucose monitoring methodology.

The approach used to implement the nocturnal alarm was to collect transmission spectra continuously from the tissue of the patient during the sleep period, followed by the application of pattern recognition methods to determine if a spectrum represents a blood glucose level that exceeds a hypoglycemic threshold. To initialize the monitor, a reference spectrum is collected and a conventional fingerstick glucose concentration measurement is made at the start of the sleep period. The ratio is then taken of each subsequent spectrum to the collected reference, forming a differential spectrum corresponding to the signed difference in concentration relative to the reference. The identification of these differential spectra as “alarm” or “non-alarm” is performed with a classification model computed with piecewise linear discriminant analysis (PLDA).[12,13] Chapter 3 discusses the steps behind the nocturnal hypoglycemic alarm algorithm in detail.

In the dissertation work, the alarm methodology was tested both with *in vitro* and *in vivo* spectral data that simulated the glucose excursions that occur during sleep. Chapter 4 begins with the discussion of the data collection and analysis methodologies used in an *in vitro* simulation study designed to test the feasibility of the proposed hypoglycemic alarm algorithm. This study employed a four-component dynamic system consisting of glucose, urea, lactate, and glyceryl triacetate in pH 7.4 phosphate buffer. Three peristaltic pump systems were used to simulate the glucose excursions that occur while a diabetic is sleeping, while also allowing the spectral background to be varied. Varying concentrations of urea, lactate and glyceryl triacetate were used to simulate the presence of varying levels of close spectral interferents. The developed alarm algorithm performed well in this study, giving no missed or false alarms in the presence of the interfering components.

A major source of spectral variance in the noninvasive glucose data collection in the NIR region is the tissue sample itself. The propagating photons will undergo scattering and absorption due to the heterogeneity of the chemical composition of the tissue. As a result, the chemical composition of the tissue matrix and the physical distribution of the principal chemical constituents should be taken into account in designing simulation studies.

Skin tissue contains two types of proteins, keratin and collagen. Chapter 5 discusses an approach to make a skin tissue phantom composed of keratin and collagen which could be used to simulate data collection in the human skin tissue matrix. In this work, two types of protein films were prepared in the laboratory. The thickness of each film was designed to match the equivalent amount of keratin or collagen observed in

spectra of human tissue. Initial studies were performed with films of uniform thickness. The consistency and stability of these films was evaluated, and protocols were developed for their use with the *in vitro* simulation system employed in the work described in Chapter 4. Follow-on studies were then performed in which films of varying thickness were created. This allowed the incorporation of variation in film thickness into the *in vitro* simulation studies

In the work described in Chapter 6, the noninvasive nocturnal alarm algorithm methodology was tested *in vitro* to simulate a sleep cycle and its performance was evaluated with the tissue phantom described above to simulate the spectral properties of human tissue. Two data sets were collected in this work. The data for the first dynamic system were collected using the films in the same orientation during all the days of data collection. This simulated a constant tissue matrix. Varying concentrations of urea were used to simulate the presence of a close spectral interferent. Two peristaltic pump systems were used to simulate the glucose exclusions that occur while a diabetic patient is sleeping.

Chapter 6 further discusses the results obtained for a second dynamic system collected in the presence of a tissue phantom with varying film thickness. Different combinations of orientations of the produced films were used to simulate skin tissue variations during an *in vivo* data collection, especially those that would be encountered when the tissue measurement site is varied from day to day. Varying concentrations of urea were again used to simulate the presence of a close spectral interferent.

In Chapter 7, the noninvasive nocturnal alarm algorithm methodology was tested *in vivo* with data collected from the skin tissue of a rat animal model to simulate

noninvasive human measurements. The data collected during single-day glucose clamp experiments from two anesthetized rats were tested with the developed alarm algorithm. The results obtained for these glucose transients are discussed in detail in the first part of Chapter 7. The second part of the chapter presents results for data collected from two awake rats over multiple days. In these studies, the first days of the data collection for each of the rats were used as calibration data, while the data for the remaining days were used to test the developed algorithm. Finally, the important results from this dissertation are summarized in Chapter 8, and future work to advance these projects is also proposed.

CHAPTER 2

NONINVASIVE GLUCOSE MONITORING IN NEAR-INFRARED SPECTROSCOPY

Introduction

In Chapter 1, an introduction to diabetes and a brief overview of current invasive methodology to measure blood glucose and noninvasive glucose monitoring by NIR spectroscopy were introduced. In this chapter, a more detailed discussion of these subjects is given so that the motivation for the dissertation research can be appreciated with better understanding. Motivation for the noninvasive glucose sensing technology will be presented, with a discussion of self-monitoring of blood glucose (SMBG) and continuous glucose monitoring (CGM) approaches. Both direct and indirect noninvasive glucose measurement methodologies will be discussed in detail, followed by a description of the use of NIR spectroscopy in glucose sensing. The principles, mathematical procedures, and instrumentation of Fourier transform infrared (FT-IR) spectroscopy will be presented.

Blood Glucose Testing

Although diabetes is a chronic disease, it can usually be managed successfully with a proper treatment plan and a change in lifestyle. The main goal of treatment is to keep blood glucose levels in the normal or near-normal range. More frequent checking of blood glucose is one of the best ways to know how well the patient’s diabetes treatment plan is working. A physician periodically orders a blood test to determine the diabetic’s blood glucose levels and hemoglobin A1c (A1C). This test gives a sense of how blood

glucose levels are controlled since it indicates the patient's average blood glucose level during the past two to three months. However, maintaining blood glucose levels at a normal level and successful treatment of diabetes also requires that the patient monitor his/her own glucose levels on a day-to-day basis.

Self-Monitoring of Blood Glucose

Self-monitoring of blood glucose allows the patient to know his/her own blood glucose level at any time and helps prevent the severe occurrences of very high or very low glucose levels. Self-monitoring also enables tighter glycemic control, which minimizes the long-term complications associated with diabetes. Self-monitoring employs glucose monitors that utilize test strips. The basic methodology includes a fingerprick, capillary blood collection, and use of a portable meter to obtain the glucose reading. The test strip is doped with enzymes termed oxidoreductases. These enzymes catalyze the oxidation of glucose to gluconic acid.[14]

The earliest generations of the test strips were designed for a visual evaluation.[15] With this approach, electrons generated from the oxidation reaction are transferred to an indicator molecule which in turn forms a color.[14] In the common procedure, a drop of blood is placed on the test strip, wiped off, and the color produced is compared with a color scale. However, colors are susceptible to fade and differences in visual acuteness across individuals make these glucose monitoring systems less accurate and precise.[15,16]

The concept of electrochemical glucose monitoring was proposed in 1962 by Clark and Lyons.[17] Most of the electrochemical glucose meters are amperometric enzymatic biosensors.[18,19] Glucose measurement in electrochemical sensors is based on three different strategies. The first detection methodology is based on measuring the

oxygen consumption during the oxidation of glucose; the second measures the hydrogen peroxide produced by the enzymatic reaction;[20] and the third scheme replaces oxygen with a nonphysiological (synthetic) electron acceptor to transfer electrons from the enzyme (e.g., glucose oxidase) to the electrode.[19-21] A schematic diagram of the three main strategies used with electrochemical glucose sensors is shown in Figure 2.1.[21]

These electrochemical blood glucose meters are widely applied for SMBG owing to their high selectivity and sensitivity for glucose. In addition, these meters are portable and easy to use. Various types of test strips for electrochemical glucose meters are commercially available today including Accu-Check Aviva, TRUEtrack, Precision Xtra, Freestyle Lite, Nova, Arkray, One Touch Ultra, Agamatrix, etc.

Self-monitoring using these enzyme-based electrochemical monitors has three major disadvantages or limitations: (1) the method requires blood sampling which is painful and has a risk of infection; (2) the test strips are expensive which can limit the frequency of testing for hypo- and hyperglycemia; and (3) these discrete measurements do not provide information about fluctuations in blood glucose concentrations. Current research directed to the development of the next generation of glucose monitoring devices is focusing on overcoming these limitations.

Continuous Glucose Monitoring

Continuous glucose monitoring systems use a glucose sensor implanted subcutaneously to determine the level of glucose in the interstitial fluid (ISF). The sensor measures glucose based on an electrochemical reaction and wirelessly transmits results to a small recording device. The continuous glucose monitor displays the diabetic's blood

glucose level every few minutes and thus allows monitoring of glucose fluctuations. The receiver can also be set to alarm if the patient's glycemic level is above or below a threshold value, which is highly useful to detect dangerous hypoglycemic events.

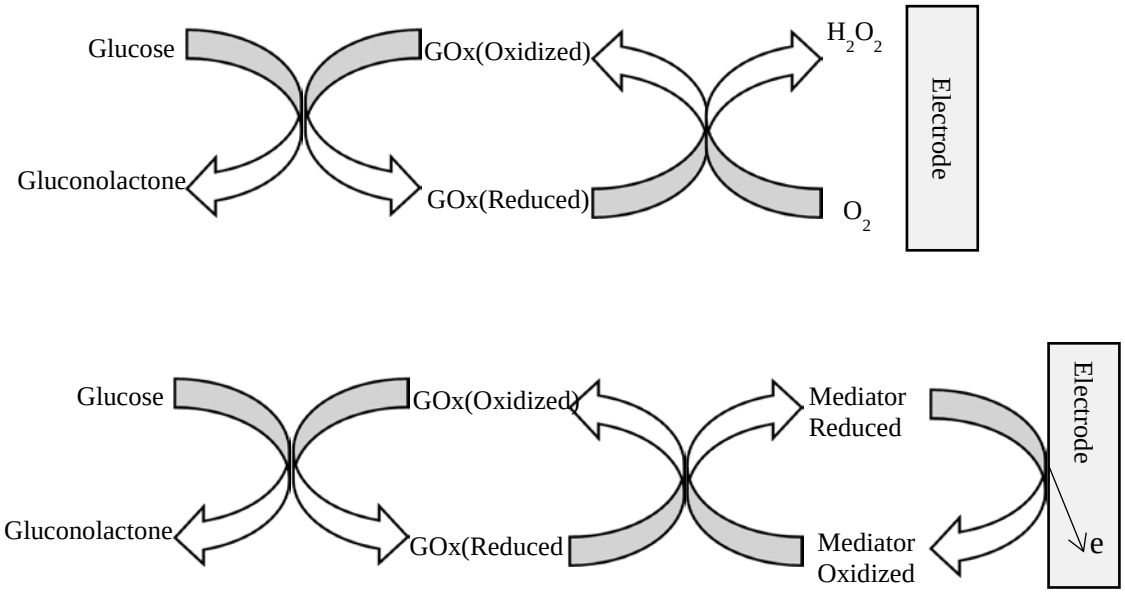


Figure 2.1 Three main strategies used in glucose measurements with electrochemical glucose sensors based on glucose oxidase (GOx). As depicted in the upper diagram, the sensor can be configured to measure either the consumption of oxygen or the production of H_2O_2. The third approach shown in the lower diagram uses a mediator molecule to transport electrons from GOx to the electrode where the electrochemical current is measured.[22]

Source: Oliver, N.; Toumazou, C.; Cass, A.; Johnston, D. *Diabetic Medicine* **2009**, *26*, 197.

The sensor should be removed and replaced in a different location on the body approximately once per week. In addition, the patient must continue to perform fingerstick glucose testing several times daily to ensure that the continuous monitor is correctly calibrated. Proper calibration helps to ensure that the measured ISF glucose

concentration accurately reflects the blood glucose level. It has been observed that the glucose levels in ISF temporally lag behind blood glucose levels.[20,23] Computer algorithms for correcting this short lag time have been developed.[24]

A key issue with these devices is still the robustness of the calibration over a period of several days. A recalibration is required to account for variations in sensitivity. In addition, these implantable biosensors are still facing issues of biocompatibility and electrochemical interferences.[19] Several implantable biosensors are available commercially, but their use is not yet widespread because of reasons such as complexity, inaccuracy, invasiveness, cost, pain, discomfort, and risk of infection.

The first CGM device approved by the United States Food and Drug Administration (FDA) was manufactured by MiniMed in 1999.[25-27] Current real-time CGM devices approved by the FDA and clinically used include: GlucoDay® (Menarini Diagnostics),[28,29] Guardian®-RT(real time) (Medtronic Minimed, Northridge, CA, 2005), Dexcom SEVEN® (Dexcom, San Diego, CA, 2007),[27,30,31] and Freestyle Navigator® (Abbott Diabetes Care, Alameda, CA, 2008)[32].

The Guardian®-RT, Dexcom SEVEN®, and Freestyle Navigator® use a subcutaneous sensor to measure ISF fluid glucose concentrations. The GlucoDay® uses microdialysis to acquire a sample for analysis.[27] In this context, a hollow dialysis fiber is commonly implanted in the subcutaneous tissue to pass an isotonic fluid across the skin. Glucose, diffusing from the tissue into the fiber, is assayed using an electrochemical method. Some research groups have developed devices for the continuous tissue glucose monitoring based on a combination of microdialysis and an electrochemical method.[33-36]

The current research focus in this area is to advance the development of what is known as an artificial pancreas or artificial pancreas device system (APDS). The proposed artificial pancreas consists of a series of devices, e.g., a CGM implanted subcutaneously to measure ISF glucose, a blood glucose measurement device for use in periodically calibrating the ISF sensor, and an insulin pump to deliver insulin as needed to achieve proper glycemic control. A computer algorithm would be in operation to communicate with all of the devices. The feedback from the glucose monitor can thus be utilized to adjust insulin levels.[37] No artificial pancreas systems have yet received FDA approval.

There are some approaches for minimally invasive or noninvasive sensors which avoid the continuous presence of a sensor inside the skin tissue.[38] Both the sensor and the sample collection device in this method are externally located. Different strategies of drawing the transdermal fluids that serve as the sample for analysis include reverse iontophoresis,[39,40]sonophoresis,[41,42] micropore technology,[43,44] microneedle technology, and a skin blister technique.[45,46]

The GlucoWatch[47] is an example of this approach and was approved by the FDA in 2001 for use by patients in tracking their glucose levels with time. It is an example of a minimally invasive near-real time continuous glucose monitoring (RT-CGM) device that is designed to be worn like a wristwatch. This device uses a low-level electrical current to extract ISF through the skin by electroosmotic flow (reverse iontophoresis). The determination of the glucose concentration in the ISF is subsequently performed with an electrochemical measurement. The method has drawbacks related to time lag between values, a complex calibration procedure which requires fingerstrick glucose

measurements, poor accuracy especially near the hypoglycemic range, and problems associated with skin irritations.[48] Despite initial interest in the device from diabetics, the GlucoWatch ultimately failed as a product and is no longer sold.

Noninvasive Glucose Monitoring

The invasive methods for glucose monitoring are chemical-based methods that require a direct contact between the transduction element of the sensor and the representative biological fluid (e.g., ISF, capillary or venous blood). However, noninvasive methods are most commonly spectroscopic methods that require no contact between the representative fluid and the sensor. Furthermore, they do not require the extraction of the representative fluids from the body. The common procedure is to probe a specific area of the body tissue with a selected portion of the electromagnetic spectrum and extract glucose information from the acquired response.[9]

The development of an accurate noninvasive CGM device would revolutionize the treatment of diabetes. These devices would provide a painless and inexpensive continuous monitoring capability, thereby giving the patient the ability to maintain better glycemic control. Ultimately, improved glycemic control should reduce the medical complications associated with diabetes and also lower the costs of health care.[49] However, despite extensive efforts by many research groups and companies, to date, no one has developed a reliable method to measure glucose noninvasively.

There are several key challenges to developing a successful noninvasive monitoring device. The device should match the accuracy obtained with enzyme-doped test strip methods, especially in the clinically important hypoglycemic concentration

range (2-5 mM).[9,50] Test-strip based methods have a relative accuracy of 6-10 %,[51] and any reliable noninvasive method must be matched with this accuracy to claim the ability to measure glucose noninvasively.

One of the key issues in considering the performance of a noninvasive measurement approach is chemical selectivity. The invasive methods achieve selectivity by the use of chemical reagents (e.g., glucose oxidase) that specifically target the glucose molecule and thereby generate a signal that is known to be a direct measurement of the glucose concentration. Because the noninvasive methods do not extract a sample from the body, they cannot employ selective chemical reagents. Selectivity must come solely from the characteristics of the measured response. For example, if the noninvasive method is based on collecting a spectrum of tissue, success depends on either (1) the presence of a spectral signature of the glucose molecule that can be reliably extracted from the overall measured response or (2) the presence of a spectral signature that, while not directly originating from the glucose molecule, is reliably and reproducibly present and correlates with changes in glucose concentration. The ability to extract these spectral signatures and establish that reliable correlation with glucose concentration exists remains the central challenge in the development of a successful noninvasive glucose method.

The above discussion motivates two fundamentally different approaches to the development of a successful noninvasive glucose sensing methodology: the direct approach and the indirect approach. The direct approach is based on measuring a unique property of the glucose molecule, while the indirect approach measures the changes in one or more physical, physiological, or chemical parameters that track glucose fluctuations.

Optical coherence tomography (OCT) is one of the indirect measurement approaches which uses the effect of glucose on the backscattering of light.[52-55] There is a strong correlation between the glucose concentration and the scattering coefficient of tissues. An increase in the glucose concentration induces a corresponding decrease in the scattering coefficient. The refractive index (n) of the ISF increases with an increase in glucose concentration which results in a decrease in the skin tissue scattering coefficient. This glucose-induced fluctuation in the tissue scattering is used in OCT to measure glucose noninvasively.

The OCT measurement uses an interferometer with a low coherence light source, a reference arm, sample arm and a photodetector to detect the signal in the NIR overtone region.[52] This method has the advantages of a high signal-to-noise ratio in the optical measurement, high spatial resolution[56], and relatively long penetration depths into tissue (scanning depth up to 1 mm). However, the OCT approach has limitations including questionable selectivity, sensitivity to motion during the measurement, and sensitivity to skin temperature.[55] It has also been observed that other skin tissue components such as albumin protein contribute to skin tissue scattering.[54]

Impedance or dielectric spectroscopy (DS) uses the correlation between changes in blood glucose levels and the electrical resistance of tissue to measure glucose noninvasively. The electrolyte balance across the membranes of the blood cells is sensitive to variations in glucose concentration and is estimated in DS. The instrumentation is inexpensive and easy to use. However, the analysis suffers from interference by several chemical and physical components, including temperature, skin moisture, and perspiration.[57]

Because they are based on the measurement of secondary effects correlated to changes in glucose concentration, the indirect methods described above are particularly susceptible to interference from a variety of sources. Their selectivity for glucose has not yet been proven reliable enough for use in a practical measurement device.

By contrast with the indirect methods, the direct approaches target unique properties of the glucose molecule and thereby seek to gain increased selectivity while also possessing sufficient sensitivity to be able to measure glucose across the required concentration range. The vast majority of the research based on the direct measurement approach has centered on the use of spectroscopy to probe transitions associated with the glucose molecule. The methods being examined include near-infrared (NIR) spectroscopy,[58-60] mid-IR spectroscopy,[61] Raman spectroscopy,[62,63] fluorescence,[64,65] and optical polarimetry[66].

Polarimetry is based on the optical activity of glucose. Glucose being a chiral molecule can rotate the plane of polarized light being transmitted through it, and thus a measure of the degree of rotation can be correlated to the glucose concentration. The key to this approach is the choice of measurement site. The use of polarimetry to measure glucose in skin is impractical owing to several reasons. The signal produced by glucose is very small in the clinically important concentration ranges, typically on the order of millidegrees of rotation. Other optically active components present in the skin such as proteins can also interfere with the measurement, and the high degree of light absorption and scattering from skin tissue can be problematic.

To overcome these limitations, polarimetry measurements for glucose have focused on the human eye.[67] There is little protein in the aqueous humor of the eye and

scattering and absorption are also low. The main optical rotatory component in the human eye is glucose. A correlation between blood glucose and that in the eye has been demonstrated. However, the drawbacks to this methodology include poor selectivity resulting from the presence of other optically active compounds such as cholesterol and the physiological lag between the glucose in blood and that in the aqueous humor.[68]

Fluorescence, being fast and extremely sensitive, is also being studied as a measurement platform for use in developing a glucose sensor. While the glucose molecule itself does not fluoresce, sensors can be constructed on the basis of the interaction of glucose with a receptor molecule. For example, glucose and a fluorescent-labeled analogue can bind competitively to a receptor site specific for both. Glucose displaces the labeled binder from the receptor site which in turn alters the fluorescence signal. The most commonly used receptor molecule is Concanavalin A (ConA) and the frequently used competitive binders are dextran, α-methyl mannoside and glycated protein. This concept has also been extended with a fluorescence resonance energy transfer (FRET) approach.[21,64] In another implementation, FRET between a fluorophore and the flavin group in the enzyme GOx has been monitored.[69]

None of these fluorescence-based methods can be considered truly noninvasive as a reagent package must be placed under the skin to allow the interactions that give rise to the fluorescence signal. The measurement hardware is external, however, and no bodily fluids are removed. In addition to issues of selectivity and sensitivity, however, this approach suffers from issues of biocompatibility in much the same way as the implanted electrochemical sensors discussed previously.

The most widely researched noninvasive methods are based on the use of vibrational spectroscopy in the form of NIR, mid-IR or Raman spectroscopies. The NIR and mid-IR methods utilize absorption spectroscopy, while Raman spectroscopy probes vibrational transitions that occur during light scattering. While interrogating many of the same vibrational transitions, Raman and absorption measurements are complementary as their signals are based on different optical phenomena.

The transitions in the NIR range are due to combinations and overtones of the fundamental vibrations associated with O-H and C-H groups of the glucose molecule.[70] By contrast, measurements in the mid-IR region focus on the fundamental vibrational transitions of the molecule. The mid-IR region has stronger and sharper absorption features compared to the NIR. However, water has a strong and broad absorption in the mid-IR which reduces the light penetration depth in this region to tens of micrometers. This limits the measurement to either thin vascular regions in the body or requires the use of techniques such as attenuated total reflectance.[71] In either case, the measurement is biased toward spectral features that are close to the surface of the skin and thus may not correlate well with blood glucose.

Different data collection configurations can be used for collecting NIR noninvasive spectra including transmission, diffuse reflectance, transflectance, and photoacoustic detection. In the transmission geometry, the light is completely propagated through the tissue and the exiting photons are collected and measured relative to the incident photons. The diffuse reflectance methodology measures photons that are backscattered by the tissue matrix at angles other than the incident angle. Such photons have penetrated into the tissue to some extent before being reflected.

The challenge with this approach is to avoid photons whose penetration has been limited to the outer surface of the skin tissue where little ISF is present and therefore little information about the blood glucose concentration is available.[72] A common strategy to overcome this problem is to displace the incident and collection optics from each other such that the only photons detected are those that have traversed some distance into the dermis. The separation distance between the incident and collection optics defines an effective path length for the measurement. Implemented in this way, the measurement geometry is termed transflectance.[73,74]

In the photoacoustic geometry, an incident IR beam modulated at an acoustic frequency generates localized heating in the sample upon absorption. This modulated heat emission creates an acoustic wave and is detected by a microphone.[75,76] The photoacoustic measurement also has concerns regarding penetration depth into the tissue and the relevance of the acquired information for use in correlating to blood glucose concentrations.

Regardless of the optical collection geometry used, NIR spectroscopy has been a very popular region for research into noninvasive glucose monitoring for several reasons. Absorption of NIR radiation by water is lower than in the mid-IR and creates three good transmission windows that have the potential for use in the glucose analysis. Because of the lower background absorbance, the NIR region offers longer light penetration depths (1-10 mm depending on the wavelength region) for the measurements, thereby providing some flexibility in the choice of measurement sites.[9,77] In humans, NIR spectroscopy has been used across different measuring sites including the oral mucosa,[58] tongue,[59] the inner portion of the lower lip,[78] and the index finger.[79] Other advantages of the NIR

region include the availability of sensitive and relatively inexpensive photoconductive detectors for this spectral range and the harmless nature of the radiation when exposed to human tissue.

However, there are some limitations to the NIR region as well. The other components in the tissue matrix show extensive spectral overlap with the available glucose absorption features. Figure 2.2 shows absorption spectra for glucose, urea, glyceryl triacetate (simulant for lipids), and sodium L-lactate (all in 100 mM concentrations) in the NIR range of 4800-4200 cm^{-1}. This is the combination region of the spectrum that includes primarily combinations of fundamental stretching and bending vibrations associated with N-H, C-H, and O-H bonds. While each compound has a distinct signature, the figure clearly demonstrates the extensive spectral overlap that occurs from common constituents of blood and tissue.

A related complication in the NIR region is the occurrence of spectral features that are relatively weak and broad.[70] Measurements in samples with high water content are also highly sensitive to the sample temperature because of the temperature dependence of the hydrogen bonding network of water.[80] There is no single wavelength that can selectively measure glucose in this region. Measurements are thus performed at multiple wavelengths and the data analysis requires multivariate calibration methods to help integrate the glucose signals and extract them from the overlapping background. Success in the noninvasive monitoring of glucose in the NIR region depends on selecting the correct experimental and multivariate model parameters such as optical path length, spectral resolution, calibration model dimensionality, and wavenumber range.

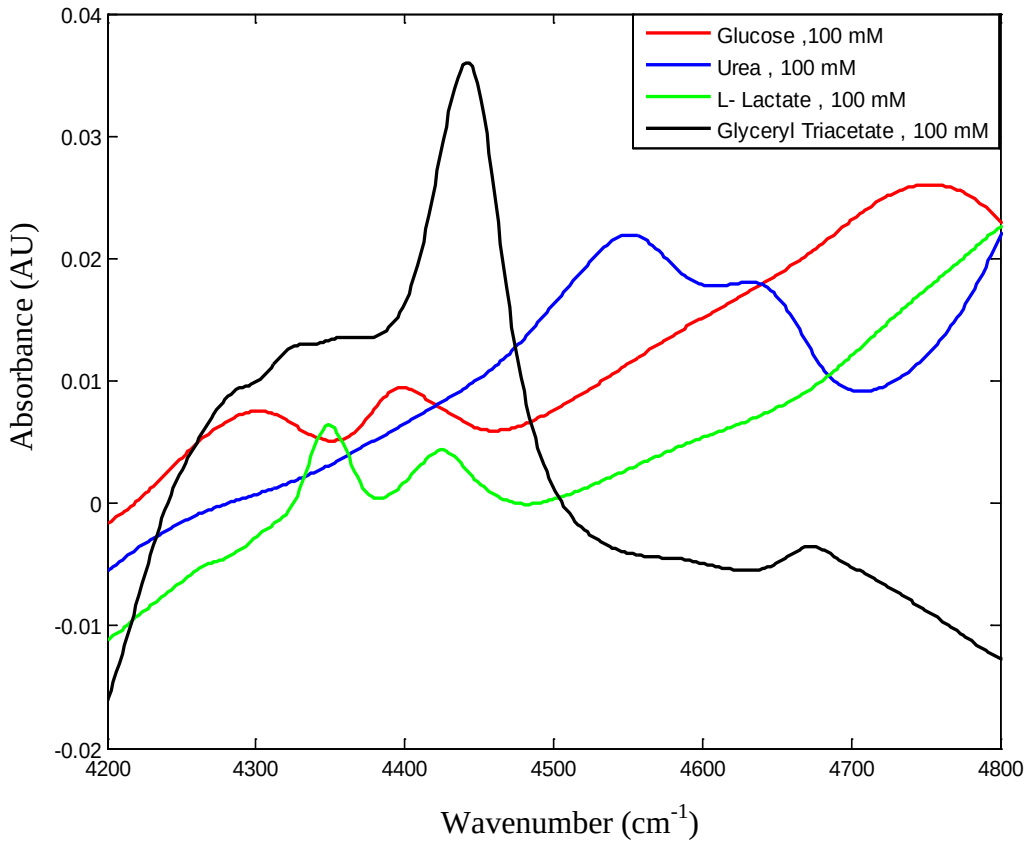


Figure 2.2 Absorbance spectra of glucose, urea, sodium L-lactate, and glyceryl triacetate collected in the 4800-4200 cm^{-1} wavenumber range. All solutions were 100 mM concentrations in 0.1 M, pH 7.4 phosphate buffer. A spectrum of the same buffer was used as the background in the absorbance calculation. Negative absorbance values are a consequence of the water concentration being higher in the buffer background than in the prepared samples.

Developing a well-performing quantitative model typically requires that significant efforts be directed to the optimization of these parameters.

Raman spectroscopy probes the fundamental vibrational transitions in a molecule that are Raman active. As noted previously in the discussion of mid-IR spectroscopy, the fundamental vibrations are sharper and exhibit less overlap than those observed in the NIR.[81] Because water has only a small Raman signal, it is also much less of an interference than in IR absorption measurements.[21] However, two key disadvantages have limited the application of Raman spectroscopy to noninvasive glucose measurements.

First, the acquired Raman signals are extremely weak because only a small fraction of the scattered photons collected in the Raman measurement carry information about vibrational transitions (i.e., the "Raman scattered photons). The great majority of photons are Rayleigh scattered at the same frequency as the incident light and thus carry no useful information about the sample constituents.

To address this limitation, various ways to enhance the number of Raman scattered photons have been explored. Increasing the incident light intensity has the effect of increasing the absolute number of Raman scattered photons, but there are clear limits to the feasibility of this approach in human tissue measurements because of laser exposure limits. Similar concerns arise in resonance Raman measurements in which the incident laser is moved to the visible or ultraviolet region where tissue damage is more likely. Perhaps more feasible is surface-enhanced Raman spectroscopy (SERS) in which an increase in the Raman signal of several orders of magnitude is realized by placing the sample in close proximity to a silver or gold metal surface. For a glucose monitoring application, however, this would entail placing a reagent package under the skin in much

the same way as described previously in the discussion of fluorescence spectroscopy. This minimally invasive approach is being actively investigated.[82,83]

The second major limitation of Raman spectroscopy is the possibility of exciting natural fluorescence in the sample which can swamp the weaker Raman signals. The presence of macromolecules such as proteins in tissue makes this a significant concern in the noninvasive glucose application. For this reason, incident wavelengths have to be shifted to the NIR to lessen the probability of inducing the electronic excitations that give rise to fluorescence. Because of the fourth power relationship between incident light frequency and scattering intensity, however, moving the incident light to longer wavelengths places additional challenges on the signal-to-noise ratio of the measurement.

On the basis of the discussion presented above, it can be argued that NIR spectroscopy provides the most practical approach currently available for pursuing noninvasive glucose measurements. For this reason, the *in vitro* and *in vivo* studies performed in the dissertation research employed NIR spectroscopy. In the sections below, a more detailed description of the theory and the application of NIR spectroscopy is provided.

Introduction to NIR Spectroscopy

Theory and Applications

The NIR region covers a wide range of the electromagnetic spectrum ranging from 0.7 -2.5 μm in wavelength units or 14,286 to 4000 cm^{-1} in wavenumber units (Figure 2.3). Water plays a key role in the IR region producing a strong background absorbance. In the NIR region water has three main absorption bands: (1) the first

overtone of the –OH stretching vibration near 6930 cm^{-1}, (2) an -OH combination band near 5190 cm^{-1},[84] and (3) a tail of the strong –OH fundamental stretching vibration can be seen in the region of 4000 cm^{-1}.[85] These three water absorption bands create three corresponding transmission windows that can potentially be used for the application of NIR measurements to samples with high water content. These three transmission windows are termed the combination (4000-5000 cm^{-1}), first overtone (5500-6500 cm^{-1}) and short wavelength (7300-14286 cm^{-1}) NIR regions.[9] A NIR absorbance spectrum of water with these three sub-regions labeled is shown in Figure 2.4.[86]

The spectral features in the combination region relate to first-order combination transitions associated with bending and stretching vibrations of C-H, N-H, and O-H bonds. The first overtones of these stretching vibrations give rise to the absorption features in the first-overtone region, and the bands in the short wavelength region are numerous related higher order combination and overtone transitions. The NIR spectral bands become weaker and broader when the wavelength of the light decreases from the combination to the first-overtone to the short wavelength regions.[70,87] These weaker and broader bands influence NIR measurement applications in two ways. First, when the bands for a particular analyte get broader, there is a higher chance they overlap with other spectral bands arising from the sample matrix. Second, the sensitivity for a particular analyte becomes an issue when the bands get weaker.[88]

The lower absorptivity of water in the NIR vs. the mid-IR results in greater light penetration depths when aqueous samples are measured by transmission. Stated differently, longer optical path lengths are possible with aqueous samples in the NIR relative to the mid-IR. Viable optical path lengths for NIR measurements are 1 mm for

the combination region, 5 mm in the first overtone region, and 1 cm in the short wavelength region. As noted previously, these longer light penetration depths allow some flexibility in the choice of measurement sites for potential noninvasive glucose measurements in human tissue. Near-infrared spectroscopy was originally considered less useful than the mid-IR because of the occurrence of weaker and broader absorption bands. The availability of chemometric methods in combination with faster computer systems and the development of sophisticated instrumentation and detectors have led to the practical use of NIR measurements in a variety of application areas.

Near infrared spectroscopy is used successfully in the pharmaceutical,[11,89,90] food,[91] chemical,[92] and petroleum[93] industries, as well as in some environmental[94] and biomedical applications.[88] The advantages of using the NIR region for these applications include the nondestructive nature of the radiation, rapidity of measurement, the ability to measure samples without pretreatment, and the relatively high water transmission.

A variety of instrumentation platforms can be used for performing NIR measurements. Initial spectrometer development for the NIR region was done as an extension of existing instrumentation used in the visible spectral range. These were typically scanning dispersive systems that used monochromators in conjunction with single-channel detectors. Filter-based instruments (both tunable and fixed wavelength) have also been widely used, especially for dedicated applications in which only a few wavelengths needed to be monitored (e.g., in the measurement of water in agricultural samples). With the development of multichannel detectors for the NIR, rapid-scanning dispersive instruments have also come into use based on standard spectrograph designs.

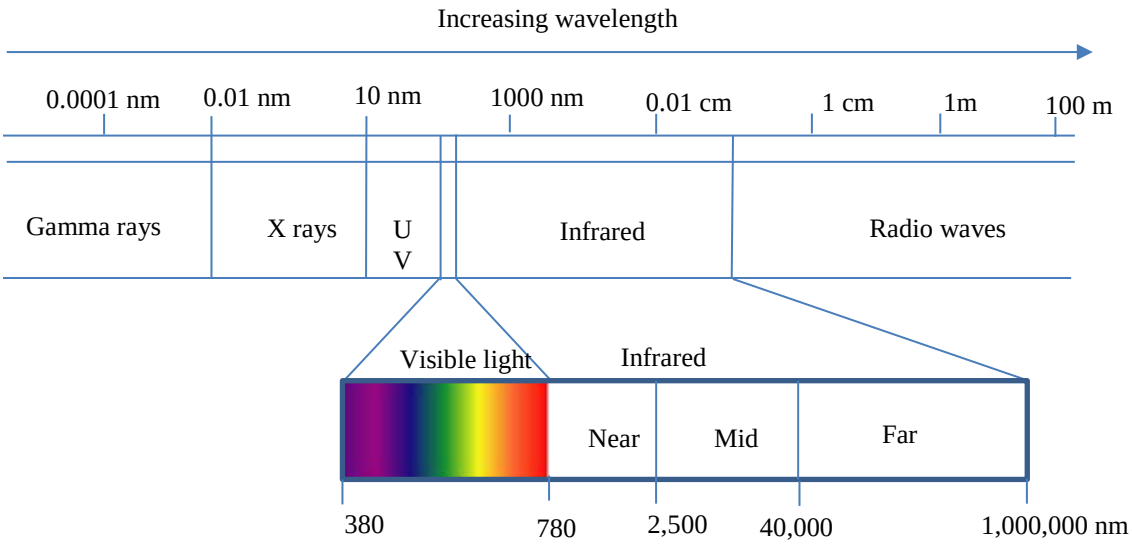


Figure 2.3 Electromagnetic spectrum showing the IR region positioned between the visible and radio wave regions. The IR region is further divided into three sub regions, NIR, Mid-IR and Far-IR.

Source: http://www.windowimage.sg/about/

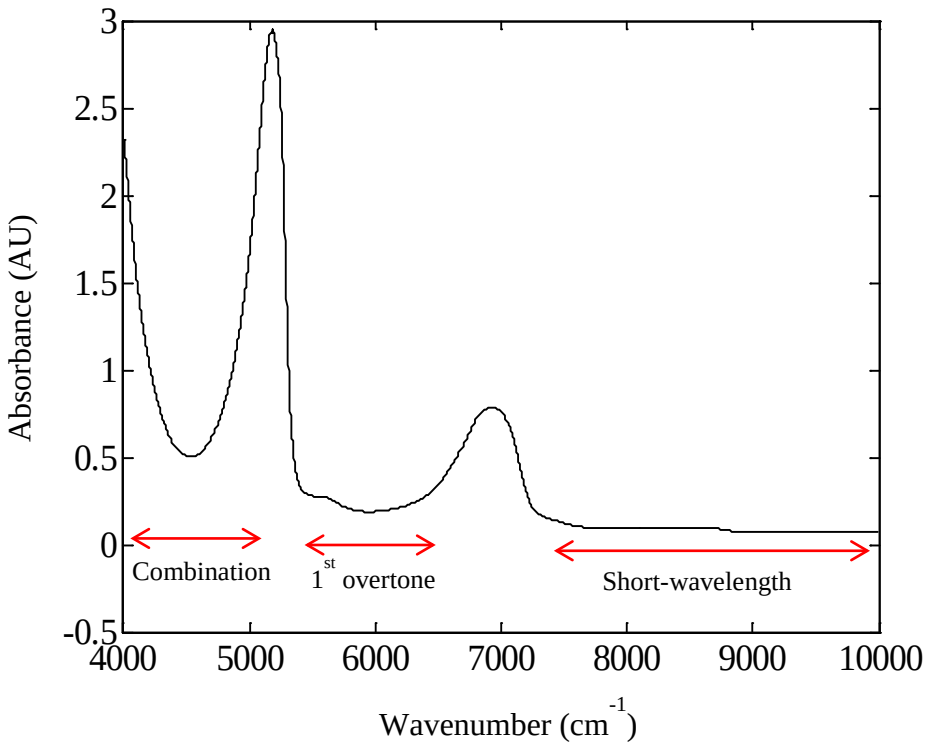


Figure 2.4 Absorbance spectrum of water collected at 37 °C with an optical path length of 0.5 mm. The three transmission windows in the NIR region, combination, first-overtone, and short wavelength, are labeled. An open-beam air reference was used as the background in the absorbance calculation.

Fourier transform (FT) spectrometers were developed initially in the 1970's for the mid-IR to provide for convenient signal averaging of weak spectral signals. They have become the dominant instrument design for the mid-IR and beginning in the 1980's were extended to the NIR region. Since that time, they have become competitive with the best dispersive instruments designed for use in the NIR. The research presented in this dissertation was performed with FT-NIR instrumentation. A detailed description of FT-NIR spectroscopy is given in the next section.

Fourier Transform Near-Infrared Spectrometry

In traditional dispersive NIR spectroscopic measurements, each resolution element (RE) is measured at any instant by a scanning mechanism. If the total scan time is T for N number of REs, the measurement time for each RE would be T/N. This traditional measurement technique suffers from limited SNR, because only a fraction of the total scan time is used to acquire the signal at a particular RE.

The alternative would be to observe each RE for the total scan time. Fourier transform instrumentation has a multiplex measurement capability that is achieved by passing multiple channels of information (i.e., information about the spectral intensity at each RE) onto a single measurement channel through the sample. This is accomplished by an encoding process in which a single optical detector is used to acquire a signal in time that carries the spectral information about each RE in a way that can be decoded later. A particular SNR can be achieved rapidly or an increased SNR can be obtained by signal averaging. The higher SNR can be advantageous when working with the weak signals observed in the NIR region.

In the FT measurement approach, a complex waveform is recorded by the detector as a function of time. This waveform is said to exist in the time domain, and carries the spectral information about the light intensity at each RE. This waveform is mathematically transformed (via the FT) to yield the single-beam spectrum (i.e., the individual light intensities across the REs). The extracted signal produced through the transform is said to exist in the frequency domain. In combination with today's computer technologies and fast algorithms for computing the transform, the FT spectrometer is commonly used in a variety of types of spectroscopic measurements including ultraviolet/visible spectroscopy, infrared spectroscopy,[88] nuclear magnetic resonance (NMR),[95] magnetic resonance spectroscopic imaging (MRSI),[96] and mass spectrometry[97]. The following sections outline the theory of FT spectrometry with a brief discussion of the mathematical operations employed, as well as the instrumentation used when applying the method to measurements in the NIR region. The advantages of FT spectrometry are also discussed.

Theoretical Background of the Fourier Transform

The FT decomposes or separates any harmonic signal or waveform into the set of underlying sine and cosine waves (defined by frequency and amplitude) that comprise it. These underlying waveforms sum to reproduce the original signal.[98] For example, a time-domain signal, $P(t)$, made up of two different cosine frequencies ν_1 and ν_2 can be expressed as,

$$P(t) = k_1\cos(2\pi\nu_1 t) + k_2\cos(2\pi\nu_2 t) \qquad (2.1)$$

In the FT, the time-domain signal, $P(t)$, is mathematically decomposed to find the amplitudes of its individual cosine waves (k_1 and k_2). A plot of the amplitudes as a function of the frequency gives the desired spectrum. Figure 2.5A shows two individual cosine signals with frequencies ν_1 and ν_2 with equal amplitudes. The sum of these two waves can be represented by a sinusoidal wave in the time domain as shown in Figure 2.5B. The constructive and destructive interference that happens in the summation of the individual waves creates a regular pattern in the resulting sinusoidal waveform. The theoretical frequency-domain signal, once the FT is applied, is given in Figure 2.5C. True line spectra (i.e., single lines at the two frequencies) are not realized in the figure because the input to the FT has a finite sampling time (i.e., the transform is not being applied to an infinitely long waveform).

Michelson Interferometer[99-101]

As discussed previously, the concept of FT spectrometry is to use a single-channel detector to acquire a composite (encoded) signal derived from all spectral frequencies. The FT is then used to extract the amplitudes of the component frequencies, thereby yielding the single-beam spectrum.

At optical frequencies, this approach is hampered by the inability of current detectors to respond fast enough to be able to see the input frequencies. To address this problem, an optical component called an interferometer is employed.

In FT-IR spectrometry, the interferometer is used to modulate the input light frequencies such that a single-channel detector can detect the composite waveform composed of each light frequency.

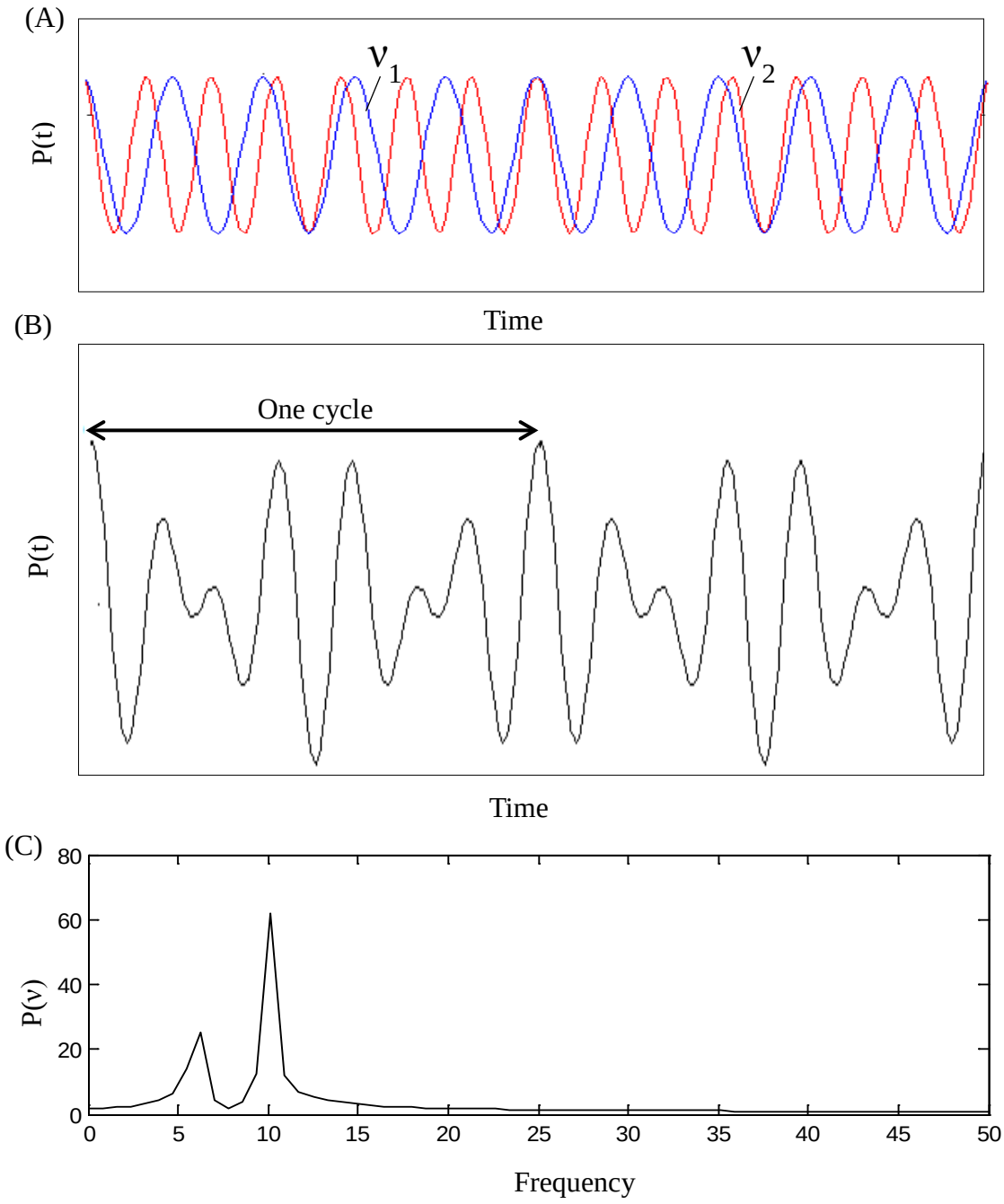


Figure 2.5 Time-domain plots for two single-frequency cosine waves, ν_1, ν_2, plotted as individual signals (A) and as an additive signal (B). Panel C displays the corresponding frequency-domain signal obtained from the application of the FT.

The detector and electronics are thus fast enough to respond to these modulated frequencies. The theory behind interferometry can most readily be understood by studying the common two-beam interferometer designed by Michelson in 1891.

The following section provides an introduction to the Michelson interferometer. A similar instrumental design was used for the research discussed in this dissertation. The simplest form of the Michelson interferometer is shown in Figure 2.6. The beamsplitter, which is ideally 50% transmissive and 50% reflective, divides the input beam of radiation from the source into two paths. One is reflected towards a stationary mirror and the other is transmitted towards a moving mirror. When the two beams return to the beamsplitter, they interfere either constructively or destructively as a consequence of the difference in the lengths of the paths traveled by the photons. They are again partially reflected and partially transmitted. One part of this returning beam is directed through the sample and ultimately detected by the detector, while the other returns to the source and is unused.

The beamsplitter is a very important component in the interferometer because of its essential function to create two beam paths from a single incident light source. The material used in constructing the beamsplitter is chosen depending on the wavelength range required. Commonly used materials in the NIR region can be fused silica (quartz) (Visible/NIR), CaF_2, BaF_2, and ZnSe (NIR/Mid-IR).[101]

The function of the interferometer can be understood by studying the beam paths in Figure 2.6. For simplicity in this discussion, a monochromatic and collimated light beam is considered. Part of the light beam strikes at the point O in the beamsplitter and is reflected towards the stationary mirror at a distance, OS. The other half of the incident beam is transmitted towards the movable mirror at a distance, OM. Once returned to the

beamsplitter, the total difference, δ, in optical path lengths followed by the two waves in Figure 2.6 is 2(OM-OS), where the factor of 2 accounts for the total path traveled (i.e., out to each mirror and back).

This optical path length difference, δ, is termed retardation. Depending on the relative position of the movable mirror, the reflected beams can have different retardations and the waves can interfere either constructively or destructively. When $\delta = n\lambda$, where $n = 0, 1, 2, 3, \ldots$ (where n is an integer) the two light waves are in phase and thus interfere constructively. In this case, the intensity of the light experienced by the detector is equal to that of the intensity of the light source. A zero intensity is experienced by the detector when δ is a half-integral multiple of the wavelength ($\delta = n\lambda/2$, where $n = 1, 3, 5, \ldots$). In between these two cases, the beams undergo partial interference.

For a monochromatic light source, when the movable mirror is moved at a constant velocity, the intensity of the light experienced by the detector is a sinusoidal waveform which has a sequence of maxima and minima depending on the degree of interference (i.e, constructive, destructive, or partial). The plot of light intensity vs. retardation, δ, is called an interferogram.

For a monochromatic light source, the intensity of this sinusoidal wave (cosine wave) experienced by the detector is given as [99,100]:

$$I(\delta)=B(\bar{v})\cdot\cos\left(\frac{2\pi\delta}{\lambda}\right)=B(\bar{v})\cos(2\pi\bar{v}\delta) \qquad (2.2)$$

where $I(\delta)$ is the intensity of light reaching the detector as a function of δ and $B(\bar{v})$ is a constant corresponding to the intensity of the light source, efficiency of the beam splitter, and response of the detector. The term B depends on the wavenumber of the light.

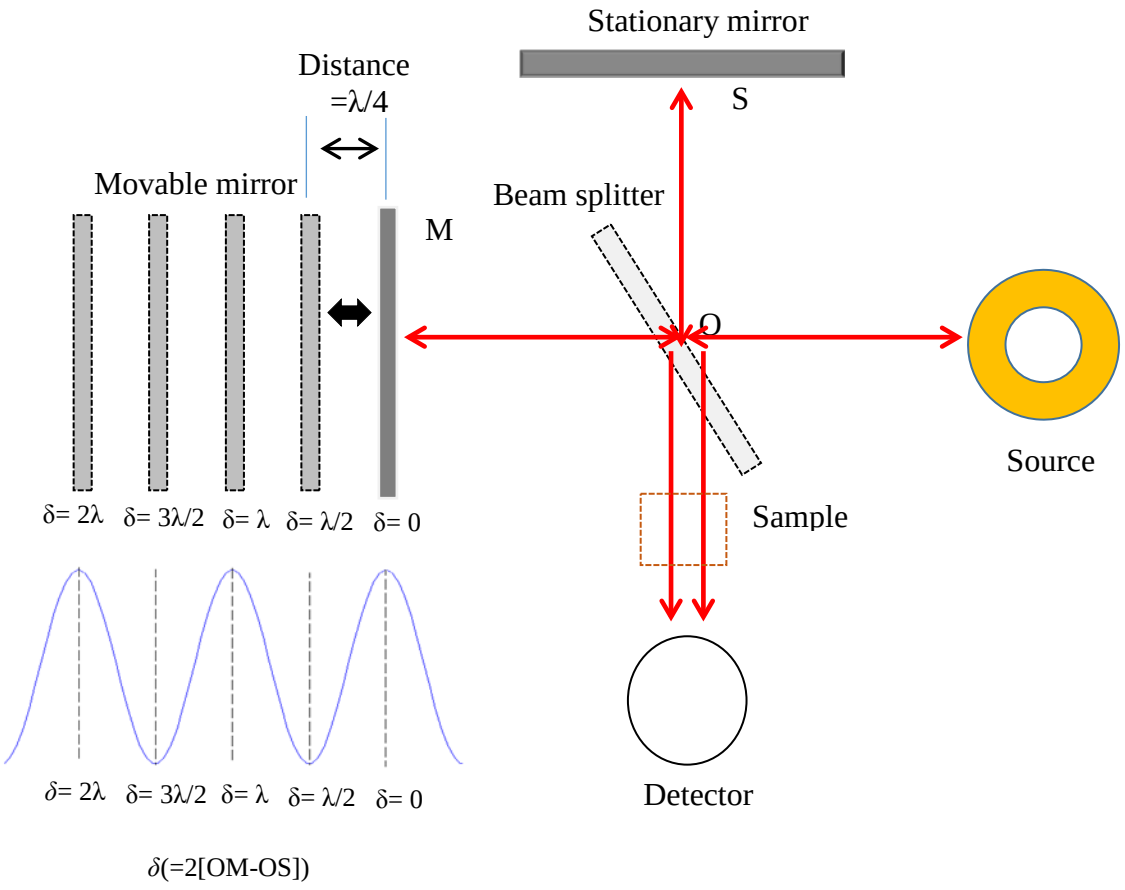


Figure 2.6 Schematic diagram of a Michelson interferometer. The input light from the IR source is directed towards the beamsplitter where half the beam is directed towards a fixed mirror and half is transmitted to a movable mirror. The reflected beams from the two mirrors return to the beamsplitter where they undergo interference either constructively or destructively on the basis of the differences in path lengths traveled. One-half of the resulting beam travels to the source and the other is directed towards the sample and detector where its intensity is recorded. The detector signal recorded as a function of the total path length difference of the two beams is termed an interferogram.

For polychromatic light sources (the typical case), the interferogram can be represented by the integral

$$I(\delta) = \int_{-\infty}^{+\infty} B(\bar{v})\cos(2\pi\bar{v}\delta)\mathrm{d}(\bar{v}) \quad (2.3)$$

The detector experiences the maximum light intensity when the movable and stationary mirrors are at an equal distance from the beamsplitter (i.e., where $\delta = 0$). This is called the point of zero path difference (ZPD) or the interferogram centerburst. As the moving mirror moves away from the centerburst position, the resulting interferogram exhibits an exponentially decaying envelope due to the destructive interference of the many frequency components. The wider the width of the input spectral band, the narrower is the width of the envelope of the interferogram. For broadband spectral sources, the decay is very fast and the envelope of the interferogram is very narrow. By contrast, for a perfect monochromatic source, an infinitely wide interferogram results (i.e., it will be a pure cosine wave as there is no destructive interference resulting from the co-addition of the waveforms of other frequencies).

To convert the interferogram into the frequency-domain spectrum, the FT is performed by evaluation of the following integral expression:

$$B(\bar{v}) = \int_{-\infty}^{+\infty} I(\delta)\cos(2\pi\bar{v}\delta)\mathrm{d}(\delta) \quad (2.4)$$

The FT can be represented in terms of corresponding expressions in the time and frequency domains. For this reason, Eqs. 2.3 and 2.4 are called Fourier transform pairs. Because $I(\delta)$ in Eq. 2.3 is an even function, Eq. 2.4 can be rewritten as:

$$B(\bar{v}) = 2\int_{0}^{+\infty} I(\delta)\cos(2\pi\bar{v}\delta)\mathrm{d}(\delta) \quad (2.5)$$

Resolution and Sampling

According to Eq. 2.5, if the movable mirror moves from 0 to +∞ cm in infinitesimally small steps, a complete spectrum from 0 to +∞ (in cm^{-1}) could be measured with an infinitely high resolution. Moving the mirror to infinity is impractical and thus it can be moved only a finite distance. The interferogram is measured over a limited retardation which causes the single-beam spectrum to have a finite resolution.

The resolution, defined as the ability to distinguish two closely located peaks ($\Delta\bar{\nu}$), is related to the maximum retardation, Δ_{max}, by the equation, $\Delta\bar{\nu} = 1/\Delta_{max}$. The high retardation increases the number of points in the interferogram which ultimately yields a higher number of points in the spectrum. Because the maximum spectral frequency is constant, having more points increases the resolution of the spectrum.

Figure 2.7-A, B, C, D illustrates how the resolution of a spectrum measured interferometrically depends on the number of points or the interferogram length. Figure 2.7A shows an interferogram composed of two cosine waves having frequencies at $\nu_1 = 6$ Hz and $\nu_2 = 10$ Hz which correspond to $\bar{\nu}_1$ and $\bar{\nu}_2$ in wavenumber units. Figure 2.7B corresponds to the FT of the interferogram shown in Figure 2.7A using only 32 points. The resolution is not high enough to distinguish the two peaks at $\bar{\nu}_1$ and $\bar{\nu}_2$. When the length of the interferogram is increased by increasing the number of points to 64, the two peaks just start to resolve as shown in Figure 2.7C. Figure 2.7D shows that if 128 points are used for the transform, sufficient resolution is obtained to resolve the two peaks at $\bar{\nu}_1$ and $\bar{\nu}_2$.

A complete spectrum from 0 to ∞ cm^{-1} can be obtained by sampling the interferogram with an infinitely small spacing in retardation (i.e., with an infinitely high

sampling rate). In practice, however, the interferogram is sampled discretely at a fixed and finite sampling rate. The maximum frequency that can be recognized for a given sampling rate is mathematically explained by the Nyquist sampling theorem. According to the Nyquist theorem, a continuous signal should be digitized with a sampling frequency greater than or equal to twice the maximum frequency of the signal. If the sampling rate is given as 2ν in frequency, any frequency region having a bandwidth of ν can be sampled explicitly with this sampling rate. On the other hand, a feature whose true frequency is higher than ν will induce a lower apparent frequency that will appear below ν. This occurrence is known as folding or aliasing. This phenomenon is explained by Figures 2.8-A, B, C and D.

Figure 2.8A shows a sine wave that has a theoretical frequency of 3 Hz. This signal is sampled with a sampling frequency of 20 Hz which is more than twice the signal frequency. This satisfies the Nyquist criterion and the FT of this signal is the feature shown in Figure 2.8B, a peak correctly centered at 3 Hz. The peak deviates from its pure line spectrum that is desired due to the limitations described previously in the discussion of the effect of finite resolution. A sine wave that has a theoretical frequency of 16 Hz which is sampled with the same sampling rate of 20 Hz is shown in Figure 2.8C. Because the sampling rate for this signal does not meet the Nyquist criterion, the FT produces a spectrum with the peak located at a lower frequency (4 Hz) than the true frequency. These figures highlight the importance of using the correct sampling rate during the analysis to prevent the acquisition of an inaccurate spectrum. If the entire spectrum is to be computed correctly from 0 to $\bar{\nu}_{max}$ in wavenumber with a resolution of $\Delta\bar{\nu}$, the required number of points, *Ns*, to be sampled in the interferogram is given by

$$Ns = \frac{2\bar{\nu}_{max}}{\Delta\bar{\nu}} \tag{2.6}$$

In the *in vitro* simulation studies described in Chapters 4, 5, and 6, the sampling rate is controlled by a reference He-Ne laser with a frequency of 15,798 cm^{-1} (wavelength of 632.8 nm). The laser was directed through the interferometer and the modulated frequency was detected by a visible light detector. The signal output from the visible detector was connected to a sampling circuit which controlled the acquisition of data from the infrared detector. Interferograms were sampled at every zero-crossing of the laser interferogram (i.e., twice per cycle of the sinusoidal signal), which ensures proper sampling according to the requirements of the Nyquist theorem.

Phase Correction

According to Eqs. 2.3 and 2.5, the theoretical representation of the interferogram is symmetric around the point of zero path difference (centerburst). However in reality, a correction needs to be made to the phase angle, $2\pi\bar{\nu}\delta$, to describe the actual measured interferogram. The deviations from symmetry may arise because of optical, electronic, or sampling effects.[100] The addition of new terms into the phase angle has the effect of adding sine components to the cosine interferogram. The process correcting the asymmetric nature of the cosine interferogram, or removing their effects from a spectrum, is known as phase correction.

The phase correction is commonly applied to both the interferogram and single-beam spectrum. The Mertz method which is performed in the spectral domain was used for this research.[102-104] A complex Fourier transform is performed on a symmetric section of the interferogram that spans the centerburst.

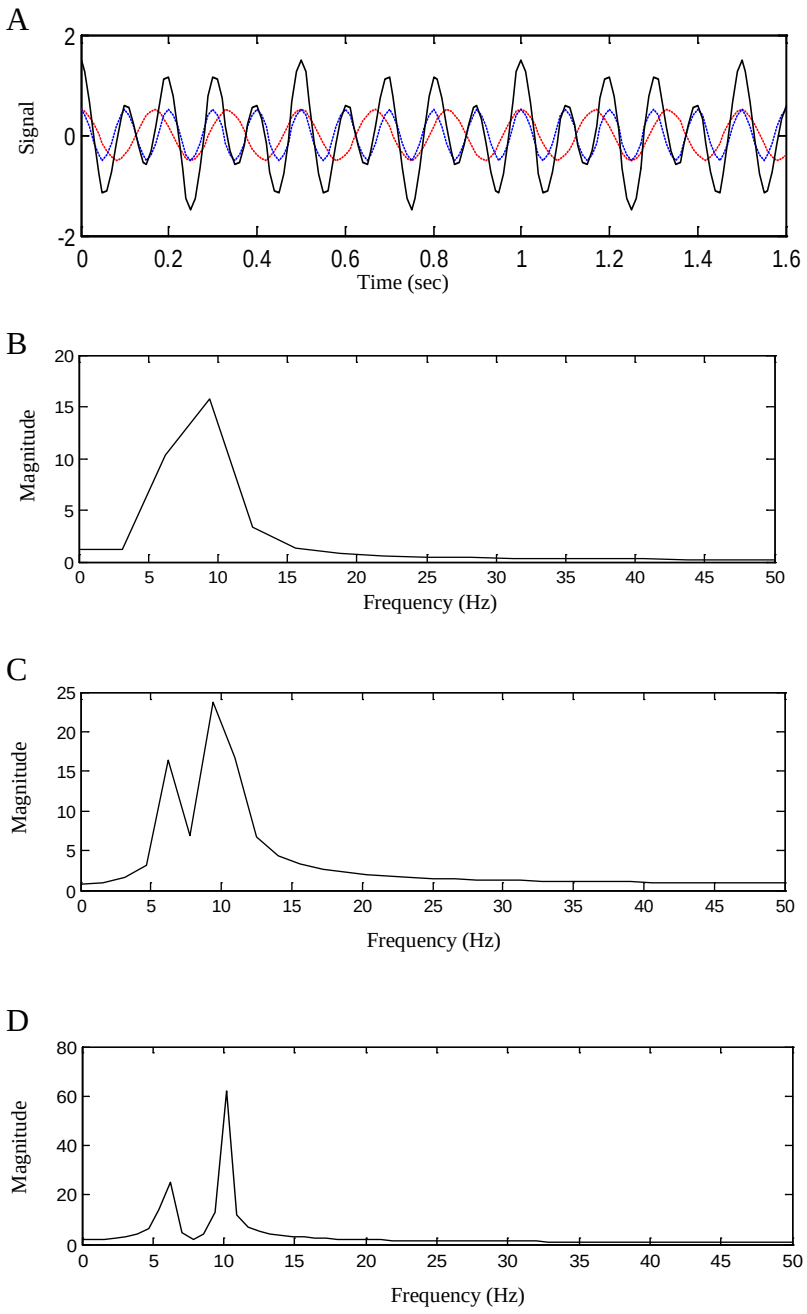


Figure 2.7 Two individual cosine waves having frequencies ν_1= 6 Hz and ν_2=10 Hz are shown in plot A, plotted as individual signals (dotted lines) and one additive signal (solid line). The FT of the additive signal using 32, 64 and 128 points are shown in plots B, C and D, respectively. These plots clearly demonstrate the effect of the length of the sampled interferogram on the ability to resolve two closely spaced peaks.

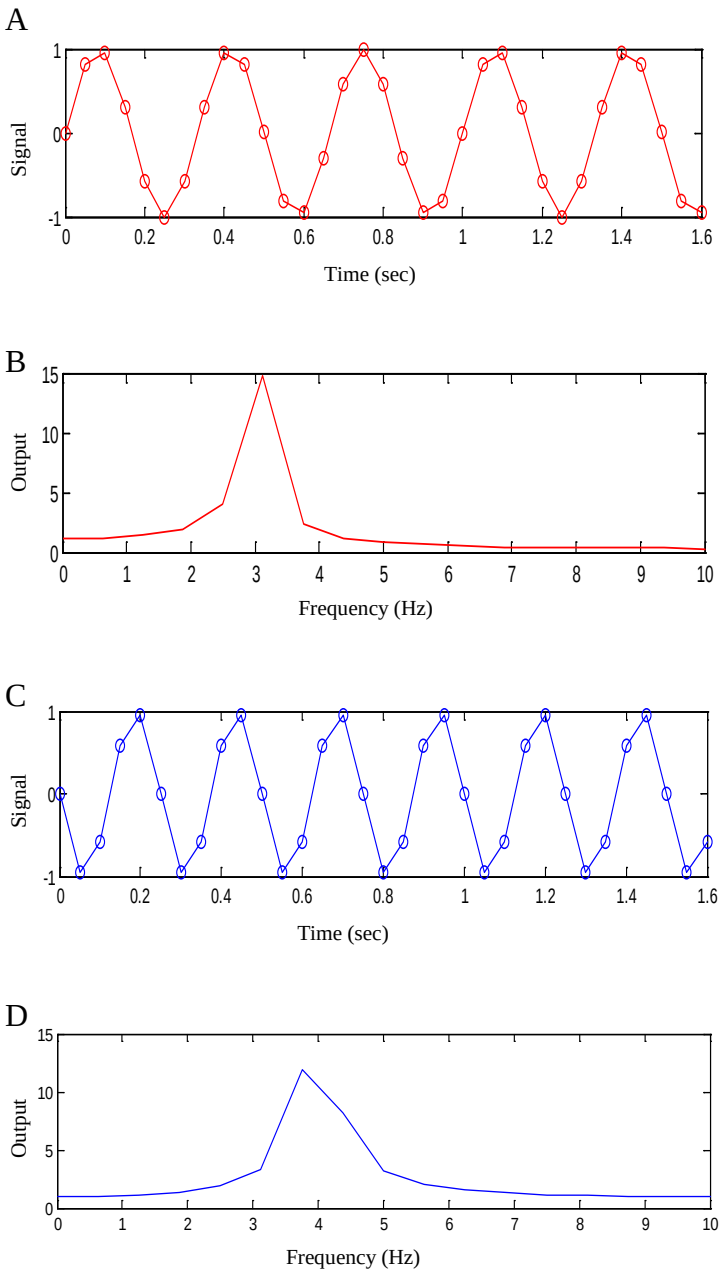


Figure 2.8 Plot A corresponds to a 3 Hz sine wave sampled at a 20 Hz sampling frequency. The FT of this sine wave is shown in plot B. A peak centered at its true frequency can be seen in the plot. This sampling frequency satisfies the Nyguist sampling requirement. Plot C shows a 16 Hz sine wave sampled at the same sampling frequency of 20 Hz. The FT of the signal is shown in plot D, revealing a peak that is aliased to 4 Hz (i.e., located below the true frequency).

The component computed with the cosine transform is called the real part, $\mathrm{Re}(\bar{v})$, while the sine component computes the imaginary part, $\mathrm{Im}(\bar{v})$. The complex FT can be written as:

$$B(\bar{v}) = \mathrm{Re}(\bar{v})\cos\theta(\bar{v}) + \mathrm{Im}(\bar{v})\sin\theta(\bar{v}) \qquad (2.7)$$

In Eq. 2.7, $\theta(\bar{v})$ is the component of the phase shift, which varies with the wavenumber. This phase error function is defined as:

$$\theta(\bar{v}) = \tan^{-1}\left(\frac{\mathrm{Im}(\bar{v})}{\mathrm{Re}(\bar{v})}\right) \qquad (2.8)$$

Once the phase error function has been estimated by use of the symmetric interferogram, Eq. 2.7 can be used to correct the computed spectrum.

Apodization

As discussed previously, the interferogram is collected by restricting the maximum retardation of the interferogram to Δ cm. Experimentally, this corresponds to restricting the translation of the moving mirror of the interferometer to a finite range of retardation. From a mathematical perspective, the equivalent result is obtained by multiplying the theoretically infinitely long interferogram by a truncation function, $D(\delta)$, which is unity between -Δ and +Δ, and zero at all other points:

$$D(\delta) = \begin{cases} 1 \text{ if } -\Delta \leq \delta \leq +\Delta \\ 0 \text{ if } \delta > \Delta \end{cases} \qquad (2.9)$$

This function is termed the boxcar truncation function. With this truncation function Eq. 2.4 can be rewritten to represent the interferogram that is actually sampled:

$$B(\bar{v}) = \int_{-\infty}^{+\infty} I(\delta)D(\delta)\cos(2\pi\bar{v}\delta)\mathrm{d}(\delta) \qquad (2.10)$$

The multiplication of two functions is the convolution of the FT of each function.[105] The product of $I(\delta)$ measured with an infinitely long retardation and the boxcar function $D(\delta)$ results in a spectrum that is the convolution of the FT of each. The FT of $I(\delta)$ yields the true spectrum (e.g., an infinitely narrow line for a monochromatic light source), whereas the FT for $D(\delta)$ is the sinc function (i.e., $\sin x/x$), expressed as:

$$f(\bar{\nu}) = 2\Delta \operatorname{sinc} 2\pi(\bar{\nu})\Delta \qquad (2.11)$$

In addition to a broadening of the central peak, a ramification of finite sampling is that the presence of the sinc function adds artifacts such as oscillating side-lobes to the final FT spectrum.

The spectral resolution associated with the measurement is determined by the width of the boxcar function. This corresponds to the maximum retardation of the interferogram as discussed previously. For the boxcar function, the full width at half-maximum height of the observed band is $0.605/\Delta$. The FT of the boxcar function also causes a "leakage" of the spectral intensity, i.e. the spectral intensity is not strictly localized but also contributes to the side lobes.[106]

To reduce the side lobes caused by the boxcar function, other sampling functions, $A(\delta)$, such as the triangular or Happ-Genzel function are often used.[107] These sampling functions are also termed windowing or apodization functions.

Unlike the boxcar function which has an infinitely sharp transition to zero in the edges, these alternate sampling functions gradually approach zero on the edges, resulting in fewer side lobes when transformed to the final spectrum. This procedure of reducing the side lobes by taking the product of the interferogram with an artificial sampling function is termed apodization.[108]

The Happ-Genzel apodization function can be represented as:

$$A(\delta)=\begin{cases} a+b\cos\frac{\delta\pi}{2\Delta}, & \text{if-}\Delta\leq\delta\leq+\Delta \\ 0, & \text{others} \end{cases} \quad (2.12)$$

In Eq. 2.12, the parameters *a* and *b* are optimized to minimize the first (largest) side lobe. The choice of a particular apodization function depends on the particular application. As mentioned above, compared to the boxcar function, other apodization functions minimize the side lobes but there is also a tradeoff in spectral resolution. This effect derives from the width of the center peak of the FT of the apodization function. If a higher resolution is very important, the boxcar function should be chosen. If the resolution is less important than concern about the presence of the side lobes, other apodization functions can be utilized. In practice, the boxcar function is typically only used for gas samples where the spectral bands are narrow. For the *in vitro* simulated experiments in Chapters 4, 5 and 6, the Happ-Genzel apodization function was used. In Chapter 7, triangular apodization was used.

Advantages and Disadvantages of FT-IR Spectrometry

There are several advantages that an FT-IR spectrometer has compared to the traditional single-channel dispersive instrument.[99,100,103,109] The FT-IR spectrometer offers the multiplex advantage (Fellgett advantage). In contrast to dispersive measurements in which the spectrum is examined by scanning one RE at a time, in an FT-IR measurement all the frequencies in a spectrum are acquired simultaneously. The Fellgett advantage offers the ability to collect the complete spectrum very rapidly and many scans can be averaged in the time taken for a single scan of a dispersive spectrometer that is based on a single-element detector. While modern multichannel dispersive instruments based on

array detectors have overcome the limitation in scan speed discussed above in the ultraviolet and visible spectral regions, the FT design is still preferable in the IR region due to the expense and lack of reliability of IR multichannel detectors.

Another advantage the FT-IR measurement offers is the throughput advantage or Jacquinot advantage. A FT-IR instrument does not use a slit to limit the amount of light reaching the sample and detector as a dispersive instrument does. In addition, the fewer mirror surfaces in the FT-IR spectrometer minimize reflection losses. Overall, a higher amount of light energy reaches the sample and hence the detector in an FT-IR spectrometer compared to a dispersive spectrometer. This in combination with the multiplex advantage means that the signal-to-noise ratio achieved with an FT-IR spectrometer is often higher than that obtained with a dispersive instrument. This high throughput advantage is somewhat limited since too much light reaching the detector may saturate the detector and thereby exceed its linear dynamic range. Consequently, for less absorbing samples, the light output from the interferometer may need to be attenuated. However, the high throughput of light reaching the detector is extremely valuable for highly absorbing or scattering samples such as those encountered in aqueous or *in vivo* measurements.

Another advantage of FT spectrometry is the Connes advantage, which allows greater wavenumber accuracy to be achieved than in a conventional dispersive design based on a monochromator. In modern interferometers, this is accomplished by using a very stable helium-neon laser to control the sampling of the interferogram.

There are several limitations of FT-IR spectrometry as well. The interferometer has moving parts and requires precise optical alignment in order to function. The

ruggedness of the spectrometer may thus be insufficient for some application environments. The need for the interferometer also makes it difficult to miniaturize the spectrometer. Finally, in systems that are source-noise limited, the Fellgett disadvantage applies. This arises because all the regions of the spectrum are collected simultaneously (i.e., all light frequencies are integrated at each interferogram point). Therefore, if noise occurs in one spectral region, it will contaminate each interferogram point with the end result being the noise will be spread throughout the spectrum. Overall, however, the advantages of the FT instrumental design outweigh the disadvantages. For this reason, FT spectrometers have become the instrument of choice in the mid-IR region and their use has also become very widespread for NIR measurements.

CHAPTER 3

DATA ANALYSIS TECHNIQUES

Introduction

In the previous chapter, the theory and instrumentation associated with FT-NIR spectroscopy were discussed. Near-infrared spectra are difficult to interpret due to the weak and broad nature of the peaks and the extensive spectral overlap that occurs when multi-component samples are measured. The extraction of useful spectral information thus typically requires the aid of computer-based data analysis tools from the field of chemometrics.

The ultimate goal of the research described in this dissertation was to develop an alarm algorithm for the detection of nocturnal hypoglycemia from NIR spectra collected continuously from a diabetic patient during the sleep period. In this context, the data analysis problem can be considered qualitative in nature. For each spectrum collected, the task is to classify the spectrum into one of two data categories: (1) corresponding to a glucose concentration below a threshold concentration, thus triggering an alarm or (2) corresponding to a glucose concentration above the threshold, thus requiring no alarm. For simplicity, these data categories will be termed “alarm” and “non-alarm”.

While the decision-making task described above is qualitative in nature, different steps of the algorithm require quantitative steps which link the spectrum to quantifiable properties of the sample. For these purposes, multivariate calibration tools are required. This chapter provides an overview of the qualitative and quantitative methods employed in the research, as well as the diagnostic tools used for assessing data quality and method

performance. A detailed description of the steps behind the alarm algorithm will also be provided.

Noise Analysis

In the nocturnal hypoglycemic alarm applications discussed here, spectra were collected continuously and the mean root-mean-square (RMS) noise level was computed routinely as a diagnostic check on data quality and overall instrumental performance. The calculation was performed in three steps. First, ratios were taken between consecutive single-beam spectra, then the absorbance values were computed, and the RMS noise was computed over the 4500-4300 cm^{-1} spectral range. This range was selected because it encompasses the glucose C-H combination band at 4400 cm^{-1}. The ratio of two consecutive replicate single-beam spectra is termed a 100% line because in the absence of measurement variance all values should be at 100% transmittance. In the corresponding absorbance spectrum, all values should be at 0.0 absorbance units (AU).

The shape of the 100% line can be used to identify possible sources of variance that occur during the data collection. Offset of the line from zero indicates possible instrumental drift while the temperature fluctuations in the sample solutions can result in baseline curvature. Random fluctuations in the line about 0.0 AU are caused by the intrinsic noise of the instrumental measurement.[86]

The magnitude of each noise source can be estimated by changing the origin of the RMS noise calculation shown in Eq. 3.1

$$\mathrm{RMS}=\sqrt{\frac{\sum_{i=1}^{p}\left(A_i-b_o-b_1\bar{\nu}_i-b_2\bar{\nu}_i^2-b_3\bar{\nu}_i^3\right)^2}{p-q}} \qquad (3.1)$$

In Equation 3.1, RMS denotes the estimated noise level calculated over the 100 % line containing *p* resolution elements corresponding to wavenumber values, $\bar{\nu}_i$,and absorbance values, A_i. The coefficients b_o to b_3 in the equation determine the magnitude of the noise source and *q* gives the number of nonzero values among b_o to b_3. If b_o to b_3 = 0, Eq. 3.1 uses only the absorbance values, A_i, for the calculation and its magnitude includes all three sources of variance described above. If b_o is set to the mean of the *p* absorbance values and all other coefficients are zero, the absolute shift from zero absorbance will be corrected and the magnitude of RMS will include the effects of baseline variation and random fluctuations. Assuming a third-order polynomial regression can fit the baseline curvature, A_i and $\bar{\nu}_i$ over *p* resolution elements can be fitted to this model with b_o to b_3 in Eq. 3.1 denoting the regression coefficients. The third-order polynomial fit will then correct for the baseline variation and the RMS noise value will be due to the intrinsic measurement noise only.

Multivariate Calibration Methods

To determine the amount of an analyte present in a sample, a calibration model is required to generate the mathematical relationship that links the recorded signal (e.g., NIR spectra) with the parameter of interest (e.g., concentration). Two types of calibration models can be identified on the basis of the dimensionality of the data used for the analysis. The univariate model is the simplest and most traditional form, which involves relating one variable to predict the value of another. For example, the spectroscopic signal collected at a single wavelength is related to the analyte concentration through a simple one- or two-parameter model. By contrast, a multivariate model uses the

information derived from two or more measurements. For example, spectra collected at multiple wavenumbers or wavelengths are related to the analyte concentration through a multi-parameter model.

Given the extensive spectral band overlap and the weak absorption features observed in NIR spectra, multivariate techniques are generally required to extract target analyte information from the background signal produced by the other components of the sample matrix. In this research, all the data collected consisted of multiple measurements across a wavenumber range. The multivariate regression methods used in the dissertation research are described below.

Classical Least-Squares

Classical least-squares (CLS) analysis[110-113] is a multivariate linear calibration model computed by applying multiple linear regression (MLR). A linear mixture model that is used in the MLR can be represented as

$$y = b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_hx_h + e \qquad (3.2)$$

where the measured response (y) is based on a linear sum of contributions from h species and x_i is the "unit" response of species i. The regression coefficients (b_i) are weighting factors that encode the contribution of each species i to the measured response.

In NIR absorption spectroscopy, a model is built that fits the measured absorbance to a linear sum of contributions from a series of known spectral components. Here, the absorbance values measured across a spectral range define the dependent variables while the series of underlying spectral components collected over the same spectral bandwidth are treated as the independent variables. The regression coefficients

that form the models give the relative amount of each spectral component represented in the measured absorbance. If the pure-component spectrum of a particular analyte is taken as one of the independent variables, the regression coefficient for that term calculated from the model gives the relative amount of that component represented in the measured spectrum.

For multi-component absorbance analysis at a given wavelength, Eq. 3.2 can be written as:

$$A_i = \sum_{j=1}^{l} k_{i,j}\ b_i + e_i \tag{3.3}$$

where A_i is the total absorbance at wavelength i and $k_{i,j}$ is the absorbance in the pure-component spectrum for component j collected at wavelength i. This term is called the sensitivity term which relates the measured absorbance to the analytical parameter of interest (e.g., concentration, thickness, temperature). The term b_i in Eq. 3.3 is the calculated regression coefficient. This is the quantitative parameter that is desired from the analysis. The lack of perfect fit between the left and right sides of the equation is expressed in the residual absorbance, e_i. In Chapter 5, CLS is used to calculate the relative thickness of the protein films prepared. The relative thickness values were evaluated by calculating the regression coefficients from the fit.

If the absorbance data are collected over p resolution elements for a group of n samples, $p \times n$ total absorbance values will be accumulated and the term A_i in Eq. 3.3 becomes a matrix, **A**. The columns of **A** comprise the spectra of the n samples. If the p elements are contiguous and evenly spaced across p resolution elements and there are a total of l components contributing to the spectrum, the term $k_{i,j}$ in Eq. 3.3 now becomes a $p \times l$ matrix, **K**, which is defined sometimes as the sensitivity matrix. For l number of

components and *n* number of samples, the term in the equation will be changed into an *l*×*n* matrix, **B**, which is the regression coefficient matrix.[86]

For this type of multi-wavelength and multi-component analysis, Eq. 3.3 can be written in matrix notation as:

$$\mathbf{A} = \mathbf{KB} + \mathbf{E}_A \tag{3.4}$$

Analogous to e_i in Eq. 3.3, the $p \times n$ matrix $\mathbf{E}_A$ is placed in the equation to account for the information in the spectral data matrix **A** that is not explained by **KB**. The matrix $\mathbf{E}_A$ is termed the spectral residuals from the fit. This information includes noise or other spectral artifacts that do not fit the assumptions of the linear mixture model specified in Eq. 3.2. The desired analytical information resides in **B** and can be obtained by use of standard MLR calculations. The least-squares solution to **B** in Eq. 3.4 in matrix form can be given as:

$$\hat{\mathbf{B}} = \mathbf{K}^T (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{A} \tag{3.5}$$

where $\mathbf{K}^T$ indicates the transpose of the matrix **K** and the hat symbol in the regression coefficient matrix, **B**, denotes that it is estimated on the basis of the experimental data in **K** and **A**.

In the CLS methodology described above, no terms were used in the model for correcting for background changes, baseline variation, etc. If these artifacts are present, they will cause errors in the predicted parameters. To correct this, Haaland and coworkers have developed a series of strategies to add flexibility to the CLS model termed augmented CLS (ACLS).[114-117] The solution for this is to augment **K** with additional columns that model the “shapes” of the artifacts. For the thickness calculations in Chapter 7, the ACLS method is used.

Multivariate Inverse Calibration Methods

The success of the CLS method depends on accurately defining the sensitivity matrix, **K**, in Eq. 3.4. The CLS procedure will likely perform poorly if all the species that contribute to the response cannot be characterized. This is often the case with complex samples of the type encountered in biological or environmental applications. In this case, the **K** matrix in Eq. 3.4 is not adequately defined and the computed regression coefficients in **B** will be inaccurately estimated.

To model samples in which the sample matrix is complex, variable and difficult to characterize completely, an inverse regression method can be employed in which the concentrations become the dependent variable ("y") and the measured spectral intensities become the independent variables ("x_i"). Spectral intensities are taken at multiple resolution elements for concentrations of a set of calibration samples and then the concentration is modeled as a function of the measured instrumental responses by MLR. This is termed an inverse calibration method.[113,118-120]

In an inverse regression model for an analyte of interest in a sample, the concentration of that analyte can be modeled as a function of the spectral intensities at p resolution elements as

$$c_i = b_0 + b_1x_{1,i} + b_2x_{2,i} + \ldots.. + b_px_{p,i} + e_i \qquad (3.6)$$

where c_i is the concentration of the analyte of interest in sample i, $x_{1,i}$, $x_{2,i}$,…, $x_{p,i}$ denote the spectral intensities from p resolution elements, and b_0,b_1,…, b_p are the regression coefficients estimated from MLR. The p resolution elements used in the model are selected from the pool of available independent variables based on a wavelength selection

procedure. Here, e_i denotes the error or residual associated with the fit. For a set of n samples containing h components, Eq. 3.6 changes to matrix form as

$$\mathbf{C} = \mathbf{XB} \tag{3.7}$$

where **C** is an $n \times h$ matrix for n samples containing h components, **X** is an $n \times p$ matrix which contains spectral intensities collected for n samples at p wavelengths, and **B** is an $n \times h$ matrix which contains regression coefficients resulting from the MLR calculation.

To generate the model, a set of calibration samples is obtained and spectra are measured at multiple wavelengths for each sample. These data allow **C** and **X** in Eq. 3.7 to be defined. The regression coefficient matrix, **B**, can then be estimated by a generalized inverse as:

$$\mathbf{B} = \mathbf{X}\,(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{C} \tag{3.8}$$

Estimated concentrations for unknown samples can be obtained by using the regression coefficient matrix calculated with Eq. 3.8 and the spectral intensities matrix for the unknown samples. The estimated concentration matrix can be obtained by

$$\widehat{\mathbf{C}} = \mathbf{XB} \tag{3.9}$$

The calculation of the regression coefficient matrix by Eq. 3.8 involves the inversion of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. If collinear relationships exist among the columns and rows in the **X** matrix, poor conditioning to inversion of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ occurs. These collinear relationships are very common in the intensities associated with the same spectral band. Depending on its severity, this ill-conditioning in the spectral data matrix can lead to imprecision in computing the inverse of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. This leads to imprecision in the computed regression coefficients.

The number of resolution elements, *p*, that can be used in the model is limited by the number of independent samples, *n*, or the degree of freedom available within the calibration data set. If the number of wavelengths is too large compared to the number of samples available, chance correlations between **X** and **C** occur and can lead to an erroneous estimation for **B**. On the other hand, if the numbers of resolution elements are too low, the instrumental response may not be selective enough for the analyte.

According to the ASTM Standard E 1655 on infrared multivariate calibration methods, a minimum *n/p* ratio of 6.0 is highly recommended.[121] For example, if the spectral data acquisition is from 4900 to 4100 cm^{-1} with a point spacing of 8 cm^{-1}, there are 101 resolution elements in each of the spectra collected. To satisfy the ASTM standard, $6 \times 101 = 606$ independent samples are thus recommended to build the calibration model. A drawback of this calibration method is that such a high number of calibration samples is not feasible in most applications.

The use of MLR to build calibration models for the data collected in NIR spectroscopy faces some critical issues. As noted previously, the spectral bands are broad and there is a high degree of overlap of the spectra of the sample components. The individual resolution elements not only show poor selectivity for the analyte of interest, they will be correlated with adjacent data points. This leads to a spectral data matrix that is ill-conditioned to inversion.

One potential approach to overcome this problem would be to identify key wavelength points and build the model only with those subsets of spectral points. However, the extensive spectral overlap in NIR spectra can make it difficult to select subsets of spectral points that will build a precise calibration model. Due to these

limitations, latent variable techniques such as principal component analysis (PCA) and partial least-squares (PLS) are often used. The next section will be used to discuss those latent variable methods used in the dissertation research.

Latent Variable Methods

The basic idea of the latent variable methods is to calculate a new set of independent variables based on linear combinations of the original variables. The new variables are called latent variables because they reflect an underlying structure in the data matrix. In the context of spectroscopy, these latent variables are empirically derived spectral shapes that explain the variance in the original spectra. The relative contribution of each of these shapes to the original spectrum can be given by a set of scores. The goal of latent variable methods is to extract the key information contained in the original data but to do so with a lower dimensionality. Whereas *p* points were originally collected to represent the spectral information, it is hoped that many fewer than *p* latent variables can be found that carry the same information. If this reduction in dimensionality is realized, fewer terms will be required in the calibration model, thereby reducing the number of calibration samples that are required to satisfy the ASTM standard.

The most efficient form for the latent variables is that they be orthogonal, as orthogonal variables have no redundant information. If this is achieved, the new variables will have no collinearity and therefore there will be no issues regarding the ability to invert $\mathbf{X}^T\mathbf{X}$.[122-124] This leads to a more precise determination of the regression coefficients obtained by use of Eq. 3.8.

The relationship between the measured responses and the calculated latent variables can be given as

$$\mathbf{T} = \mathbf{RS} \tag{3.10}$$

where **R** is an $n \times p$ matrix for the measured responses and **S** is a $p \times h$ matrix in which the columns contain the h empirically derived spectral shapes that are also termed loadings or latent variables. The product of **R** and **S** yields **T**, an $n \times h$ matrix of scores. If **S** is an orthogonal matrix, each column can be considered an orthogonal basis vector in an h-dimensional coordinate system, and row i of **T** represents the projections (coordinates) of spectrum i onto this new basis.

The calibration model given in Eq. 3.6 can now be built with the new h-dimensional score vector as:

$$c_i = b_0 + b_1 t_1 + b_2 t_2 + \ldots\ldots + b_h t_h + e_i \tag{3.11}$$

The success of this approach depends on constructing the latent variables to be more efficient than the original variables in representing information in **R**. A variety of methods are available to calculate the loadings and scores in Eq. 3.10. The next section discusses the two approaches used in the dissertation research, PCA and PLS.

Principal Component Analysis[110,124-126]

The strategy of PCA is to factor the response matrix, **R**, into the product of a score matrix, **T**, and an orthogonal loading matrix, **V**, as shown below:

$$\mathbf{R} = \mathbf{TV}^{\mathrm{T}} \tag{3.12}$$

In Eq. 3.12, **R** is the ($n \times p$) response matrix and **T** is the score matrix ($n \times p$). The response matrix may or may not be mean-centered (i.e., the mean response subtracted

from each row) depending on the user's choice. The columns of **V** ($p \times p$) are the loading vectors. Once **V** has been computed, the individual loading vectors can be examined to evaluate the relevance of the information they carry. In PCA, the goal is to extract the relevant information from **R** while leaving the less relevant information behind. For a calibration modeling application, this translates to the use of only the relevant scores to build the calibration model. This approach can be represented in matrix form as:

$$\mathbf{R} = \mathbf{T}\mathbf{V}^{\mathrm{T}} + \mathbf{E} \tag{3.13}$$

In Eq. 3.13, the score matrix, **T**, is now $n \times h$ and **V** is now $p \times h$, where $h < p$. The residual matrix, **E**, is an $n \times p$ matrix that contains the information in **R** not extracted by the h loading vectors in **V**.[125]

The loading matrix **V** can be calculated from **R**, and **R** and **V** together can be used to calculate the scores (**T** = **RV**). The singular value decomposition (SVD) is the computational algorithm most often used to compute the loadings.[127,128] The application of the SVD to **R** results in an orthogonal matrix of loadings in which the columns are the eigenvectors of $\mathbf{R}^{\mathrm{T}}\mathbf{R}$. The eigenvalue associated with each loading is proportional to the magnitude of the loading vector before normalization. The useful eigenvectors that extract highly relevant information in **R** are typically those with the largest eigenvalues. These eigenvectors explain the most variance in **R**. The eigenvectors of $\mathbf{R}^{\mathrm{T}}\mathbf{R}$ with the largest eigenvalues are the loadings used in building the calibration model. These loading vectors are called principal components (PCs).

The computed loading vectors and the response matrix **R** can be used to calculate the score matrix. For each loading vector $\mathbf{v}_i$, the corresponding score vector is

$$\mathbf{t}_i = \mathbf{R}\mathbf{v}_i \tag{3.14}$$

The score vectors computed in this way are orthogonal since **V** is an orthogonal matrix. Thus, if collinearity exits in **R**, the orthogonality of the $\mathbf{t}_i$ offers the ability to explain the key information in **R** using fewer variables.

For a calibration modeling application, the goal is to use the scores as independent variables in an MLR model such as that described by Eq. 3.11. The $h \times 1$ vector of regression coefficients, **b**, can be obtained using both the score matrix and **c**, the $n \times 1$ vector of concentrations of the prepared calibration samples for a particular analyte. The calculation is the standard least-squares solution:

$$\mathbf{b} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{c} \tag{3.15}$$

When applied in this way, the combination of PCA and MLR is termed principal component regression (PCR).

A calibration model built with PCR can be used for future predictions. The spectra for the prediction samples are collected. If mean-centering was used originally, the mean of the calibration responses is used to center the prediction spectra. The scores for these spectra are then calculated using the **V** matrix obtained from the calibration data. The computed scores are used together with the regression coefficients obtained from Eq. 3.15 to estimate the concentrations for the prediction samples. The calculation is analogous to Eq. 3.9 except the matrix **B** is replaced by the vector **b** (i.e., regression coefficients for a single analyte are used).

Two other points regarding the use of PCA are noteworthy. First, two- or three-dimensional scatter plots can be created from the computed scores. These score plots are a very useful way to picture relationships among the observations in multi-dimensional

data. These plots can be used to observe outliers and to uncover clustering or grouping among the data. Score plots of this type are used in the dissertation.

Unusual observations can be identified in the context of the PCA model by either examining the residual spectra remaining after the subtraction of the contributions of the principal components or assessing the degree to which the computed principal component scores are unusual relative to the scores of the other observations.

The matrix of spectral residuals, **E**, in Eq. 3.13 contains the information in the input spectra data matrix, **R**, that is not captured by the *h* computed principal components. Each row of **E**, $\mathbf{e}_i$, contains the residual spectrum for observation *i*. We can define Q_i as the sum of squares of the elements in $\mathbf{e}_i$. This value is the squared magnitude of residual spectrum *i* and is a measure of the degree to which observation *i* fits the principal component model. Large values of Q_i signal unusual observations.

The sum of normalized squared scores, known as Hotelling's T^2 statistic[129], is a measure of the variation in each sample within the PCA model. The T^2 is defined as

$$T_i^2 = \mathbf{t}_i \mathbf{\Lambda}^{-1} \mathbf{t}_i^T \tag{3.16}$$

where $\mathbf{t}_i$ refers to the i^{th} row ($1 \times h$) of the principal component score matrix, **T**, and $\mathbf{\Lambda}$ is an $h \times h$ diagonal matrix containing the eigenvalues of $\mathbf{R}^T\mathbf{R}$. The eigenvalue in element (*j*, *j*) of $\mathbf{\Lambda}$, λ_j, is the squared magnitude of principal component *j* (i.e., the squared magnitude of the loading vector). The function of the eigenvalue in Eq. 3.16 is to weight the contribution of each principal component according to its magnitude. Samples with large T^2 define unusual observations within the model.

Finally, in PCA, the loading vectors that encode the largest sources of variation in **R** are typically used to calculate the scores. If the analyte is a minor component in the

sample matrix with respect to the overall variance in the spectra, the calculation of the PCA scores may not necessarily have extracted the analyte information most efficiently.

For example, in the case of *in vivo* spectra collected from tissue, glucose is a minor component with respect to the overall spectral variance. In this case, the glucose information may reside in one of the loading vectors associated with a small eigenvalue, while other latent variables with large eigenvalues may not be useful in building a calibration model for glucose. Thus, part of the necessary protocol in building a model with PCR is to assess which PCs are most useful in model building. This characteristic of PCA in which no guidance is provided regarding the specific information to be extracted from the input data motivates the use of PLS, a related latent variable method. The PLS technique is discussed in the next section.

Partial-Least Squares

Partial least-squares is a latent variable method widely used in analytical chemistry applications such as the quantitative analysis of ultraviolet,[130] NIR,[90] chromatographic[131] and electrochemical[132] data. In PLS, construction of the latent variables employs both the concentration information and the experimental (spectral) information as opposed to PCA which uses only the spectral information. The PLS method computes the components using both spectral and concentration data simultaneously, maximizing the covariance between the two while in PCA only the variance of the spectral data is considered. [123] The PLS calculation assumes errors associated with both the experimental information and the concentration information are of equal importance while PCA assumes that the concentration data are error free.[110] This

assumption is very important because the preparation of the calibration samples is subject to error (e.g., dilution and weighing) just as much as the spectral data collection. This is especially true in cases in which the calibration samples are not prepared but rather are collected and then analyzed with a reference method. Data collected *in vivo* from tissue would fit this category.

The decomposition of the spectral and concentration information can be implemented by

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{3.17}$$

$$\mathbf{c} = \mathbf{T}\mathbf{q} + \mathbf{e} \tag{3.18}$$

In Eq. 3.17, **X** is the $n \times p$ matrix representing the instrumental responses and **c** in Eq. 3.18 is the $n \times 1$ vector of concentrations. The **X** matrix is typically mean-centered as described previously in the discussion of PCA. The matrices, **T** ($n \times h$), **P** ($p \times h$) and **E** ($n \times p$) in Eq. 3.17 are analogous to those introduced previously in the discussion of PCA. The columns in matrix **P** contain h empirically derived spectral shapes that are termed spectral loadings or latent variables, the same as in PCA.

The vector **q** ($h \times 1$) in Eq. 3.18 contains concentration loadings. These define the relationship between the scores and concentrations and are analogous to the regression coefficients in Eq. 3.6. The key characteristic of PLS is that the same score matrix, **T**, is common to both Eqs. 3.17 and 3.18, and the calculation of **T** employs both **X** and **c**. The empirically derived h latent variables are thus used to extract the key information in the spectra that relates to the concentrations in **c**. The terms, **E** and **e**, in Eqs. 3.17 and 3.18 represent the spectral and concentration information, respectively, that is not extracted. These terms are called spectral residuals and concentration residuals, respectively. The

scores, analogous to those in PCA, are orthogonal. In the variant of PLS discussed, here, the loadings are not guaranteed to be orthogonal and are not normalized.

Among the different PLS algorithms available for calculating the scores and loadings, this thesis used PLS1, which uses another matrix called the "weight" matrix that has close analogies with the spectral loading matrix, **P**.[123,133] The decomposition of **X** and **c** computes a set of loading weights, **w**, for each spectral loading. The mean-centered spectral and concentration data are used for the calculation of the first set of loading weights as

$$\mathbf{w}_1 = \frac{\mathbf{X}^T\mathbf{c}}{\|\mathbf{X}^T\mathbf{c}\|} \tag{3.19}$$

where $\|\mathbf{X}^T\mathbf{c}\|$ represents the norm of the vector.

The projection of **X** onto $\mathbf{w}_1$ computes the first score vector, $\mathbf{t}_1$, as shown in Eq. 3.20, and the first spectral loading $\mathbf{p}_1$ is computed as in Eq. 3.21. Similarly the first concentration loading is given by Eq. 3.22.

$$\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1 \tag{3.20}$$

$$\mathbf{p}_1 = \frac{\mathbf{X}^T\mathbf{t}_1}{\|\mathbf{t}_1{}^T\mathbf{t}_1\|} \tag{3.21}$$

$$\mathbf{q}_1 = \frac{\mathbf{t}_1{}^T\mathbf{c}}{\|\mathbf{t}_1{}^T\mathbf{t}_1\|} \tag{3.22}$$

The information in **X** and **c** that is not extracted by $\mathbf{t}_1$ can be computed by

$$\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1{}^T \tag{3.23}$$

$$\mathbf{e}_1 = \mathbf{c} - \mathbf{t}_1\mathbf{q}_1 \tag{3.24}$$

The next loading weight can then be computed by replacing **X** and **c** in Eq. 3.19 with $\mathbf{E}_1$ and $\mathbf{e}_1$, respectively.

The algorithm can iterate in this manner to calculate *h* factors. Taken together, the *h* loading weights define the *p*×*h* weight matrix, **W**. Once the PLS scores are calculated, the calibration model is now built using the scores as independent variables. The procedure is identical to that discussed previously for PCA.

The same procedure used in PCA is employed with PLS1 for the prediction of unknown concentrations. Responses for *m* unknown samples are collected to produce $\mathbf{R}_{pred}$ (*m*×*p*), and the prediction response matrix is centered using the mean of the calibration data. The first PLS score vector, $\mathbf{t}_{1,pred}$**,** for the prediction data can be calculated using the first loading weight vector from the calibration. The contribution from the computed first prediction PLS score is removed to calculate the residuals. This calculation uses the first calibration spectral loading. These steps can be summarized by Eqs. 3.25 and 3.26 as:

$$\mathbf{t}_{1,pred} = \mathbf{R}_{pred}\,\mathbf{w}_1 \tag{3.25}$$

$$\mathbf{R}_{1,pred} = \mathbf{R}_{pred} - \mathbf{t}_{1,pred}\,\mathbf{p}_1^{T} \tag{3.26}$$

This process can iterate until *h* prediction PLS scores are computed. The unknown concentration can then be calculated by use of the regression coefficients computed in the calibration step. This can be done in exactly the same way as described previously for models based on the use of PCA.

An alternate implementation of the prediction step is to use **W**, **P**, and **q** to compute a regression coefficient vector, **b**:

$$\mathbf{b} = \mathbf{W}(\mathbf{P}^{T}\mathbf{W})^{-1}\mathbf{q} \tag{3.27}$$

The inner product of this regression coefficient vector and the centered (using the mean calibration response) response vector for the unknown sample produces the

estimated concentration of the unknown. If **c**, the vector of concentrations of the calibration samples was centered prior to the calculation of **W**, **P**, and **q**, the estimated concentration must be translated to the original scale by adding the mean calibration concentration.

Because the PLS method uses both spectral and concentration information to build the calibration model, there is a danger of overfitting the model to the calibration data. If the proper number of latent variables and the appropriate wavelength range are not used in the model, false calibrations based on spurious correlations between concentrations and spectral features within the data set can occur. Even random noise or spectral artifacts may be modeled and assumed to represent valid spectral features. Conventional statistical testing for the significance of model terms may fail because the extracted latent variables, by the nature of the PLS algorithm, do in fact correlate with the dependent variable of concentrations.

While many factors may appear to be significant when building the regression model with the calibration data, not all of those factors will likely be useful when the model is applied to future data. This issue is complicated by the common use of optimization procedures that vary both the spectral range (i.e., the variables input to PLS) and number of latent variables computed by PLS. Investigating many combinations of input variables and computed latent variables increases the likelihood that apparently well-performing models will be found that happen to incorporate fortuitous correlations.

Thus, model testing and critical evaluation are essential elements of the use of PLS. For example, the selected wavenumber range must be consistent with knowledge of the spectral responses of the analyte of interest and the known components of the sample

matrix. If a spectral region used in building a model is high in noise, there is a good chance that PLS has found chance correlations among the noise features. Examination of the components in the **W** and **P** matrices is important to assess the degree to which they appear to incorporate noise rather than valid spectral features. Finally, testing models with data not included in the calibration calculations is important with any modeling methodology, but especially important with PLS.

Performance Diagnostics

For any of the modeling methods discussed previously, the quality of the calibration model is assessed by computing the error in concentration associated with it. The calibration error is termed the root-mean-squared error of calibration (RMSEC) or standard error of calibration (SEC). The SEC for the analyte of interest is

$$\text{SEC} = \sqrt{\frac{\sum_{i=1}^{n}\left(c_i - \hat{c}_i\right)^2}{n-h-1}} \tag{3.28}$$

where c_i is the reference concentration for sample i , and $\hat{c}_i$ is the concentration estimated by the calibration model. The root mean-squared error is computed for the n samples of the calibration set to produce the SEC. The n-h-1 term in the denominator in Eq. 3.28 reflects the degrees of freedom for n number of samples, h number of independent variables used for the regression and assuming an intercept (constant term) is fitted to the model. The “1” is deleted from Eq. 3.28 if the model is fitted without an intercept.

When the model is used to predict the concentration of m known samples that were not used in the calibration, the concentration error is computed as the standard error

of prediction (SEP) or root-mean-squared error of prediction (RMSEP). The SEP can be obtained by

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^{m}\left(c_i - \hat{c}_i\right)^2}{m}} \tag{3.29}$$

The denominator in Eq. 3.29 has no loss of degrees of freedom since the *m* samples were not used in the estimation of the model.

Model Validation

The ultimate goal of modeling is to use the computed calibration model to predict the concentrations of spectra collected in the future. Model validation is an important step in model development to help guarantee a precise and accurate prediction. The computed calibration models can be validated with either internal or external approaches.

Internal Validation

In internal validation, only the calibration data are used to test the model. The calculation of SEC is one of the approaches to internal validation which provides a rough estimation of model performance. However, especially with the PCR and PLS approaches, the SEC is almost always improved by increasing the model dimensionality. Therefore, internal validation alone can be misleading. One other common approach to internal validation is cross-validation (CV).

In CV, a subset of the calibration data is withheld to evaluate the prediction performance of the model. A leave-one-out CV involves leaving one sample out for the computation of the standard error while the rest of the samples are used to build the calibration model.[134] This step is cycled through the calibration data such that each

spectrum is withheld from building the model and predicted once. The standard errors calculated from each step using Eq. 3.29 are pooled together to compute a standard error of cross-validation (SECV). This statistic is also termed the root-mean-squared error of CV (RMSECV).

If the calibration data set contains a large number of samples, a leave-fraction-out cross validation can be performed to save time (e.g., 10 groups of 10 % of the total samples). The fraction may be contiguous groups in time, randomly selected groups, or subsets defined through a structured selection pattern (e.g., every 10^{th} spectrum). Cross-validation is a technique used throughout this research.

To optimize the terms used in the calibration model, CV is performed with different model sizes and different combinations of the independent variables. For models built with PCR or PLS, the typical procedure is to begin with the first latent variable and sequentially add terms up to a maximum of *h* latent variables. Typically, the SECV values level off and then increase after the optimal number of components has been calculated. The reason for this is that the later components model noise, and so samples withheld in each of the calibration steps used in CV are predicted worse when more components are used in the model.[110] This is termed overfitting. Overfitting has the effect of adding predictors that perform no useful function and therefore increases the possibility for undetected errors that can lead to mistakes in future predictions.

An *F*-test can be performed on the SECV values to evaluate whether a lower SECV value is statistically different from one based on a smaller model size (i.e., whether adding additional terms to the model has decreased the SECV in a significant way). In the work described here, *F*-tests were performed at the 95% probability level to

compare values of SECV to the minimum SECV found ($SECV_{min}$) . The equation for the *F*-test can be given as:

$$F = \frac{(\text{SECV})^2}{(\text{SECV}_{min})^2} \quad (3.30)$$

The significance of this *F*-value can be tested with the critical value at (*n, n*) degrees of freedom, where *n* is the number of calibration observations (i.e., the number of CV predictions). In chemometrics, smaller models with equivalent performance are generally thought to be superior. It is thought that as this model relies on fewer fitted or computed parameters, it should be both more reliable and easier to interpret.

External Validation

With internal validation, using a subset of the calibration data set for assessing model performance may result in an overly optimistic evaluation. This result arises primarily because the calibration data are typically collected over a short time span. Thus, issues of instrumental drift or other time-based sources of data variance have not been incorporated into the assessment of model performance. By contrast, if the prediction data set is collected outside the timeframe of the calibration data, if time-correlated issues, such as instrumental drift or sources of chemical or physical variances are present, the calibration model may perform poorly since these variances have not been incorporated into the model building step. Hence, an external validation is an important way to assess the impact of these sources of time-based variance and thereby obtain a more realistic estimate of model performance. In external validation, the model performance is tested by use of one or more sets of data outside of the calibration data set and computing the SEP as given in Eq 3.29

Model Optimization

Optimization of the calibration model is a very important component of the model building procedure in order to identify the best spectral region to use in constructing the model and, for models based on PCR or PLS, the correct number of latent variables.[135-137] The optimal spectral region / latent variable combination is typically obtained using an internal validation method.

Grid search analysis is one of the common optimization methods in which a grid of spectral region / latent variable combinations is used, and the optimal combination is selected by performing CV. The combination that produces the lowest SECV is then used in subsequent work.[86] The research presented here used a grid search analysis and leave-fraction-out CV to find the optimal combination of spectral region and latent variables for use in constructing the calibration model. Here, the optimization procedure followed in Chapter 4 for the *in vitro* simulation data is taken as an example to discuss the stepwise procedures employed in grid search analysis.

In Chapter 4, a grid search is used based on sliding a window of fixed spectral width in 50 cm^{-1} increments across the 4900-4100 cm^{-1} range. At each step, a PLS model is constructed using 3-16 latent variables. By repeating this procedure with windows of different widths, the combination of window size, location, and the number of PLS factors can be examined. At each step, a leave-fraction-out CV procedure is applied, withholding 10 % of the data at each iteration for the prediction. The value of SECV is then computed for each combination of variables tested, and the *F*-test is used at the 95% level to assess the number of latent variables whose corresponding SECV is not statistically different from the minimum SECV found for that spectral segment. This

procedure produces a grid of SECV results that, when sorted, yield the most promising spectral regions and corresponding optimal numbers of latent variables.

Classification Methods

Qualitative approaches in NIR-spectroscopy use pattern recognition methods to classify samples according to their NIR spectra.[90] These classification methods are highly important in chemistry,[138] biology[139] and food sciences.[140,141] Two main types of classification methods can be identified: (1) supervised methods and (2) unsupervised methods. In unsupervised methods, samples are classified without a prior knowledge of their categories, whereas in supervised methods, a prior knowledge of the category of each sample is used to construct a classification model.[141]

The technique of PCA can be used as an example of an unsupervised classification method. A series of data objects (patterns) can be subjected to PCA, and the resulting scores can be plotted to identify which patterns are located in the same region of the data space. This approach can be used to uncover similarities and differences among a group of patterns. For example, in an NIR application, if spectra collected on a given day cluster together in a particular part of the score plot, the investigator learns that there is some commonality among these data patterns that distinguishes them.

There are many supervised pattern recognition methods available for use in building classification models. These include parametric and nonparametric linear discriminant analysis (LDA), artificial neural networks, and modeling techniques based on PCA such as SIMCA.[90] In supervised methods, a classification model is built first

with a set of samples with known categories (classes) of membership. This process is termed training and the samples used comprise a “training set”.

Two general strategies can be used to implement supervised pattern recognition methods. In the first approach, each data class is modeled to create a mathematical representation of the corresponding category. The SIMCA method referenced above employs this approach. For example, if the goal of the analysis were to identify alcohols on the basis of their spectral signatures, spectra of a set of known alcohols would be used to define an alcohol model. An unknown spectrum could then be classified on the basis of the degree to which it fits the alcohol model. By use of either parametric or non-parametric measures of the quality of fit, a decision can be made regarding whether or not the test object is a member of the alcohol data class.

The second general strategy for implementing a supervised classifier is to build a model that defines the regions between the data classes rather than the classes themselves. Methods of this type are termed discriminant methods because they serve to discriminate between the data classes that are defined by the training set. The classification model can be viewed as defining separating surfaces between the data classes.[12] The computed separating surfaces can be used to assign the membership of an unknown sample on the basis of its orientation relative to the separating boundary. In the analogy used previously, two data classes might be alcohols and non-alcohols. The classification model would define a boundary in the data space that separates the alcohol and non-alcohol classes. If an unknown spectrum is oriented on the alcohol side of the boundary, it would be assigned to that data class.

The LDA methods define linear separating surfaces and these methods are efficient for data classes whose boundaries can be represented effectively with a linear function. Often, however, the data used are more complex and it is hard to define linear separating surfaces that accurately represent the boundaries between the classes. In this case, artificial neutral networks or other nonlinear modeling techniques can be used to define nonlinear separating surfaces between the data classes. While very powerful, these nonlinear methods also suffer from slow optimization speed for large data sets, complexity in defining the architecture of the model (i.e., require optimization of parameters that define the model form), and susceptibility to becoming trapped in local optima during the optimization.[12]

In the dissertation research, piecewise linear discriminant analysis (PLDA) is used as an alternative to the nonlinear classification methods. To discriminate two data classes, this approach uses multiple linear functions to define a piecewise linear approximation to a nonlinear discriminant function.[142,143] On the basis of our experience, the PLDA technique optimizes faster, has a simpler architecture, uses fewer numerical coefficients, and is less susceptible to the effects of local optima than the true nonlinear classifiers. Next, a detailed description of PLDA will be provided.

Piecewise Linear Discriminant Analysis

The PLDA method is an extension of nonparametric LDA and is used to compute multiple linear separating surfaces (discriminants). Taken together, these discriminants collectively approximate a nonlinear separating boundary to distinguish two data classes. In PLDA, the individual separating boundaries are computed in a stepwise manner. The

calculation of the first linear discriminant is similar to the procedure used in conventional nonparametric LDA. Each linear separating boundary separates different portions of the data set and collectively forms the separating surface.[12,13,143]

The methodology behind PLDA can be understood by use of LDA as a basis. For the purpose of this discussion, let **x** be an *h*-dimensional vector (pattern) corresponding to an observation that belongs to one of the two classes, w_1 or w_2. For the research discussed here, PLS scores were used as the patterns for the classification. Thus, **x** contains *h* PLS scores.

For the nocturnal hypoglycemic algorithm studies discussed here, w_1 and w_2 correspond to patterns for glucose concentrations above an alarm threshold value (non-alarms) and patterns that correspond to concentrations below the threshold value (alarms). Patterns identified as belonging to the alarm class would signal the patient is becoming hypoglycemic and should be wakened.

To separate these two classes, a linear discriminant function is constructed as

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_{\mathrm{o}} \tag{3.31}$$

The (*h-1*)-dimensional separating surface between the two data classes is defined by its *h*×1 normal vector, often termed the "weight vector". The weight vector is denoted by **w** in Eq. 3.31. Also present in the equation is w_{o}, which is a scalar bias or offset term that removes the requirement that the separating boundary pass through the origin of the data space.

The separating boundary, also called a discriminant or discriminant function, is a hyperplane in multidimensional space. The inner product between the weight vector and the pattern vector is termed the discriminant score (*y* in Eq. 3.31) and is used to assign

class membership to a pattern. For the particular application discussed in this thesis, if the discriminant score is higher than zero (i.e. $y > 0$), the pattern is assigned to the alarm class; otherwise it is assigned to non-alarm class.

The discussion presented above is based on the calculation of a single linear discriminant function. The PLDA method extends this approach by calculating additional linear discriminants that define the piecewise linear sections of an approximately nonlinear separating boundary.

The principles of PLDA can be illustrated with a two-dimensional representation as shown in Figure 3.1. In this example, each pattern would be represented by a two-dimensional vector (e.g., contain two PLS scores) and can thus be plotted in a simple *x*-*y* coordinate system. Here, the alarm patterns are represented by triangles while the non-alarm patterns are represented by ovals. The first discriminant marked as "1" is computed to separate as many alarm and non-alarm patterns as possible, with the constraint that the alarm side of the discriminant boundary must be "pure" (i.e., contain only alarm patterns). This leaves a mixed set of alarm and non-alarm patterns on the other side of the boundary. A discriminant that obeys this requirement is termed "single-sided".

Patterns separated on the pure side of the first discriminant are removed from the computation of the second discriminant, labeled "2" in Figure 3.1. The second discriminant is computed to separate as many alarm patterns as possible, again leaving a pure side and a mixed side. This process can continue, with each step producing a single-sided discriminant and a remaining mixed class of alarm and non-alarm patterns. Typically, the stopping point for the discriminant calculation is determined by the

number of additional alarm patterns separated by the new discriminant. The calculation stops if the number of patterns separated by the new discriminant is judged insignificant.

For the results presented for a fairly simple sample matrix in Chapter 4, the first discriminant was found to separate a large portion of the data. For more complex samples investigated in Chapters 6 and 7, two or three discriminants were found to be significant. Presumably, the more complex samples produced a data space in which the class boundaries were correspondingly more complex.

In the implementation of PLDA used in this research, a Bayes classification algorithm was used to calculate an initial approximation to the discriminant.[113,144] This is a direct calculation that employs the class means and the covariance matrix formed from the patterns (encoding an approximation to the class shapes). If this initial approximation was not single-sided, a simple translation algorithm was used to move the boundary such that a pure alarm side was achieved.

The discriminant positioning was further refined by use of a modified simplex method.[12,13,145,146] In the simplex optimization, a response function is calculated that numerically encodes the degree to which the discriminant is optimally positioned in the data space (i.e., separates as many alarm patterns as possible from those remaining in the mixed-class subset). The calculation of the response function is performed in an iterative manner, employing a set of rules to move the discriminant location in search of a better placement (i.e., one that produces a higher value of the response function).

The implementation of the simplex algorithm has several parameters that relate to the step sizes used in altering the discriminant position and the number of iterations employed. Depending on the choice of parameters, the discriminant can become trapped

in a local optimum. For this reason, three separate piecewise linear discriminants were computed with different sets of training parameters. This yielded three replicate discriminants that were used together to form a committee of classifiers. The formulation of rules for operating these classifiers will be discussed in a subsequent section.

The discriminants computed with the training set are used to determine the class membership of an unknown pattern. A discriminant score for an unknown pattern is computed for all linear discriminants using Eq. 3.31. If any one of the multiple discriminants classifies the unknown pattern as an alarm, the pattern is placed in the alarm class. This can be understood if one considers that each discriminant function is responsible for modeling a different section of the separating boundary. If none of the discriminants classifies a pattern as alarm (i.e., it is on the non-alarm side of every section of the boundary), it is classified to be in the non-alarm class. If more than one discriminant is involved in determining the membership (i.e., the pattern is on the alarm side of multiple sections of the boundary), the highest discriminant score is used to represent the classification.

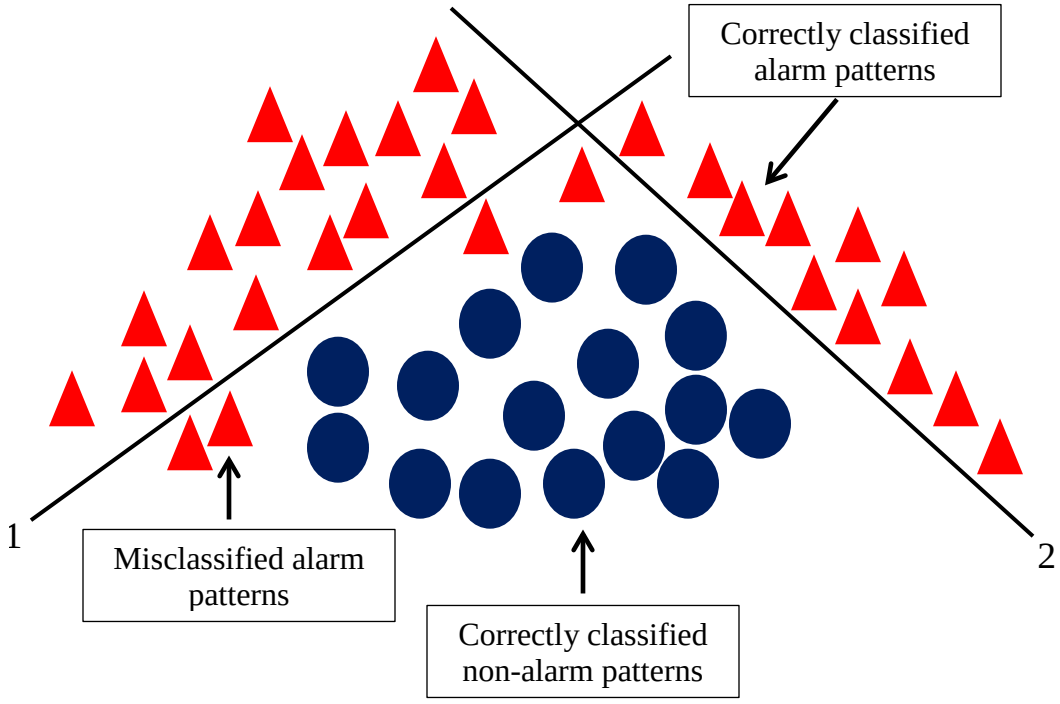


Figure 3.1 Pictorial representation of PLDA using two linear discriminants marked “1” and “2”. These discriminants collectively form an approximation to a nonlinear separating surface between the two data classes (alarm and non-alarm). The alarm and non-alarm patterns are represented as triangles and ovals, respectively.

Nocturnal Hypoglycemic Alarm Algorithm

The objective of the research discussed in this dissertation is to evaluate the potential for implementing a noninvasive nocturnal hypoglycemic alarm with NIR spectroscopy. The key elements of this algorithm are: (1) collection of a calibration database of NIR spectra and associated reference glucose concentrations that can be used subsequently in the construction of classification models that allow spectra to be assigned membership in the hypoglycemic (alarm) and non-hypoglycemic (non-alarm) data classes, (2) collection of a reference glucose concentration by use of a conventional fingerstick measurement at the start of the sleep period, (3) collection of a reference NIR spectrum at the same time the reference glucose concentration is measured, (4) use of this reference spectrum as the spectral background in the calculation of absorbance values for spectra collected subsequently, (5) definition of a critical glucose concentration that specifies the change in glucose concentration relative to the reference concentration that will cause the patient's glucose level to reach the alarm threshold concentration for hypoglycemia, (6) use of the calibration database to construct a classification model that allows spectra to be grouped into the alarm and non-alarm classes on the basis of the critical glucose concentration, and (7) collection of spectra continuously during the sleep period and classification of these spectra in real time as belonging to the alarm or non-alarm data classes. If a spectrum is placed into the alarm class, the patient is judged to be in the hypoglycemic state and an alarm is sounded.

As indicated above, the spectra submitted to the classification algorithm are in absorbance units relative to the reference spectrum collected at the start of the sleep period. The motivation for the use of this reference spectrum as the absorbance

background is to remove common spectral features that originate from the sample matrix. A complication that arises, however, is that glucose is present in both spectra and thus the absorbance spectrum that results from taking the ratio no longer has a glucose signal intensity that corresponds to its original concentration. We term this generated spectrum a differential absorbance spectrum and specify its new effective concentration as the differential concentration.

The effective analyte concentration in the differential spectrum is equal to the concentration differences of the two original spectra that are used in the absorbance calculation. This concept can be explained using the derivation shown below. Given two single-beam intensities, I_1 and I_2, collected for samples 1 and 2, respectively, the absorbance for the two samples can be calculated using a background single-beam intensity, I_0 , according to the Beer-Lambert law

$$-log_{10}\left(\frac{I_1}{I_0}\right)=abc_1 \tag{3.32}$$

$$-log_{10}\left(\frac{I_2}{I_0}\right)=abc_2 \tag{3.33}$$

In Eqs. 3.32 and 3.33, the terms c_1 and c_2 correspond to the concentrations of the analyte in samples 1 and 2, respectively, and *a* and *b* denote the absorptivity and path length. The wavelength dependence of I_1, I_2, I_0 and *a* is omitted for simplicity. For the sake of this derivation, these equations further assume that the samples contain only a single absorbing species.

The difference between the two equations is given by

$$-log_{10}\left(\frac{I_1}{I_0}\right)+log_{10}\left(\frac{I_2}{I_0}\right)=abc_1-abc_2 \tag{3.34}$$

Equation 3.34 can be rearranged to

$$-log_{10}(I_1)+log_{10}(I_0)+log_{10}(I_2)-log_{10}(I_0)= ab(c_1-c_2) \quad (3.35)$$

Canceling the terms containing I_0 and rearranging the equation yields:

$$-log_{10}\left(\frac{I_1}{I_2}\right) = ab(c_1-c_2) \quad (3.36)$$

As shown in Eq. 3.36, taking the negative logarithm of the ratio of two single-beam spectra computes a differential spectrum in absorbance units and the concentration is equal to the differences in concentrations of the corresponding spectra (i.e., numerator concentration – denominator concentration).

This calculation can be further extended for multicomponent systems that follow a linear mixture model. In this case, the negative logarithm of the ratio of two single-beam spectra will compute a differential spectrum corresponding to a sum of differences in concentrations of each absorbing species in the two samples.

The differential spectral calculation is based on several assumptions. The derivation assumes that the optical path length *b* does not change across the data collection. This suggests that special care should be taken to minimize path length variation in noninvasive glucose measurements. The derivation further assumes that the background information is the same for all the spectra collected. If this is not the case, instead of a single I_0 term, there would be $I_{0,1}$ and $I_{0,2}$. A differential background term would then be introduced into the computed absorbance spectrum.

Spectra can be used interchangeably in the numerator and denominator in the calculation of the differential spectrum. As a consequence, the differential concentrations can be either positive or negative. In the work described here, differential spectra were computed such that only negative concentration differences resulted. This was done for

simplicity when implementing the nocturnal hypoglycemic alarm algorithm in which it is assumed that the reference concentration taken at the beginning of the sleep period is always above the hypoglycemic concentration threshold. Thus, differential spectra computed relative to this reference will have effective negative concentrations on both sides of the alarm threshold.

For example, consider the case in which the alarm threshold concentration is 3.0 mM and the reference concentration obtained at the start of the sleep period is 5.0 mM. If the reference spectrum corresponding to the reference concentration is used in the denominator of the absorbance calculation, differential spectra just below and just above the alarm threshold will have differential concentrations near 3.0 – 5.0 = -2.0 mM.

Two example differential spectra corresponding to differential glucose concentrations of 20.0 mM and –20.0 mM are shown in Figure 3.2. Both positive and negative spectral features can be observed depending on the differences in the concentrations of the two spectra whose ratio is computed.

The steps used in building the calibration database are shown in Figure 3.4. For *in vivo* studies, this would involve an experiment in which the blood glucose level is caused to change over a specified range, conventional invasive reference glucose measurements are made at fixed time intervals, and NIR spectra are collected continuously. In the ultimate implementation with a human subject, this calibration step would be performed in a physician office by use of a protocol analogous to an oral glucose tolerance test (OGTT). During an OGTT, glucose fluctuations can be induced by administering glucose or insulin as required to achieve the desired excursions in concentration.

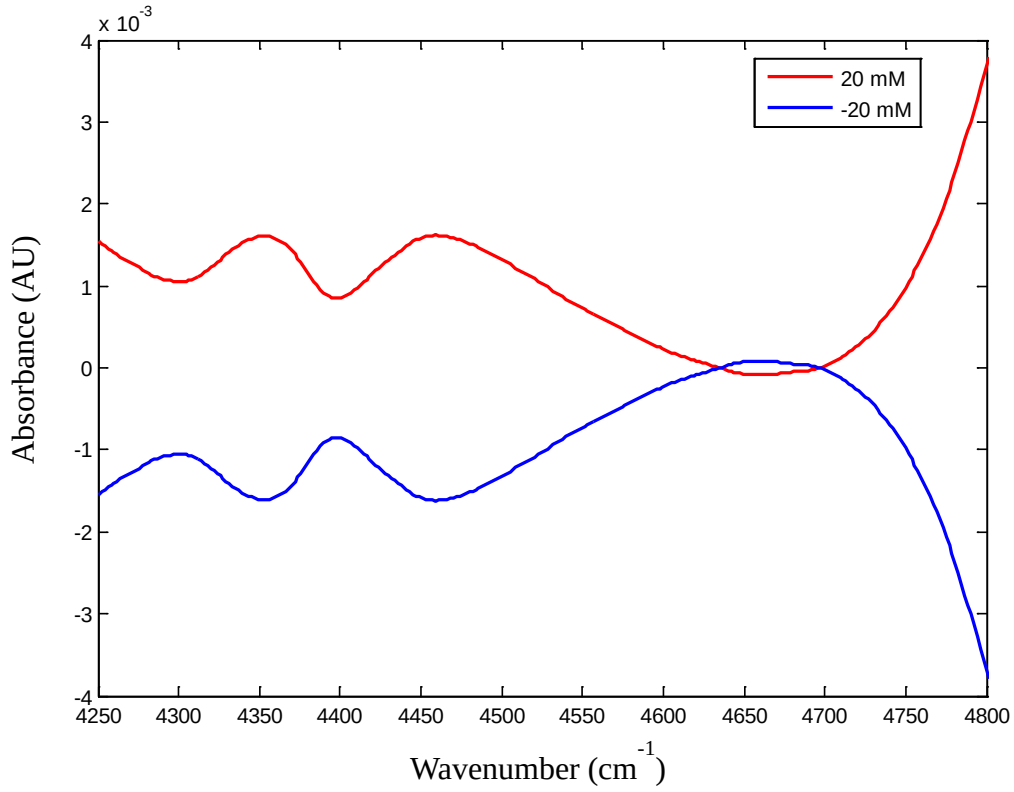


Figure 3.2 Differential spectra of glucose in 0.1 M, pH 7.4 phosphate buffer for both positive and negative concentrations of 20.0 mM, computed by taking the negative logarithm of the ratio of two single-beam spectra. The spectral features can be either positive or negative depending on the concentrations corresponding to the spectra used in the numerator and denominator of the absorbance calculation. Glucose combination bands near 4300 (C-H), 4400 (C-H), and 4650 (O-H) cm^{-1} are visible in the spectra.

To account sufficiently for background variations and incorporate these variations into the calibration database, an OGTT might have to be repeated several times.

Single-beam spectra are collected at a specified level of signal averaging and stored in blocks, which are contiguous groups of spectra corresponding to a selected time window. The block size specifies a time window in which the background variation is assumed to be negligible. Differential spectra computed within a block are assumed to have matching backgrounds and thus constant background features will have been reduced to zero absorbance.

Because of the need for reference glucose measurements made by a conventional invasive method, the number of spectra in which the actual glucose concentration is known will be limited. In the ultimate implementation of the methodology with a human subject, the need for a protocol analogous to an OGTT will also limit the number of available spectra as the patient will only be willing to be subjected to the procedure for a limited time. Thus, the calibration database acquired will have fewer spectra, fewer glucose levels, and fewer reference glucose measurements than would be desirable from the standpoint of data analysis.

The use of differential spectra has an additional advantage for the case in which the calibration database is limited. By computing differential spectra from all combinations of the single-beam spectra collected within a time block, the calibration database is expanded to fill in additional levels of glucose concentration, as well as some additional variation in the non-constant background components that are not removed by the differential absorbance calculation.

The PLS algorithm is then applied to the differential spectra in the calibration database in combination with the differential concentrations to reduce the dimensionally of the original spectra to an *h*-dimensional PLS score matrix (**T** matrix in Eq. 3.17). The reduction in dimensionality reduces the time required for the steps required in building the classification model used to identify alarm and non-alarm spectra. The computed spectral loadings (**P** matrix in Eq. 3.17) and loading weights (**W** matrix in Eq. 3.27) are saved for the calculation of PLS scores for unknown spectra collected in the future when the classification model is put into operation.

Once the calibration database is assembled, the next step is to calibrate the alarm algorithm. The steps of this procedure are summarized in the flow chart shown in Figure 3.5. The alarm threshold concentration, C_{alarm}, is user specific, but for this research the hypoglycemic threshold was defined to be 3.0 mM. If a spectrum represents a blood glucose level that is equal to or lower than this hypoglycemic threshold, an alarm would trigger to wake the sleeping patient. Once the alarm threshold concentration is defined, the next step is to partition the calibration database (i.e., as represented by the PLS scores) into alarm and non-alarm groups.

The differential concentrations can be used to identify the alarm and non-alarm patterns. As discussed previously, the calculation of the differential spectra is performed to yield negative differential concentrations. To identify the alarm and non-alarm spectra within the differential concentrations that comprise the calibration database, a negative threshold concentration needs to be defined. As defined in Eq. 3.36, the difference between the alarm threshold concentration and the reference concentration (C_{ref}) measured at the start of the sleep period is defined as the critical concentration, C_{crit}:

$$C_{crit} = C_{alarm} - C_{ref} \quad (3.37)$$

Assuming that C_{ref} is not in the hypoglycemic range, C_{crit} will always be negative in sign. This critical concentration identifies the alarm point in the context of the future differential spectra computed with respect to the reference spectrum. As an example, if C_{ref} is 4.0 mM, C_{crit} = 3.0 mM – 4.0 mM = -1.0 mM. Any differential spectra having a differential concentration below -1.0 mM will trigger an alarm.

The calibration database is partitioned into alarm and non-alarm classes on the basis of C_{crit}. If any differential concentration is lower than this critical concentration, the corresponding PLS score vector (pattern) is placed into the alarm class; otherwise it is placed into the non-alarm class. The alarm decision is thus a classification problem in which patterns are classified into either the alarm or non-alarm classes.

Once the calibration database is partitioned, the next step is to compute the separating surfaces between the alarm and non-alarm classes by use of PLDA. As described previously, three replicate classifiers are computed by changing the parameters associated with optimizing the positioning of the individual discriminants. The obtained weight vectors (**w** in Eq. 3.31) are saved for the determination of the class memberships of unknown patterns.

The steps in the operation of the alarm algorithm are summarized in Figure 3.6. Spectra are collected continuously over time while the patient is sleeping. The ratio of each spectrum to the collected reference is taken, forming a differential spectrum corresponding to the signed difference in concentration relative to the reference. After projecting each differential spectrum collected at time, *t*, to the previously computed PLS factors, an *h*-dimensional spectral pattern (i.e., the PLS score vector), $\mathbf{t}_{dif,t}$ is obtained.

Using the previously computed discriminants, the pattern $\mathbf{t}_{dif,t}$ is classified into either the alarm or non-alarm classes.

As discussed in the description of PLDA, the discriminant score calculated using Eq. 3.31 determines the class membership of the pattern, $\mathbf{t}_{dif,t}$. If the discriminant score is higher than zero, the corresponding pattern belongs to the alarm side of the separating boundary while a zero or negative discriminant score corresponds to a pattern on the non-alarm side. This research computed three replicate classifiers. To be classified as an alarm, two of the three replicate classifiers had to place the pattern in the alarm class.

In the following chapters, the nocturnal hypoglycemic alarm algorithm is evaluated by the use of spectra collected during both *in vitro* and *in vivo* studies that simulate a continuous monitoring application.

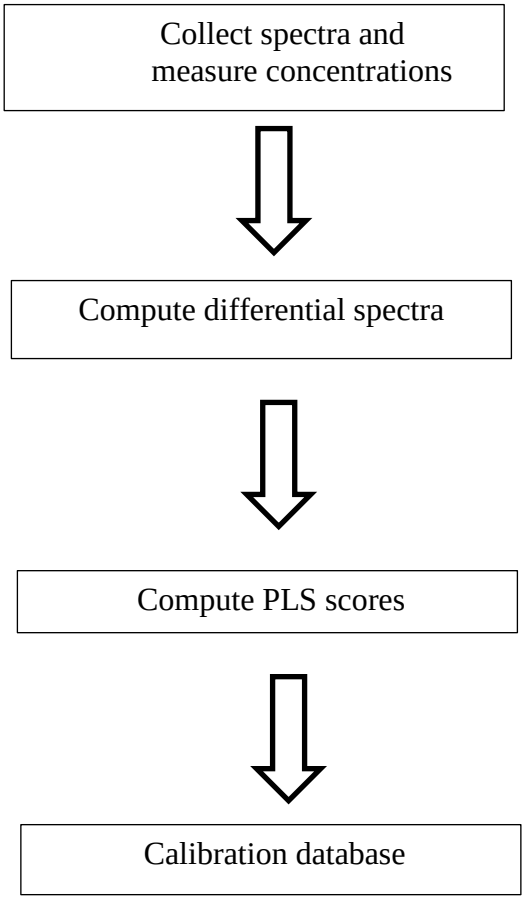


Figure 3.3 Steps used in building the calibration database. The calibration database consists of a PLS score matrix ($n \times h$) computed from using n differential spectra and corresponding differential concentrations to produce h PLS latent variables.

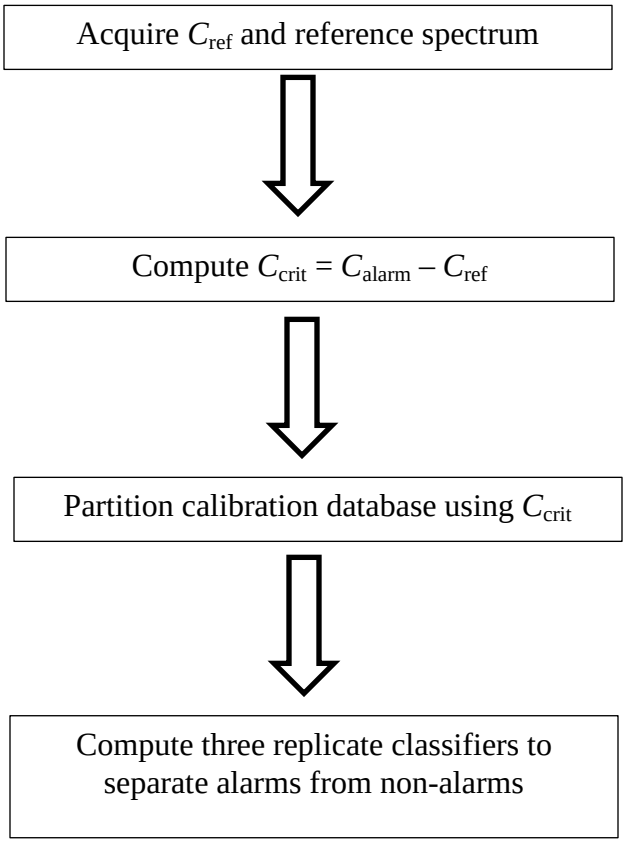


Figure 3.4 Steps in the calibration procedure for the nocturnal alarm. In the diagram, C_{ref} is the reference glucose concentration obtained at the start of the sleep period. The alarm threshold concentration, C_{alarm}, is 3.0 mM for these experiments. The difference between the C_{alarm} and C_{ref} is termed the critical concentration, C_{crit}.

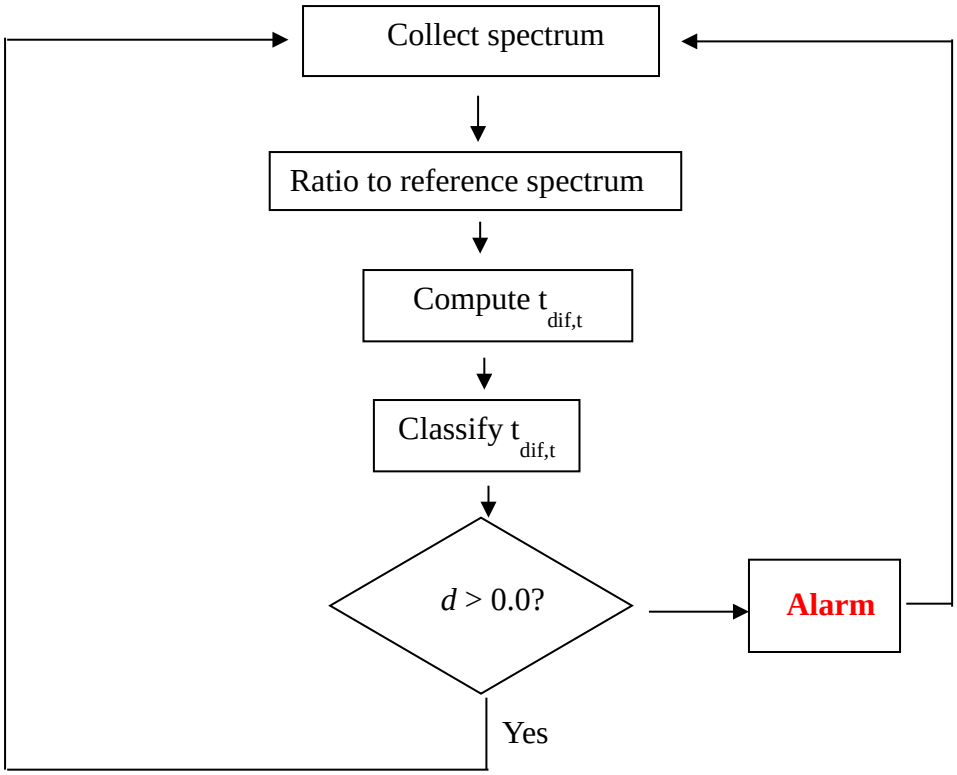


Figure 3.5 Flow chart of the operation of the alarm. A spectrum is collected at time *t* and the ratio is taken to the reference spectrum to compute a differential spectrum. Projection of the differential spectrum onto the calibration PLS factors yields a pattern (i.e., $\mathbf{t}_{dif,t}$) which is classified using the previously computed discriminants. If the pattern is classified into the non-alarm class, the process repeats. If the pattern is placed into the alarm class, an alarm is sounded to wake the patient.

CHAPTER 4

IN VITRO SIMULATION STUDIES FOR THE DEVELOPMENT OF A NOCTURNAL HYPOGLYCEMIC ALARM ALGORITHM BASED ON NEAR-IFRARED SPECTROSCOPY

Introduction

As described in Chapter 2, NIR spectroscopy is a promising approach for noninvasive glucose sensing. While the ability to measure glucose noninvasively must be demonstrated ultimately in *in vivo* experiments, feasibility studies based on *in vitro* simulations can provide valuable information during the development of experimental protocols and data analysis methodology.

In this chapter, an initial evaluation of the nocturnal hypoglycemic alarm algorithm described in Chapter 3 was performed. A dynamic system based on varying concentrations of glucose, urea, lactate, and triacetin in phosphate buffer was implemented by use of a set of peristaltic pumps. While this chemical system was not designed to mimic a specific biological matrix, the species studied are either blood constituents or they mimic blood constituents as in the case of triacetin (glyceryl triacetate), a soluble short-chain fatty acid ester. This matrix provides a test regarding whether the spectral features of glucose are sufficiently selective relative to other chemical components of the biological matrix to allow effective quantitation in the combination region of the NIR.

The NIR absorption bands for glucose, urea, triacetin, and lactate in the combination region were reported by Arnold *et.al* and all had absorptivities on the order

of 10^{-4}AU/mm-mM.[70] Spectra of these species were presented previously in Figure 2.2. The low absorptivity values observed in the combination region limit the detection to only major substances within the biological matrix. As a general rule, any substance must have at least a concentration of 1 mM to be quantified by NIR spectroscopy.[88]

This inability to measure components present below mM concentrations improves the measurement selectivity for glucose by minimizing the spectral contributions of other endogenous molecules within the biological matrix whose concentrations are below the mM range. Only major chemical components that show similar absorptivities and overlapping spectral features to the analyte of interest must be considered for the purpose of assessing selectivity. Thus, while the matrix studied in this work had a limited number of components, the species investigated can all be present in the mM range and thus represent realistic interferences to the measurement of glucose.

Experimental Methods

Near-infrared spectra used for this study were collected during 14 one-day data collection sessions by using two dynamic systems (DS 1 and DS 2), each consisting of four chemical components in phosphate buffer. The individual data groups will be termed runs 1 to 14. If run 1 is defined as time zero, runs 2-14 were conducted approximately 1, 2, 2, 3, 4, 21, 26, 28, 30, 56, 57, 58 and 175 weeks later. Thus, in its entirety, the data collection spanned approximately 3.7 years. To mimic the procedures described in Chapter 3 for the calibration and implementation of the nocturnal hypoglycemic alarm, the data were concatenated across the individual sessions to enable the construction of a

calibration database and to provide subsequent data sets that allowed the operation of the alarm to be simulated.

Reagents

Phosphate buffer was prepared by dissolving appropriate amounts of dry monobasic sodium phosphate ($NaH_2PO_4 \cdot H_2O$) (ACS reagent, Fisher Scientific, Fair Lawn, NJ) and a preservative, sodium benzoate ($C_6H_5COO_2Na$) (ACS reagent, Fisher Scientific) in 18.2 MΩ water purified by a Labconco water purification system (Labconco, Inc., Kansas City, MO). The buffer was titrated to pH 7.4 with 50 % w/w sodium hydroxide (Fisher Scientific, Fair Lawn, NJ). The final phosphate buffer concentration was 0.1 M and the benzoate concentration was 5.0 g/L.

Data for DS 1 were collected during five different days. Stock solutions of α-D-glucose (ACS reagent, Fisher Scientific), triacetin (ACS reagent, Sigma-Aldrich, St. Louis, MO) and urea (ACS reagent, Fisher Scientific) were prepared in the phosphate buffer. The stock solution concentrations used with each run will be given in a subsequent section. Each of the stock solutions in DS 1 contained 10 mM sodium L-lactate (ACS reagent, Sigma-Aldrich).

Dynamic system 2 was composed of α-D-glucose, glyceryl triacetate, and L-lactate prepared in phosphate buffer. Stock solution concentrations will be given below. Each of the stock solutions in DS 2 contained 10 mM urea. As noted above, the data collected from these two dynamic systems were mixed together to create an overall data set for use in testing the nocturnal hypoglycemic alarm.

Apparatus and Procedures

To simulate the glucose excursions that occur in the body, three stock solutions were mixed in different ratios using three peristaltic pumps (Rabbit-Plus and Dynamax Models, Rainin Instrument Co., Woburn, MA). A maximum flow rate of 28.2 mL/min can be obtained with these pumps, when connected with polyvinyl chloride (PVC) tubing (inner diameter (ID) of 3.16 mm) and with a maximum pump speed of 48 revolutions per minute (rpm).

A pump calibration was performed at the beginning of each day of data collection to correct for any differences in pump performance or deficiencies in tubing. Each of the pumps was connected with a buffer solution and ran at 4 rpm for 4 minutes and the volume of the buffer solution pumped was collected into a 10 mL measuring cylinder. The manufacturer's theoretical flow rate (i.e., based on the pump speed and tubing specifications) was used to compute the theoretical volume for the pump speed corresponding to each of the concentration levels. This theoretical volume and the actual volume collected at 4 rpm were used to adjust the pump speeds to compensate for any deficiencies. The total pump speed (i.e., the sum of the speeds of the three pumps) was maintained at a value of 10 rpm.

The three peristaltic pumps used for the study were computer controlled by use of Rainin Pump Control software (Version VI, Waterville Analytical, Waterville, MA) running under LabVIEW Version 7.1 (National Instruments Corporation, Austin, TX). Each stock solution was transferred to a 1 L polyethylene bottle and placed in a water bath regulated at ~55 °C to keep the flowing liquid temperature at 36.6-37.2 °C. The solutions were connected to the pumps by PVC tubing with 3.16 mm ID. Two Y-

connectors (Cole-Parmer Instrument Co., Vernon Hills, IL) were used to connect the tubing from the pumps into a single output line. The solution exiting the Y-connectors flowed through Tygon tubing containing an in-line mixer (0.48 cm in diameter and 11.6 cm in length, Cole-Parmer Instrument Co.), and then through insulated vinyl tubing to the spectrometer.

By changing the pump speeds, and therefore the flow rates of each of the solutions, the concentrations of the solutions exiting the mixer were varied. The concentrations of each component of the solution exiting the mixture can be calculated as shown in Eq. 4.1.

$$C_2 = \frac{(C_1 R_1)}{(R_1 + R_2 + R_3)} \tag{4.1}$$

In the equation, C_1 is the concentration of the glucose stock solution, R_1 is the pump speed for the glucose pump, and R_2 and R_3 are the pump speeds for the other two stock solutions. The stock solution concentration, total pump speed and glucose pump speed thus determine a desired glucose concentration. The concentrations of the other components can be computed in an analogous manner.

The data collected with DS 1 (runs 1 to 5) and runs 1 to 6 from DS two were used to form the calibration database. Runs 7 to 9 from DS 2 were used as prediction sets to simulate the operation of the hypoglycemic alarm. The concentration values for each component per sample were assigned to minimize correlation between each constituent. Tables 4.1 A and B list the correlation coefficients between each pair of components in the calibration data in DS 1 and 2, respectively, while Tables 4.1 C, D and E list the corresponding correlation coefficients for the data comprising the three prediction sets.

These levels of correlation were judged to be acceptable from the standpoint of preventing fortuitous results based on chance correlations.

Figures 4.1 and 4.2 summarize the desired concentrations for glucose, urea, triacetin, and lactate for both DS 1 and the calibration data from DS 2. Figure 4.3 summarizes the desired concentration profiles for the three prediction sets. For runs 1-14, the glucose stock concentrations were approximately 10, 25, 35, 30, 10, 20, 50, 20, 40, 25, 10, 10, 10 and 10 mM respectively. For runs 1-5, the stock concentrations for urea were 20, 15, 20, 15 and 15 mM, while for triacetin it was 45.8 mM for all. For runs 6-14, the stock solution concentrations for lactate were 10, 20, 10, 10, 10, 10, 25, 10 and 20 mM, respectively. The corresponding stock solution concentrations for triacetin in these runs were 45.8, 45.8, 22.9, 36.6, 45.8, 45.8, 22.9, 45.8 and 45.8 mM.

Table 4.1- Correlation coefficients

(A) Correlation coefficients for DS 1 (runs 1-5)

	Urea	Triacetin
Glucose	-0.34	-0.29
Urea		0.12

(B) Correlation coefficients for DS 2(runs 6-11)

	Urea	Lactate
Glucose	0.26	0.22
Lactate		0.03

Table 4.1 Continued.

(C) Correlation coefficients for prediction set 1

	Urea	Lactate
Glucose	-0.46	-0.42
Lactate		-0.62

(D) Correlation coefficients for prediction set 2

	Urea	Lactate
Glucose	-0.40	-0.44
Lactate		-0.65

(E) Correlation coefficients for prediction set 3

	Urea	Lactate
Glucose	-0.57	-0.43
Lactate		-0.50

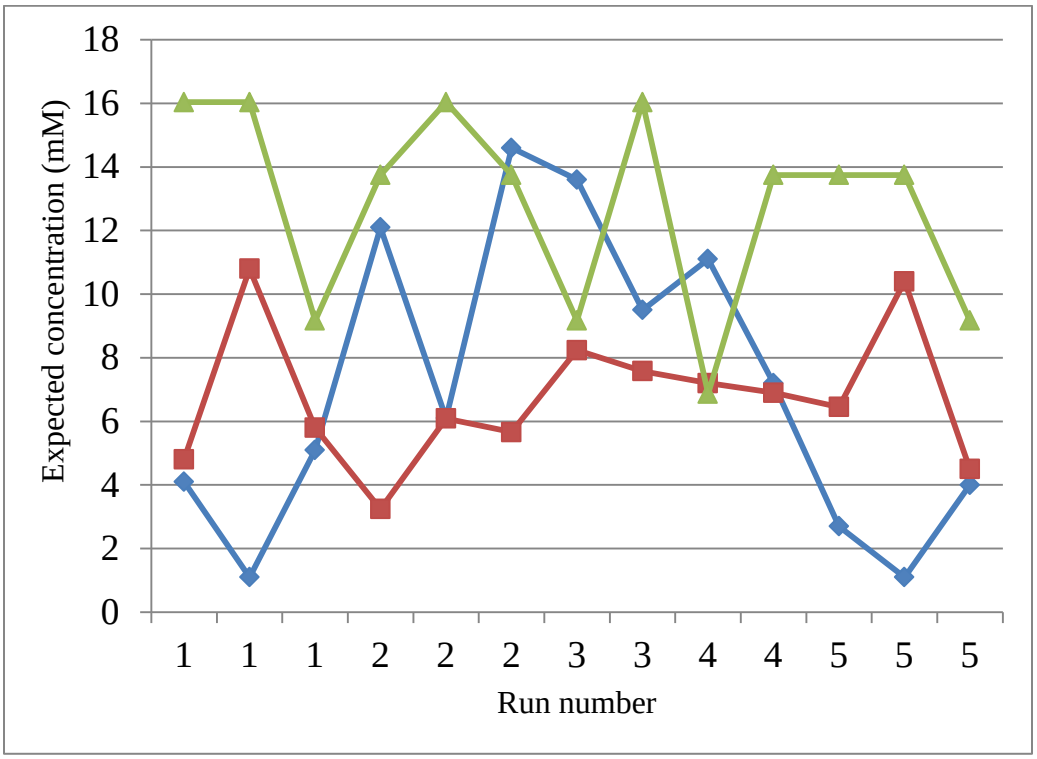


Figure 4.1 Expected concentration profile for DS 1. Each solution contained a constant amount of 10 mM lactate. Blue, red, and green traces correspond to glucose, urea, and triacetin, respectively.

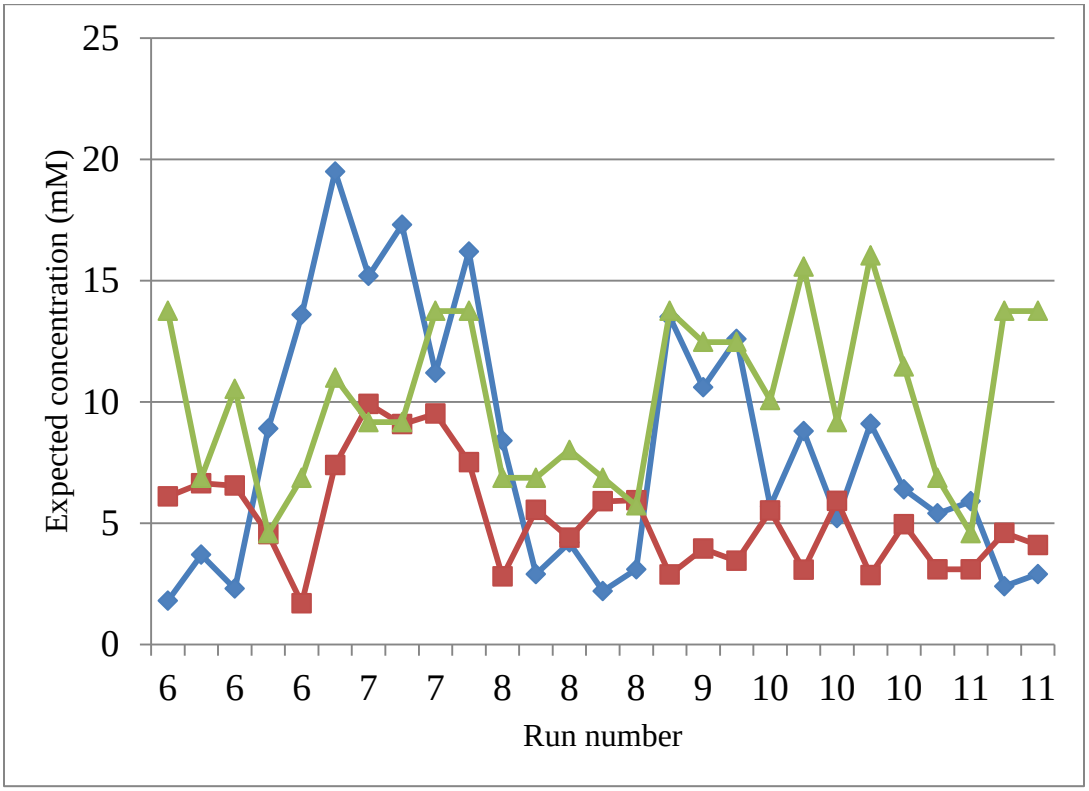


Figure 4.2 Expected concentration profile for the calibration data in DS 2 (runs 6-11). Each solution contained a constant amount of 10 mM urea. Blue, red, and green traces correspond to glucose, lactate, and triacetin, respectively.

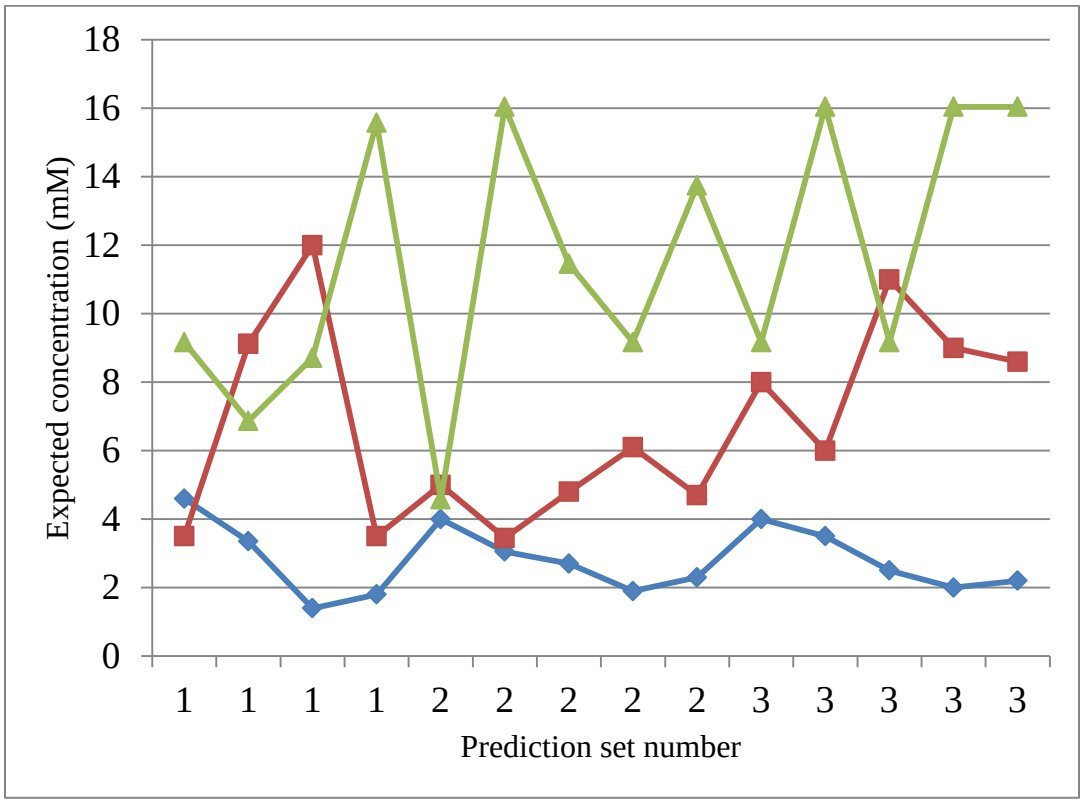


Figure 4.3 Expected concentration profile for the three prediction sets. Each solution contained a constant amount of 10 mM urea. Blue, red, and green traces correspond to glucose, lactate, and triacetin, respectively.

The solution exiting the mixer was flowed through a 20 mm-diameter circular aperture transmission cell (Model 118-3, Wilmad Glass, Buena, NJ). The sample cell employed sapphire windows (Meller Optics, Providence, RI) and was configured with a path length of 1.26 mm. The transmission cell was placed in the sample compartment of a Nicolet 6700 FT spectrometer (Nicolet Analytical Instruments, Madison, WI). The spectrometer employed a tungsten-halogen source, CaF_2 beam splitter, and a liquid-nitrogen-cooled InSb detector. A K-band optical interference filter (Barr Associates, Westford, MA) was placed before the sample to isolate the region of 5000-4000 cm^{-1}. To ensure detector linearity, an aperture setting of 100 was used and the source was further attenuated by placing a nominal 63 % neutral density filter (Rolyn Optics, Covina, CA) before the sample.

The temperature of the samples exiting the sample cell was monitored with a copper-constantan thermocouple probe (Omega Engineering Inc., Stamford, CT) inserted into a port in the vinyl tubing. An Omega Model 670 digital meter recorded the temperatures with a precision of ± 0.1 °C. For the entire study, the temperature range of the flowing liquid was maintained in a range of 36.6-37.2 °C.

After the sample exited the sample cell and passed through additional insulated vinyl tubing (89.2 mm in length), fractions were continuously collected at a rate of 1 min/tube using a Gilson FC 203B fraction collector (Gilson, Inc., Middleton, WI). The glucose concentrations of each of the fractions were verified each day with a YSI Model 2300 STAT PLUS glucose-lactate analyzer (YSI Inc., Yellow Springs, OH) which had an estimated instrumental error of ± 0.2 mM according to the YSI product specifications. The experimental setup used is shown in Figure 4.4.

The software used for the data collection and subsequent Fourier processing was Omnic (Version 7.1, Nicolet Analytical Instruments) operating on a Dell OptiPlex GX280 computer (Dell Computer Corp., Austin, TX) running under Windows 7 (Microsoft, Inc., Redmond, WA). Spectra for the liquid flowing through the sample cell were collected continuously as 64 co-added (~ 1 min) asymmetric scans consisting of 4097 points. The Fourier processing steps included one level of zero filling, Happ-Genzel apodization, and Mertz phase correction. The computed spectra had a point spacing of 1.93 cm^{-1}.This corresponded to 519 resolution elements over the range of 4000-5000 cm^{-1}.

The software recorded the time at the end of the collection of each spectrum. There was a time delay between the spectra collected in the sample cell and the fractions collected by the fraction collector for the corresponding solution. The starting time for the spectral collection, starting time for the fraction collector and the time for the solution to flow through the tubing from the sample cell to the fraction collector were used to assign a glucose concentration value for each of the collected spectra. Spectra collected while the pump speed was changing have partially equilibrated glucose concentrations and were omitted from the data analysis.

Concentrations corresponding to changes in pump speed were identified on the basis of a 95% confidence interval calculation for a particular concentration level. Specifically, the 95 % confidence interval was calculated for the concentrations measured for a particular pump speed and the concentrations above or below the confidence limits were judged to correspond to samples with insufficient equilibration. For these samples, there was no guarantee that the glucose concentration determined with the YSI analyzer

accurately reflected the concentration of the solution flowing through the sample cell when the spectrum was acquired.

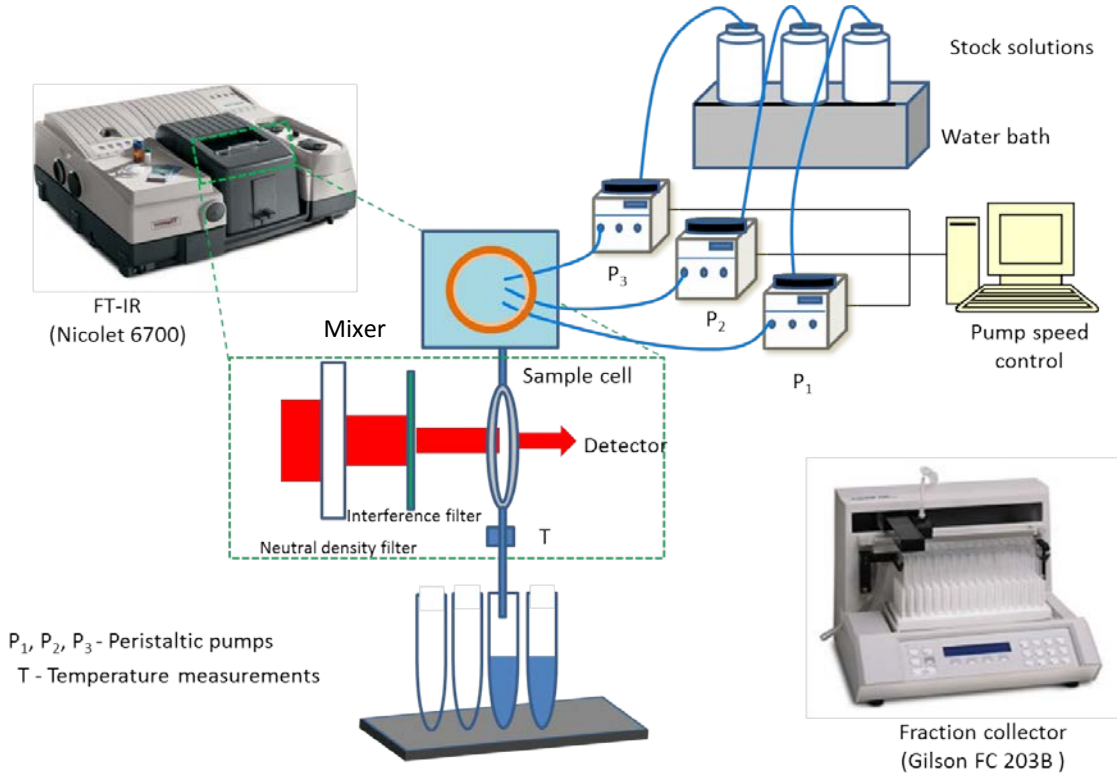


Figure 4.4 Schematic of the experimental setup. Three peristaltic pumps (P_1, P_2, P_3) and an in-line mixer were used to produce varying concentrations of glucose samples from the stock solutions. Samples flowed into the sample cell contained in the spectrometer and then were collected by the fraction collector.

The complete glucose concentration profile and the corresponding refined profile are given in Figures 4.5 A and 4.5 B, respectively. These profiles correspond to the concatenation of runs 1-14. The horizontal line at 3.0 mM in both figure panels indicates the concentration used in this work to define hypoglycemia. Superimposed on Figure 4.5 B are designations that denote the "training set" of data used for the calibration of the alarm algorithm, the "monitoring set" of data used to optimize the calibration, and the three "prediction sets" used to simulate the operation of the alarm with unknown data. The partitioning of the data into these subsets will be discussed in a subsequent section.

After collection and Fourier processing, spectra were transferred from the computer controlling the spectrometer to a Dell Precision 670 workstation (Dell Computer Corp.) running under Red Hat Linux (Version 5.3, Red Hat, Inc., Raleigh, NC). All subsequent calculations were performed on this computer using the Matlab development environment (Version 7.4.0 (R2007a), The MathWorks, Inc., Natick, MA). Software for the calculation of piecewise linear discriminants used in-house software written in Fortran and compiled with the Intel Fortran Compiler for Linux (Version 10.0, Intel Corp., Santa Clara, CA).

Results and Discussion

Spectral Noise Levels

Data for DS 1 were acquired over five days and consisted of 355 total single-beam spectra. The refined glucose concentration profile consisted of 325 spectra. Data for

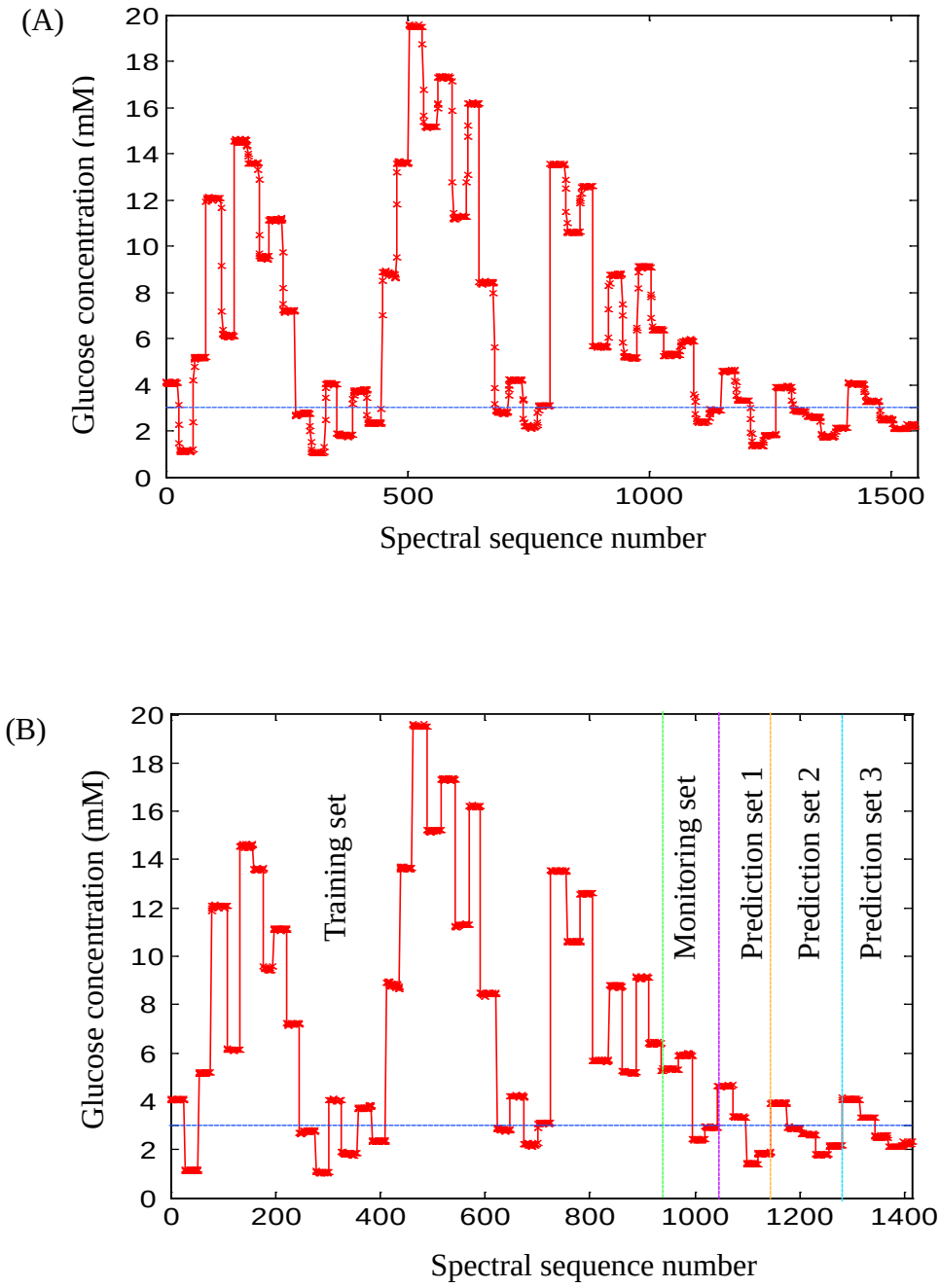


Figure 4.5 Glucose concentration profiles for the study. Panels A and B correspond, respectively, to the complete profile and the refined profile obtained after the removal of samples judged to be insufficiently equilibrated during the spectral acquisition. The labels in panel B denote the subdivision of the data into groups for calibration, calibration testing (monitoring), and external prediction. Horizontal lines in both panels denote the hypoglycemic alarm concentration of 3.0 mM used in this work.

DS 2 were collected over nine days and consisted of 1195 single-beam spectra. After omitting spectra corresponding to changes in pump speed, 1088 single-beam spectra remained.

All the refined spectra for DS 1 and the first six days of DS 2 were used to define the calibration database (calibration set) while the data for the last three days of DS 2 were used as prediction sets for testing the calibrated alarm algorithm. For each concentration level, short-term noise was evaluated by computing 100 % lines from each pair of consecutive spectra. These 100 % lines were converted to AU, and the wavenumber region of 4300-4500 cm^{-1} was fitted to a third-order polynomial model. The RMS noise was then computed about the polynomial fit to obtain the intrinsic measurement noise. The RMS noise calculated for the spectra collected over runs 1-14 at 64 co-added scans ranged from 0.71-5.4 μAU. Figure 4.6 plots the average RMS noise for each run, with error bars drawn as the average plus one standard deviation.

Assembly of Calibration Database

As described in Chapter 3, the implementation of the hypoglycemic alarm was based on the construction of differential spectra from single-beam spectra judged to have similar backgrounds. For this *in vitro* study, all spectra collected during a single run were treated as having a constant background and placed into a single data block. Thus, 11 data blocks were defined for the calibration set (runs 1-11).

Differential spectra were calculated by taking the ratios of all combinations of single-beam spectra within each block. Each ratio was oriented to produce a negative differential concentration. Those combinations that produced a differential concentration

of 0.0 mM were not used. This procedure yielded a total of 51,269 differential spectra in the calibration database. Figure 4.7 plots a histogram of the corresponding differential concentrations.

Optimization of Calibration Parameters

The PLS algorithm was used to reduce the multidimensional spectral information in the calibration database into a series of PLS scores. Two parameters that must be optimized for the implementation of PLS are the spectral region submitted to the algorithm and the number of latent variables to be computed.

For the optimization of these two parameters, the calibration set was partitioned into a training set and a monitoring set. The monitoring set is defined as a subset of the calibration set that is used as a pseudo external prediction set for the purpose of parameter optimization. For example, in the optimization of the spectral range and number of latent variables, various levels of these parameters are set, the alarm algorithm is built with the data in the calibration set, and the data in the monitoring set are used to test the calibration. Through this process, the best performing calibration is used to define the optimal levels of the parameters under investigation. This procedure allows parameter optimization to occur while retaining the independence of the external prediction data that are ultimately used to evaluate the performance of the methodology. As indicated previously, the partitioning of the data into training (runs 1-10), monitoring (run 11), and prediction (runs 12-14) sets is shown by the vertical lines in Figure 4.5 B.

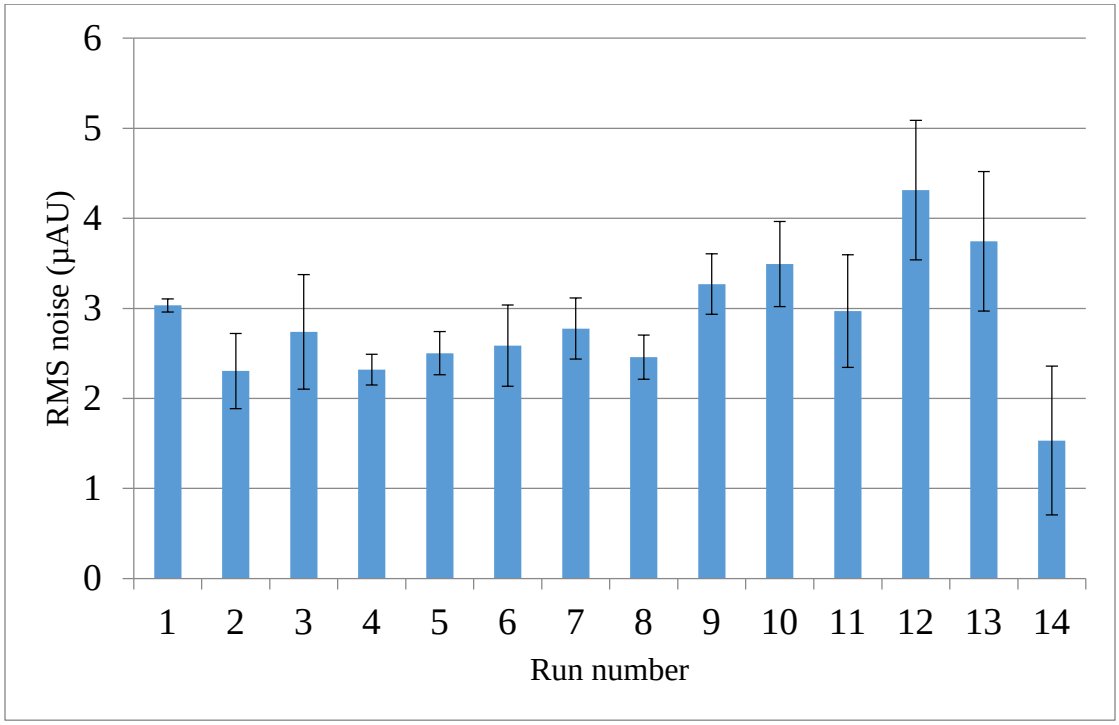


Figure 4.6 The average RMS noise (in µAU) for each run 1-14, with error bars drawn as the average plus one standard deviation.

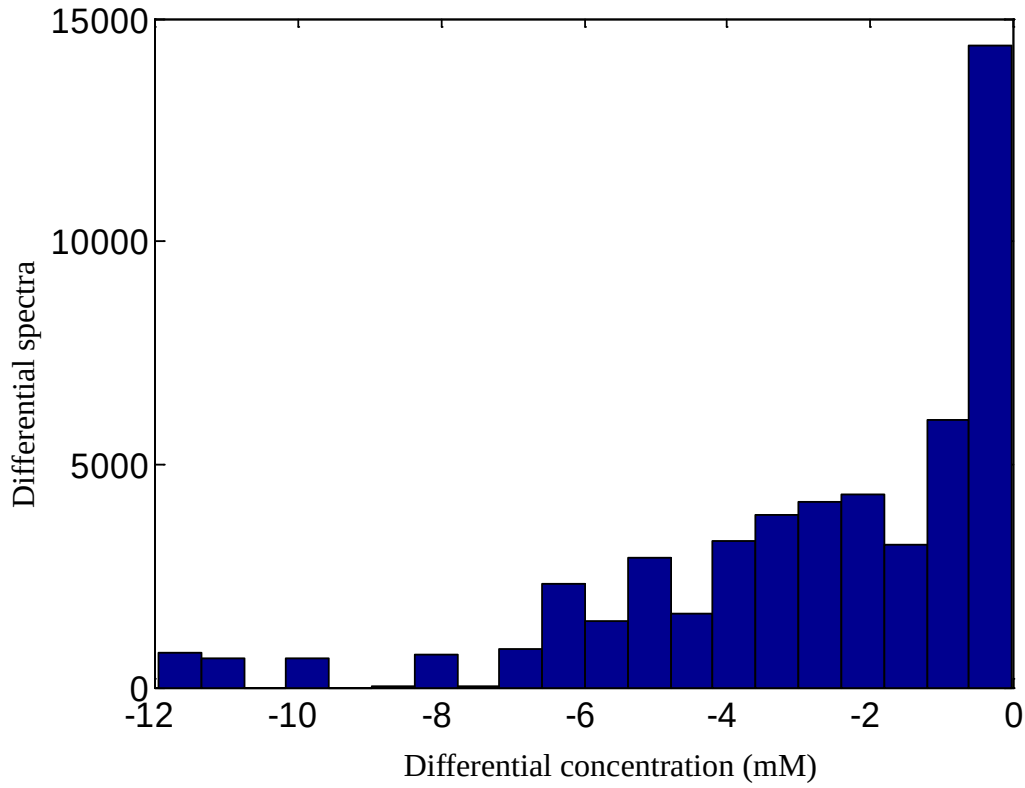


Figure 4.7 Histogram of the differential concentrations in the calibration database.

The training subset consisted of 936 single-beam spectra collected over ten different days. The calculation of differential spectra for the training set led to 45,844 spectra. The optimization of the spectral range and the number of PLS factors was performed in two steps: (1) a grid search analysis and (2) a PLDA-based optimization.

The grid search was based on sliding a window of fixed spectral width in 50 cm^{-1} increments across the 4900-4100 cm^{-1} range in the differential spectra. The starting spectral width of 100 cm^{-1} was incremented in 50 cm^{-1} increments up to 700 cm^{-1}. At each step, PLS models for differential glucose concentration were constructed using 3-16 latent variables. This produced a total of 1386 parameter combinations. As described in Chapter 3, the performance of each model was assessed by use of cross-validation. Individual cycles in the calculation involved withholding 10% of the calibration subset in contiguous blocks. Because the differential spectra were computed from all combinations of the single-beam spectra within each time block, the order of the data in the set of computed spectra did not have a time basis.

The computed SECV values were sorted, and an *F*-test was performed at the 95% level to identify the optimal number of latent variables for each spectral range. As described in Chapter 3, the optimal model size for a given spectral range was set as the number of latent variables that produced a value of SECV that was not statistically different from the minimum SECV found for that range.

Table 4.2 summarizes the four optimal wavenumber ranges and the corresponding numbers of latent variables. Because of the large number of differential spectra (i.e., large number of degrees of freedom), small differences in SECV were judged significant in the

F-test and thus the optimal model sizes tended to correspond to the models with the minimum values of SECV.

Figure 4.8 plots the values of SECV with respect to the number of latent variables for the spectral range that produced the overall lowest SECV (4650-4250 cm^{-1}). While the minimum SECV occurs at 16 latent variables, the trace is only decreasing very slowly past 11. No benefit to extending the optimization past 16 latent variables is apparent. In addition, despite the results of the *F*-test, further evaluation of model sizes less than 16 is suggested.

The grid search analysis was used to identify promising spectral ranges and corresponding numbers of latent variables for potential use in building classification models with PLDA. The performance of the monitoring set with PLDA was tested with the top four spectral ranges found through the grid search. For each range tested, the number of latent variables was varied from 6 to 11. This selection was made on the basis of plots such as Figure 4.8 that suggest little improvement in modeling performance is obtained past 11 latent variables.

As described in Chapter 3, the implementation of the hypoglycemic alarm is based on the development of a classification model that assigns spectra to alarm and non-alarm categories. Training the classification model requires the calculation of a critical concentration, C_{crit}, defined previously in Eq. 3.37.

In the real application in which nocturnal hypoglycemia is monitored, a reference spectrum would be collected and corresponding reference concentration measured at the start of the sleep period. To simulate this scenario, the first spectrum in the monitoring set was taken as the reference spectrum, and the corresponding glucose concentration (5.3

mM) was used as C_{ref} in Eq. 3.37. Thus, using an alarm concentration of 3.0 mM and according to Eq. 3.37, $C_{crit} = 3.0 – 5.3 = -2.3$ mM. The remaining 105 single-beam spectra in the monitoring set were used to compute differential spectra by taking the ratio to the reference spectrum.

The PLS loading weights and spectral loadings previously computed from the calibration data were then used to compute the scores that defined the pattern vectors corresponding to each differential spectrum. For a given spectral range under consideration, the loading weights and spectral loadings computed from that range were employed in the calculation of the PLS scores.

The critical concentration was used to partition the 45,844 PLS score vectors in the calibration subset into alarm and non-alarm classes. There were 21,364 alarm patterns and 24,480 non-alarm patterns in the training set. As described in Chapter 3, for each combination of spectral range and number of latent variables, three replicate piecewise linear discriminants were computed on the basis of using the training set in conjunction with three sets of training parameters. Each replicate classifier was based on a single discriminant function (**w** in Eq. 3.31). For these data, no benefit was realized from constructing the classification boundary with multiple linear segments.

As discussed previously in Chapter 3, the PLDA method operates on the basis of computing one or more linear discriminant functions, each of which separates some portion of the training data into a “pure” group of alarm patterns on one side of the discriminant boundary and a mixed group of alarm and non-alarm patterns on the other side. Thus, one measure of the discriminating ability of the patterns is the total number of alarm patterns separated by the set of discriminant functions. For each combination

Table 4.2 Results of grid search optimization of spectral range and latent variables

Spectral range (cm^{-1})	Latent variables	SECV (mM)
4650-4250	16	0.32_2
4650-4300	15	0.32_4
4700-4250	16	0.32_4
4700-4300	15	0.32_5

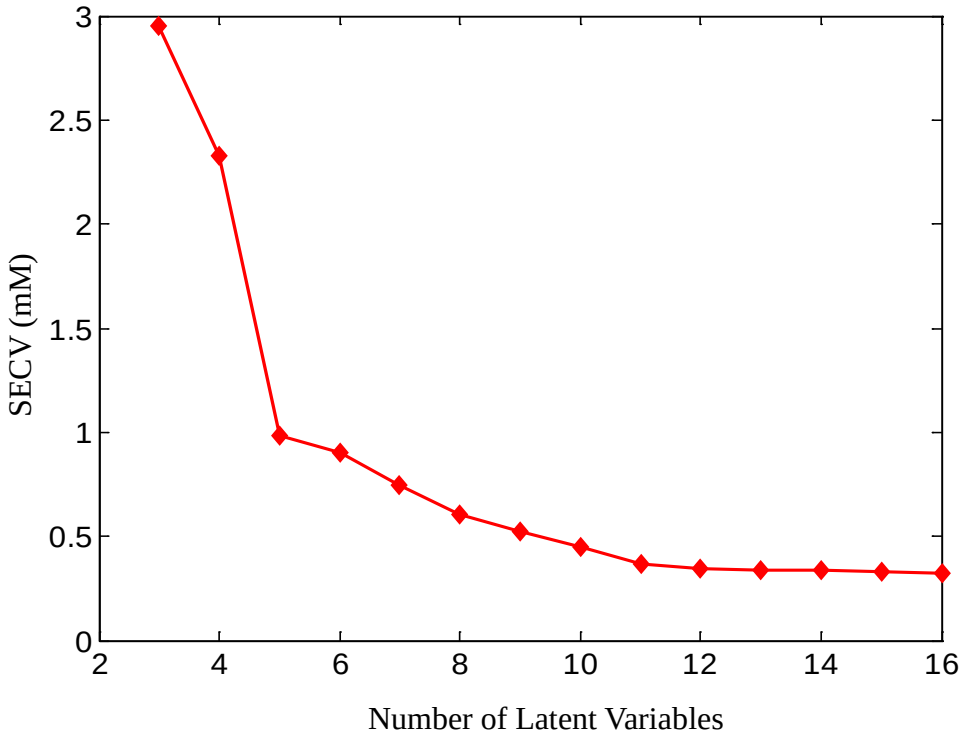


Figure 4.8 Cross-validation results (SECV) vs. the number of latent variables for the optimal wavenumber range of 4650-4250 cm^{-1}. Calibration models were based on PLS analysis of differential spectra and concentrations in the calibration subset.

spectral range and number of latent variables, Table 4.3 summarizes the percentage (average ± standard deviation) of alarm patterns separated across the three replicate classifiers.

Each replicate classifier was applied to predict the class assignment for the 48 alarm and 58 non-alarm patterns in the monitoring set. Table 4.3 further summarizes the percentage (average ± standard deviation) of missed and false alarms for each of the parameter combinations studied.

None of the combinations of spectral range and latent variables produced missed or false alarms with the monitoring set. Thus, the pattern classification results for the monitoring set do not provide clear guidance regarding an optimal choice for these parameters. This is not surprising given the similarities in modeling performance presented previously in Table 4.2.

To define a criterion for selecting a classifier for use in subsequent testing with the three prediction sets, the smallest model (i.e., the model based on the lowest dimensional patterns) that achieved an acceptable degree of separation of the training set was selected. The justification for choosing the smallest model is based on the commonly held view that a simpler model with equivalent performance often exhibits greater robustness because it depends on fewer estimated parameters.

A level of 95% separation of the training data was chosen as the criterion for acceptable performance. While somewhat arbitrary, the common use of the 95% level in statistical testing provides some degree of justification for this choice. Through the use of this criterion, the classifier based on a spectral range of 4650-4300 cm^{-1} and eight latent variables was chosen as optimal for use in subsequent testing.

The results for this parameter combination are shown in red in Table 4.3. This spectral range is logical as it encompasses the glucose C-H combination band at 4400 cm^{-1}. The glucose concentration profile for the monitoring set is given in Figure 4.9. A plot of discriminant scores produced by the optimal classifier when applied to the monitoring data is displayed in Figure 4.10.

The main assumptions in the pattern classification based on PLDA are that the alarm patterns are clustered differently in the data space compared to those of the non-alarm class and that a separating boundary can be constructed to discriminate these data clusters. To gain greater insight into the degree of clustering in the data space, the first three PLS scores calculated from the 4650 - 4300 cm^{-1} range in the monitoring spectra were plotted together. Figure 4.11 clearly demonstrates the validity of the above mentioned assumptions. A clear separation in the alarm and non-alarm PLS patterns are observed for the monitoring set.

A check of the validity of using differential spectra in the alarm algorithm implementation is to study the relationship between the discriminant scores and the differential concentrations. A plot of discriminant scores with respect to differential concentrations is shown in Figure 4.12. The discriminant scores were produced by the classifier based on the 4650 - 4300 cm^{-1} range and eight latent variables. A clear dependence of the discriminant scores on the differential concentrations is observed for the monitoring set.

Finally, an examination of the PLS spectral loadings and loading weights computed from the calibration data was conducted to verify the presence of non-noise features. The sixth, seventh and eighth loading weights and the corresponding spectral

loadings computed from the differential spectra in the calibration database are shown in Figures 4.13 and 4.14, respectively. These plots show spectral features rather than random noise. This provides further assurance that the use of eight PLS factors is reasonable.

Table 4.3 Average percentages of missed and false alarms for the monitoring set

Spectral range(cm^{-1})		Number of Latent Variables					
		6	7	8	9	10	11
4650-4250	A.M (%)[a] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.F (%)[b] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	D.S (%)[c] ± S.D	82.5 ± 0.0	85.9 ± 6.2	95.1 ± 1.4	97.7 ± 1.4	98.1 ± 0.7	99.7 ± 0.0
4700-4250	A.M (%)[a] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.F (%)[b] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	D.S (%)[c] ± S.D	75.5 ± 0.9	89.0 ± 0.1	88.2 ± 3.4	98.4 ± 0.3	97.9 ± 1.3	98.4 ± 0.9
4650-4300	A.M (%)[a] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.F (%)[b] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	D.S (%)[c] ± S.D	85.4 ± 0.6	94.8 ± 0.6	96.0 ± 0.0	99.0 ± 0.1	99.7 ± 0.1	99.8 ± 0.0
4700-4300	A.M (%)[a] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.F (%)[b] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	D.S (%)[c] ± S.D	83.2 ± 0.3	88.5 ± 0.1	93.8 ± 0.0	93.0 ± 1.7	98.7 ± 0.6	99.6 ± 0.2

[a]Average percentage of missed alarms (AM) ± standard deviation.

[b]Average percentage of false alarms (AF) ± standard deviation.

[c]Average percentage of separated alarm patterns with a single discriminant (DS) ± standard deviation.

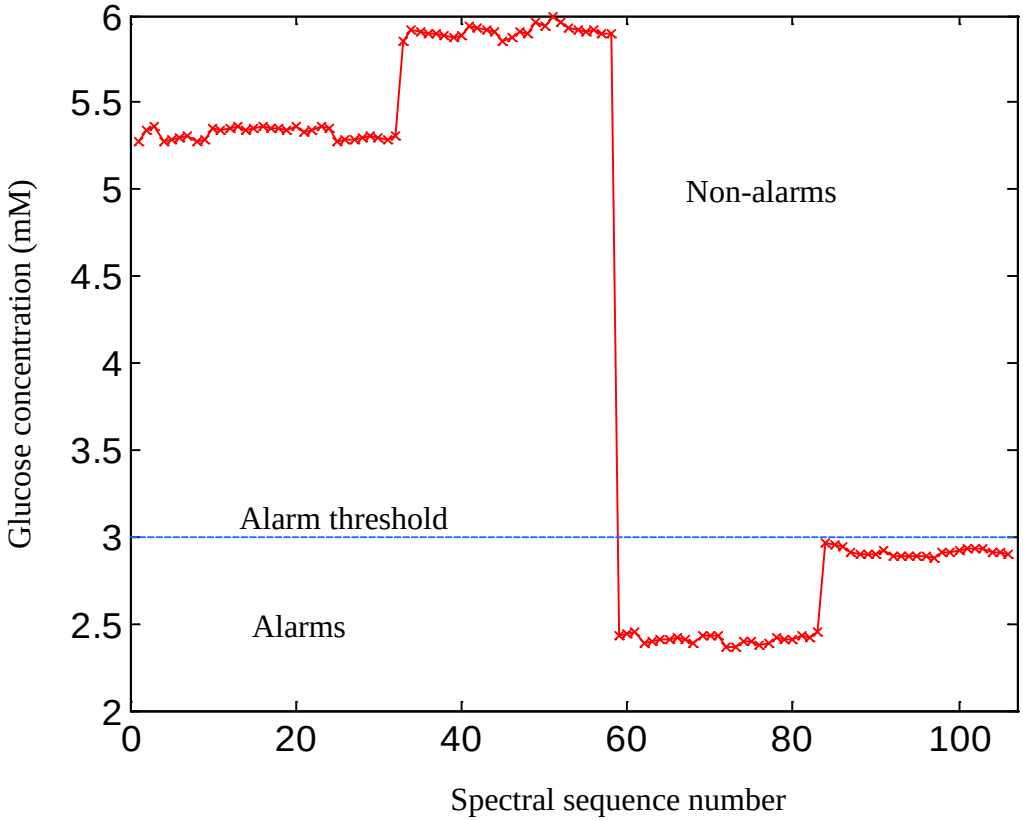


Figure 4.9 Glucose concentration profile for the monitoring set. The alarm threshold value of 3.0 mM is indicated by the horizontal line. There were 48 and 58 spectra in the alarm and non-alarm data classes, respectively.

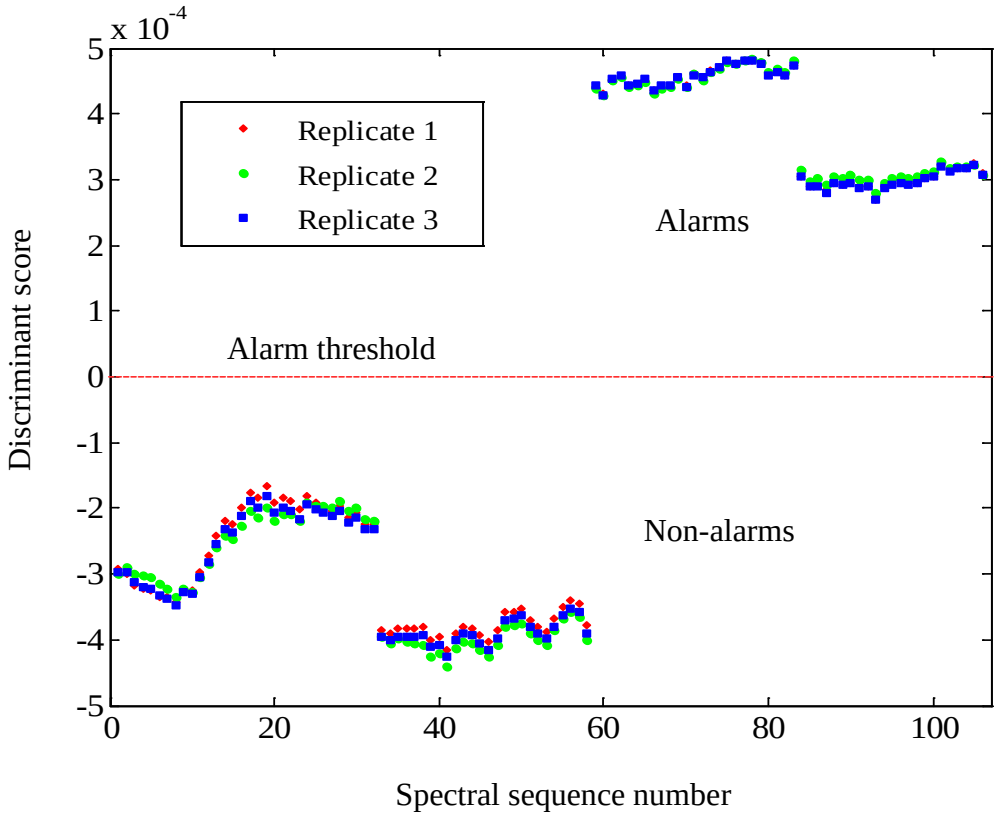


Figure 4.10 Plot of discriminant scores for the monitoring set produced by the classifier based on a spectral range of 4650 - 4300 cm^{-1} and eight PLS latent variables. Symbols shown in red, green, and blue correspond to discriminant scores produced by the three replicate classifiers.

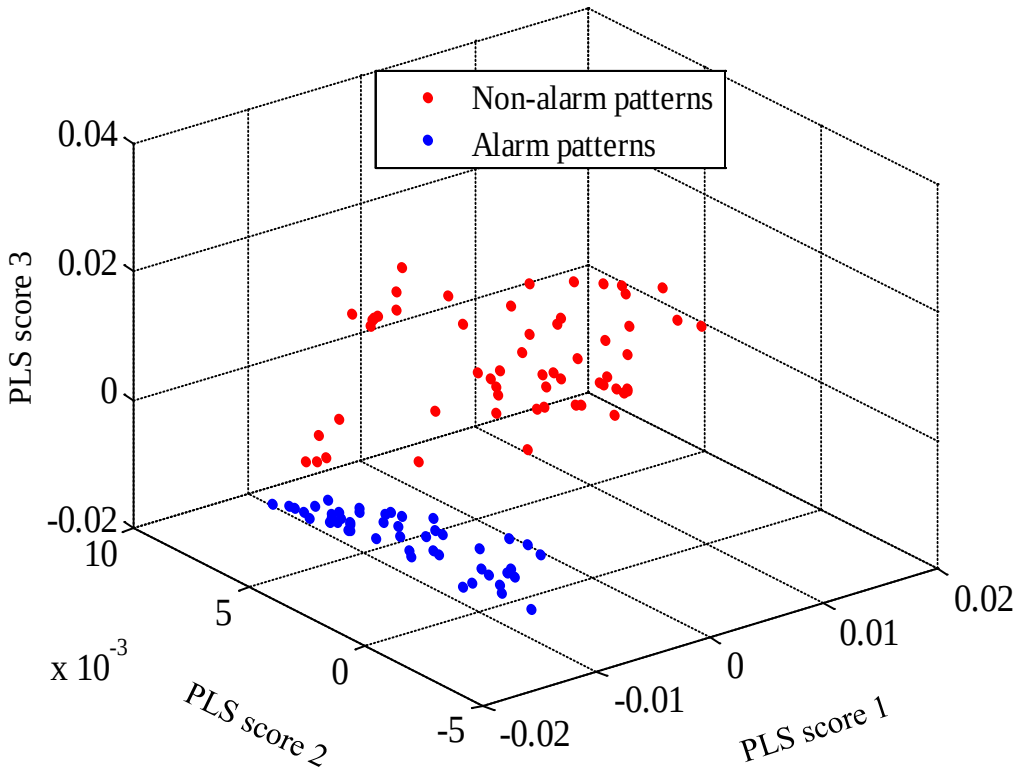


Figure 4.11 First, second, and third PLS scores for the monitoring set based on the spectral range of 4650 – 4300 cm^{-1}. Clear clustering between the alarm and non-alarm data classes is observed.

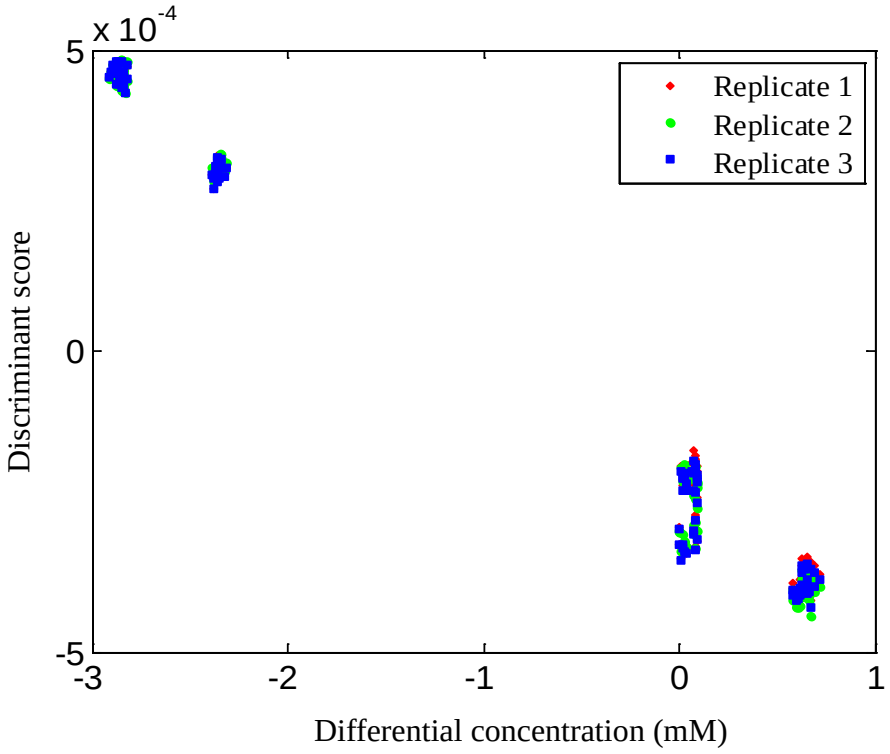


Figure 4.12 Relationship between discriminant scores and differential concentrations for the monitoring set. The classifier based on a spectral range of 4650 – 4300 cm^{-1} and eight PLS latent variables was employed. Red, green, and blue symbols denote the discriminant scores for the three replicate classifiers.

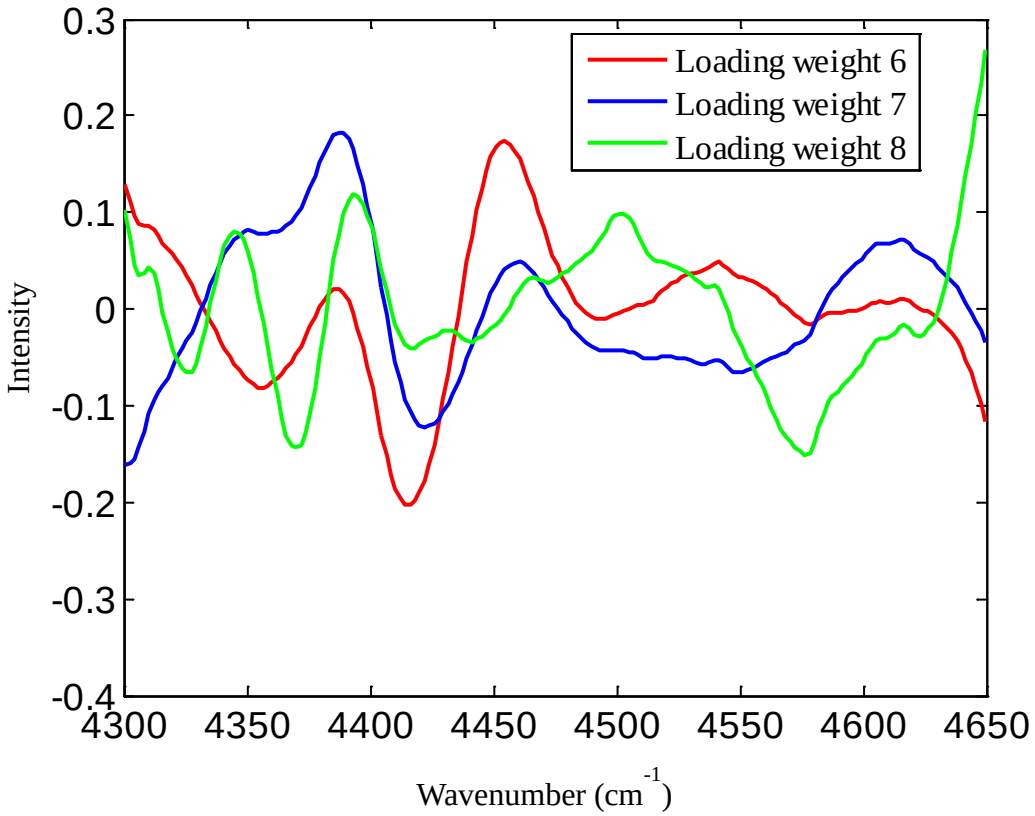


Figure 4.13 Loading weights computed from the differential spectra in the calibration database. The loading weights corresponding to the sixth, seventh, eighth PLS latent variables are shown. The features appear to derive from spectral bandshapes rather than random noise.

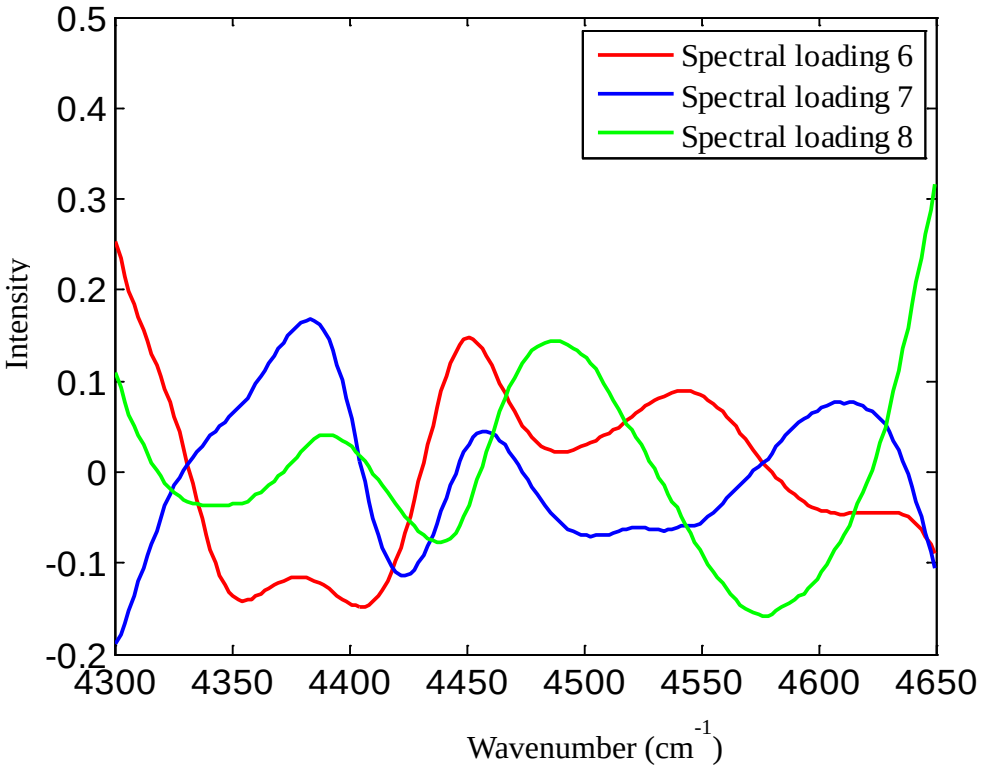


Figure 4.14 Spectral loadings computed from the differential spectra in the calibration database. The spectral loadings corresponding to the sixth, seventh, eighth PLS latent variables are shown. The features appear to derive from spectral bandshapes rather than random noise.

Classification Performance with Prediction Sets

Classifiers were next developed to test the implementation of the alarm algorithm with the three prediction sets. The parameter settings of 4650-4300 cm^{-1} and eight latent variables selected from the work with the monitoring data were again employed. The full set of calibration data based on runs 1-11 was used to define the calibration database for the development of the classifiers.

Data blocks were again defined on the basis of each run, and differential spectra were computed from all combinations of single-beam spectra within each block. Spectra corresponding to differential concentrations of 0.0 mM were again discarded. A total of 51,269 differential spectra and corresponding differential concentrations were computed through this procedure. Eight PLS factors were calculated over the 4650-4300 cm^{-1} range, and the differential spectra in the calibration set were reduced to eight-dimensional score vectors.

The reference concentration for prediction set 1 was 4.6 mM and the critical concentration was -1.6 mM. Differential spectra were generated relative to the reference spectrum and the PLS factors previously computed with the calibration data were used to compute the corresponding score vectors. The calibration patterns were partitioned based on the critical concentration into a training set containing 28,469 alarm patterns and 22,800 non-alarm patterns. Three replicate classifiers were computed with the training set. As with the monitoring set, each classifier contained a single discriminant function. Across the three replicate classifiers, an average of 85.00 ± 0.01 % of the alarm patterns in the calibration set were separated. Discriminant scores were then computed for the differential spectra in prediction set 1.

Figure 4.15 shows the glucose concentration profile for prediction set 1 in which there were 47 alarm and 55 non-alarm patterns. A pattern was judged to be in the alarm class if two of the three replicate classifiers produced positive discriminant scores. Patterns failing to meet this criterion were assigned to the non-alarm class. There were no missed or false alarms for the three replicate classifiers employed. The resulting discriminant score plot for prediction set 1 is shown in Figure 4.16.

In Figure 4.17, the fourth, fifth and sixth PLS scores are plotted together to illustrate the data space formed by the alarm and non-alarm patterns. These scores were selected because they provided the best visual distinction between the data classes. Clear discrimination is noted between the alarm and non-alarm patterns. Figure 4.18 plots discriminant scores with respect to differential glucose concentrations for prediction set 1. A clear dependence of the discriminant scores on the differential concentrations is observed. Increased variation near a differential concentration of 0.0 mM illustrates the effect on the discriminant scores when there is no glucose signal.

For a successful data classification, the space explained by the prediction PLS scores should be within that of the corresponding calibration scores. The first three PLS scores for the calibration set and prediction set 1 are plotted together in Figure 4.19. It can be observed that the calibration and prediction patterns occupy the same space.

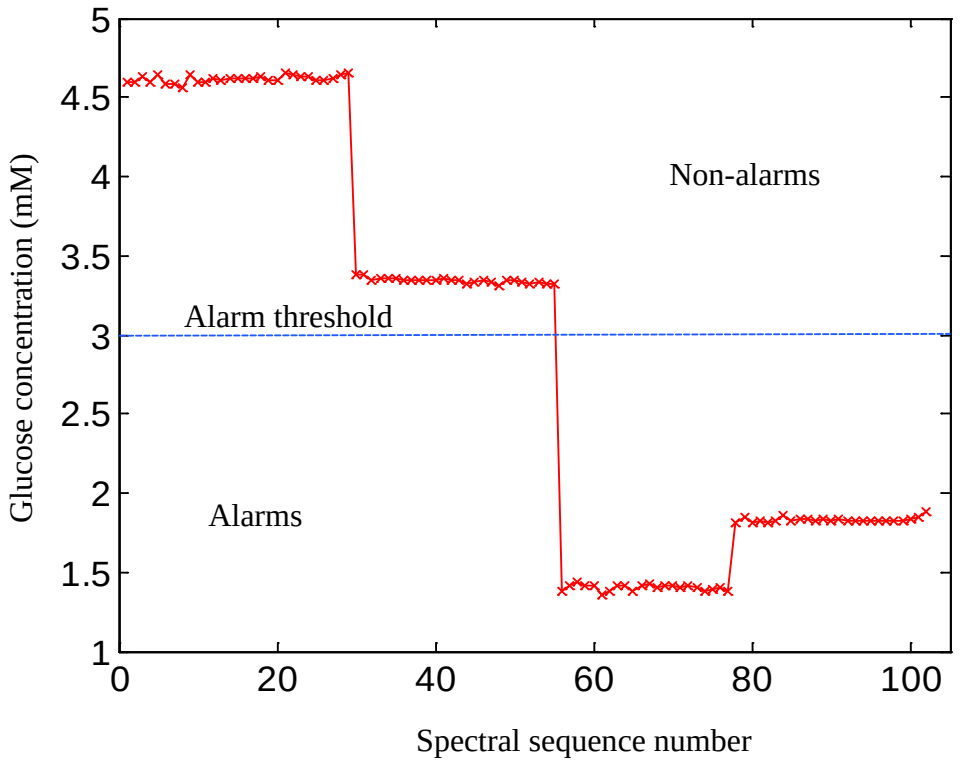


Figure 4.15 Glucose concentration profile for prediction set 1. The alarm threshold value was 3.0 mM, and there were 47 and 55 alarm and non-alarm patterns, respectively.

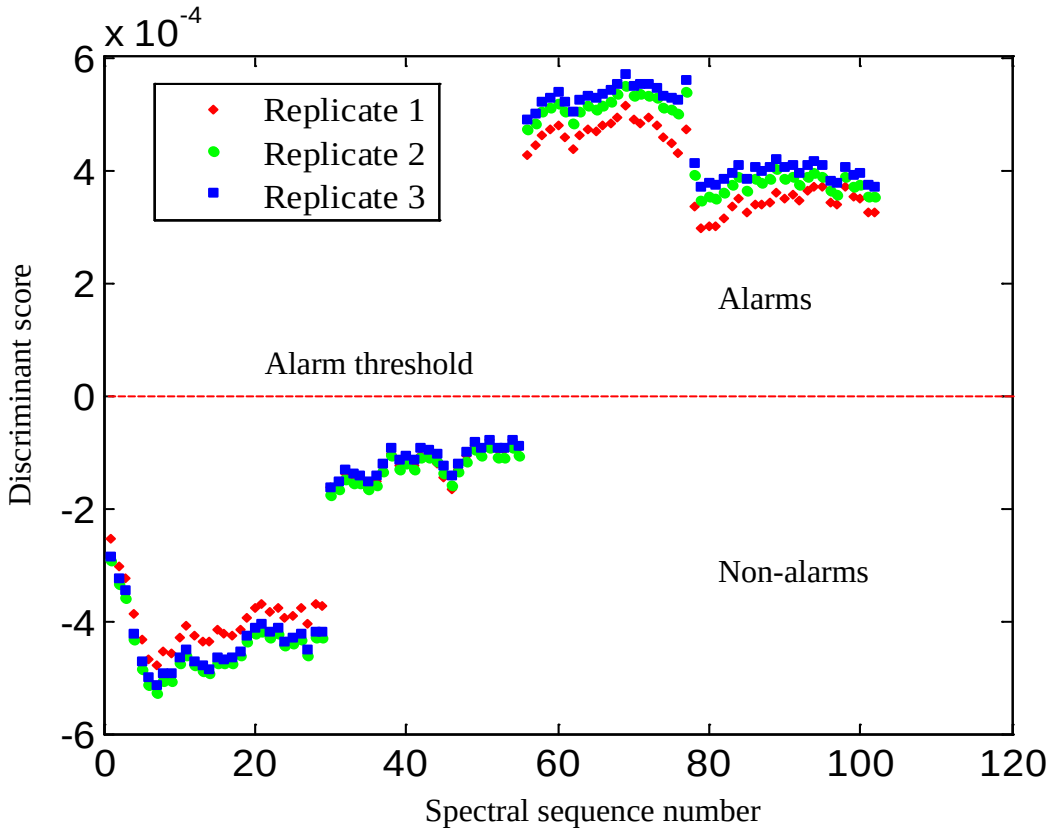


Figure 4.16 Discriminant score plot for prediction set 1. The red, green, and blue symbols denote the discriminant scores produced by the three replicate classifiers. All patterns were correctly classified.

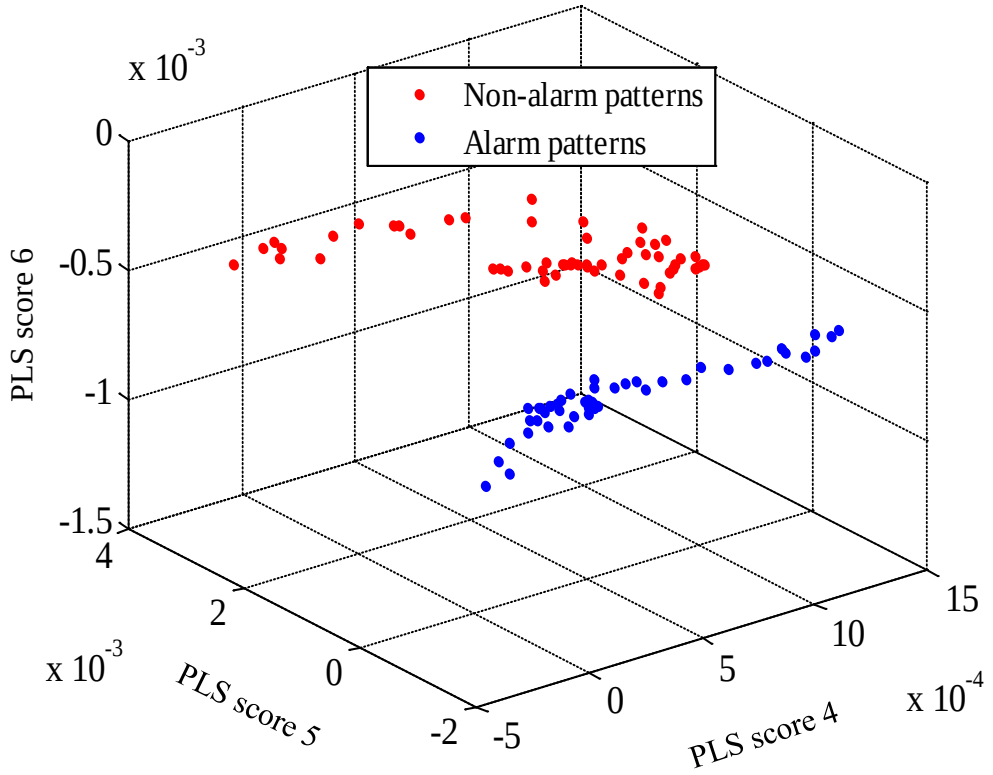


Figure 4.17 Plot of the fourth, fifth, sixth PLS scores for the patterns in prediction set 1. Blue and red symbols denote alarm and non-alarm patterns, respectively.

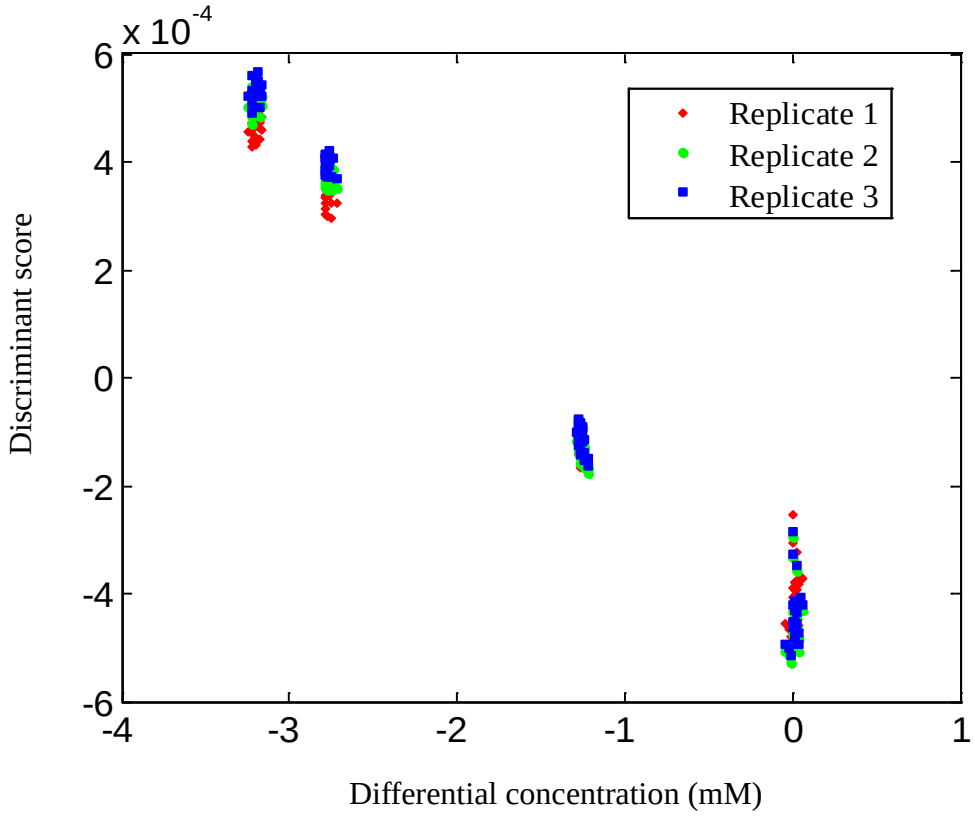


Figure 4.18. Plot of discriminant scores with respect to corresponding differential glucose concentrations for prediction set 1. Red, green, and blue symbols denote the discriminant scores produced by the three replicate classifiers. Discriminant scores greater than zero correspond to patterns placed in the alarm class.

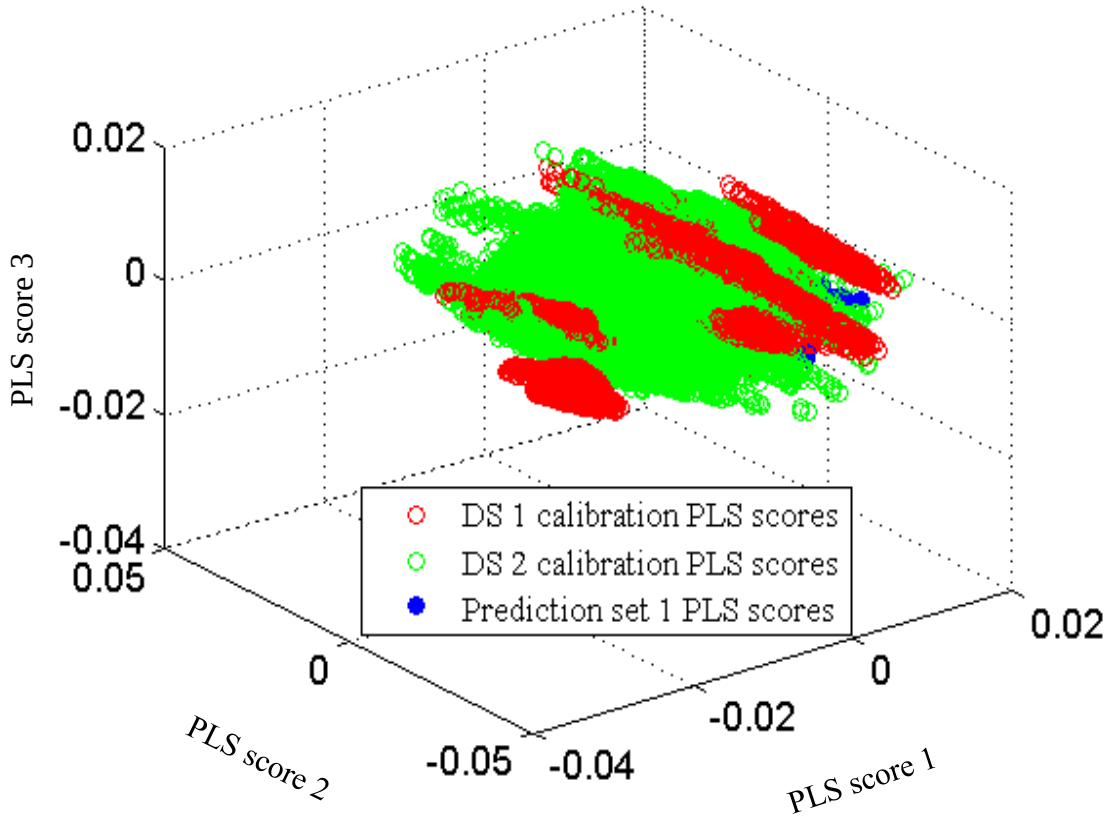


Figure 4.19 First three PLS scores plotted together for the calibration data set (DS 1 red, DS 2 green) and prediction set 1 (blue). The calibration and prediction data occupy the same space, indicating that the calibration data are effective in characterizing the prediction set.

The same procedures were used for prediction set 2. The reference glucose concentration was 3.9 mM, and the corresponding critical concentration was -0.9 mM. The data partitioning based on this critical concentration resulted in 35,062 alarm patterns

and 16,207 non-alarm patterns. Approximately 80.00 ± 0.03 % of the alarm patterns of the calibration set were separated with a single discriminant function. Figure 4.20 shows the glucose concentration profile for prediction set 2 in which there were 31 non-alarm patterns and 105 alarm patterns. The resulting discriminant score plot is shown in Figure 4.21. The second replicate discriminant was not able to classify accurately the patterns close to the alarm threshold of 3.0 mM. However, applying the decision rule of two out of three classifiers signaling an alarm, there were no missed or false alarms for this prediction set. This result underscores the importance of replicating the training of the discriminants to avoid anomalous classifications.

The first, second and third PLS scores for prediction set 2 are plotted together in Figure 4.22 to illustrate the differences in the regions of the data space occupied by the alarm and non-alarm patterns. Clear separation in observed between the data classes. A discriminant score versus differential concentration plot for prediction set 2 is shown in Figure 4.23. As observed with prediction set 1, a clear dependence of the discriminant scores on the differential glucose concentrations is noted. The first three PLS scores for the calibration set and prediction set 2 are plotted together in Figure 4.24. It can be noted that the two data groups occupy the same space, indicating good agreement between the calibration and prediction data.

Prediction set 3 was collected more than two years after the collection of the calibration data. The critical concentration for prediction set 3 was -1.1 mM, and the calibration database was partitioned into 31,575 alarm patterns and 19,694 non-alarm patterns. Approximately 77.00 ± 0.01 % of the alarm patterns of the calibration set were separated with a single discriminant function.

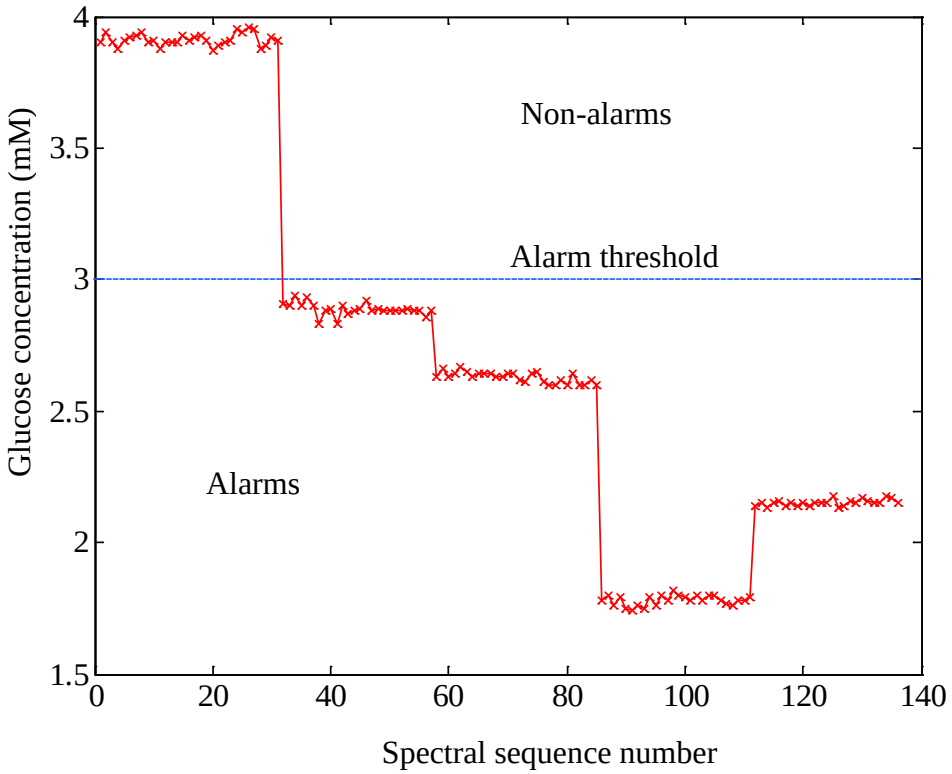


Figure 4.20 Glucose concentration profile for prediction set 2. The alarm threshold value was 3.0 mM, and there were 31 and 105 non-alarm and alarm patterns, respectively.

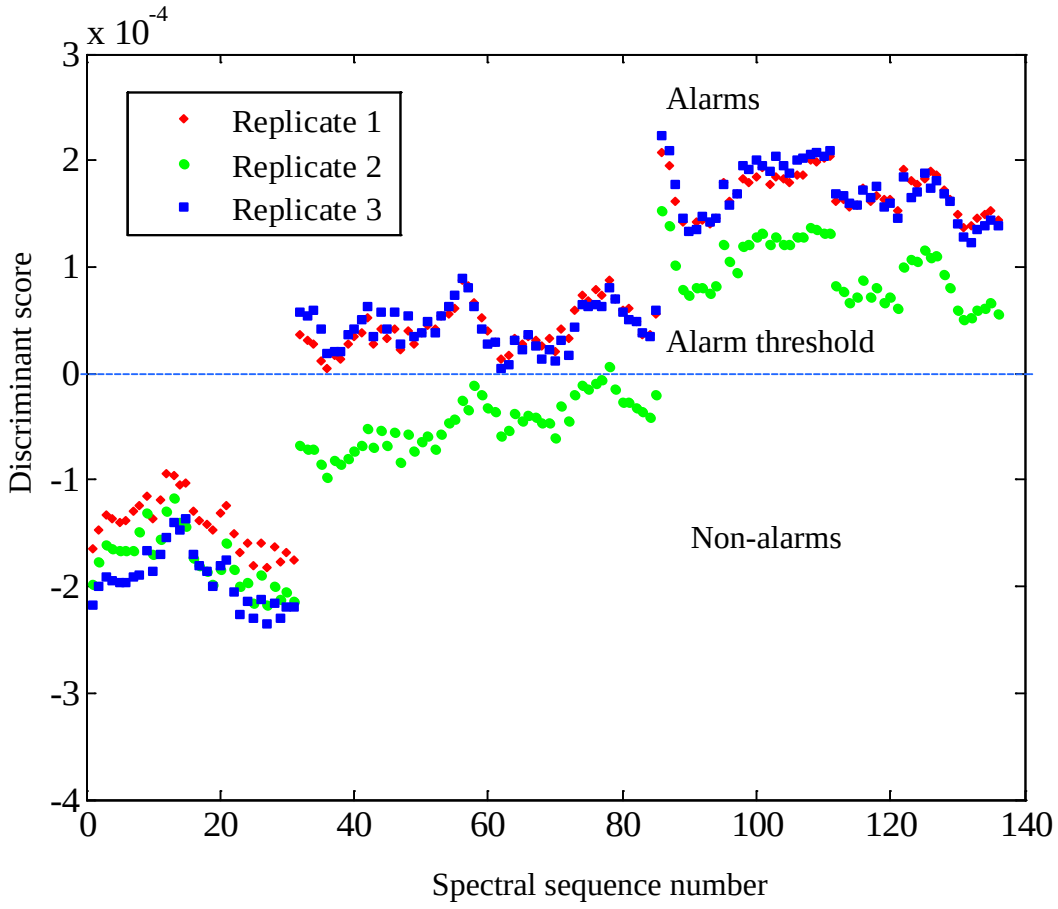


Figure 4.21 Discriminant score plot for prediction set 2. Red, green, and blue symbols denote the discriminant scores produced by the three replicate classifiers. The second replicate classifier (green) fails to recognize the alarm patterns near the 3.0 mM threshold.

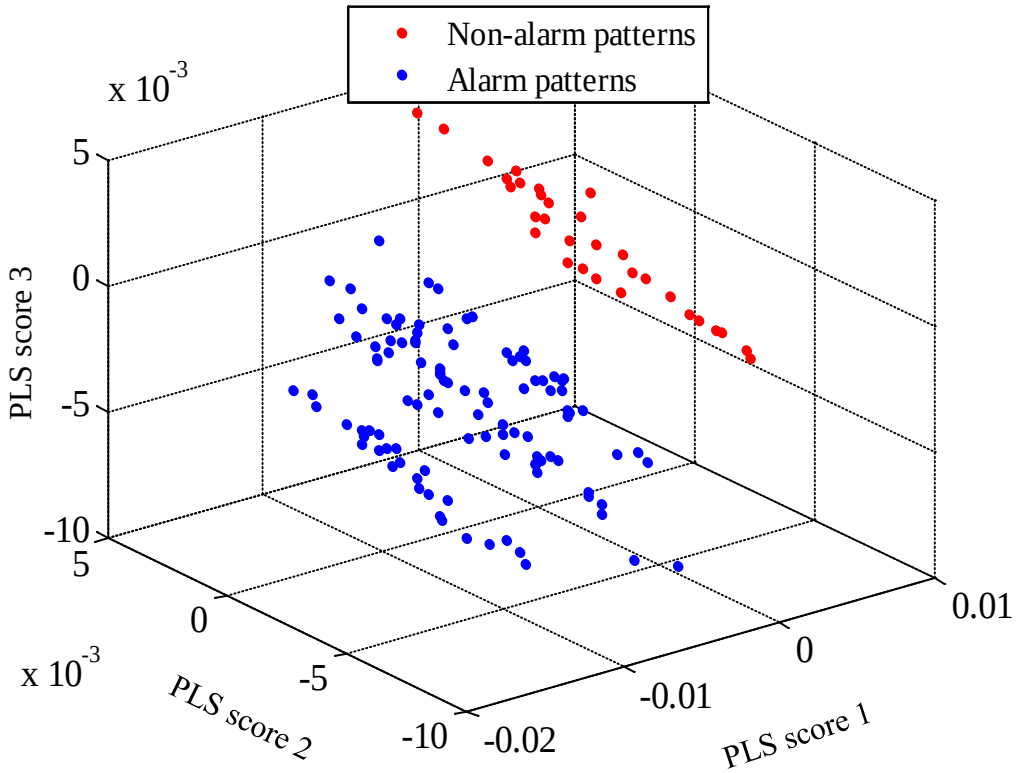


Figure 4.22 First, second, and third PLS scores for prediction set 2. Clear separation is noted between the alarm (blue) and non-alarm (red) patterns.

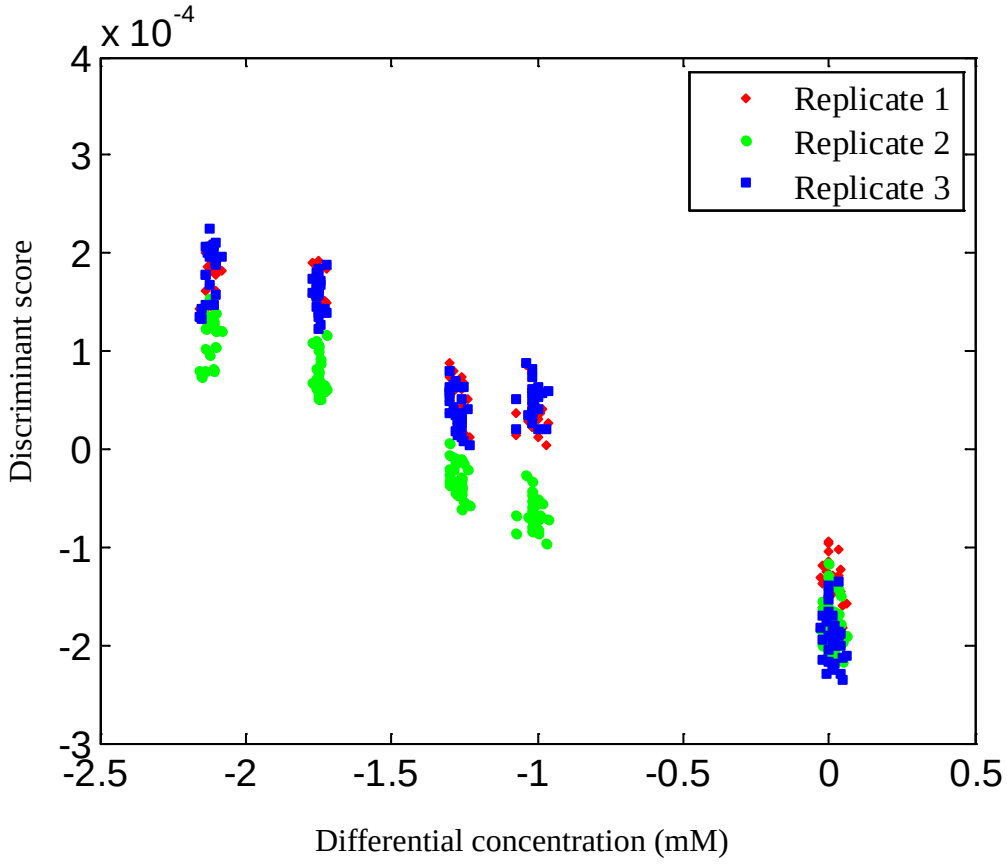


Figure 4.23 Discriminant scores plotted with respect to differential glucose concentrations for prediction set 2. Discriminant scores greater than zero correspond to patterns placed in the alarm class. The red, green, and blue symbols denote the discriminant scores produced by the three replicate classifiers. An offset of the discriminant scores from the second replicate classifier (green) can be observed. Near the critical concentration of -0.9 mM, this offset causes the classifier to fail to place the patterns correctly in the alarm class.

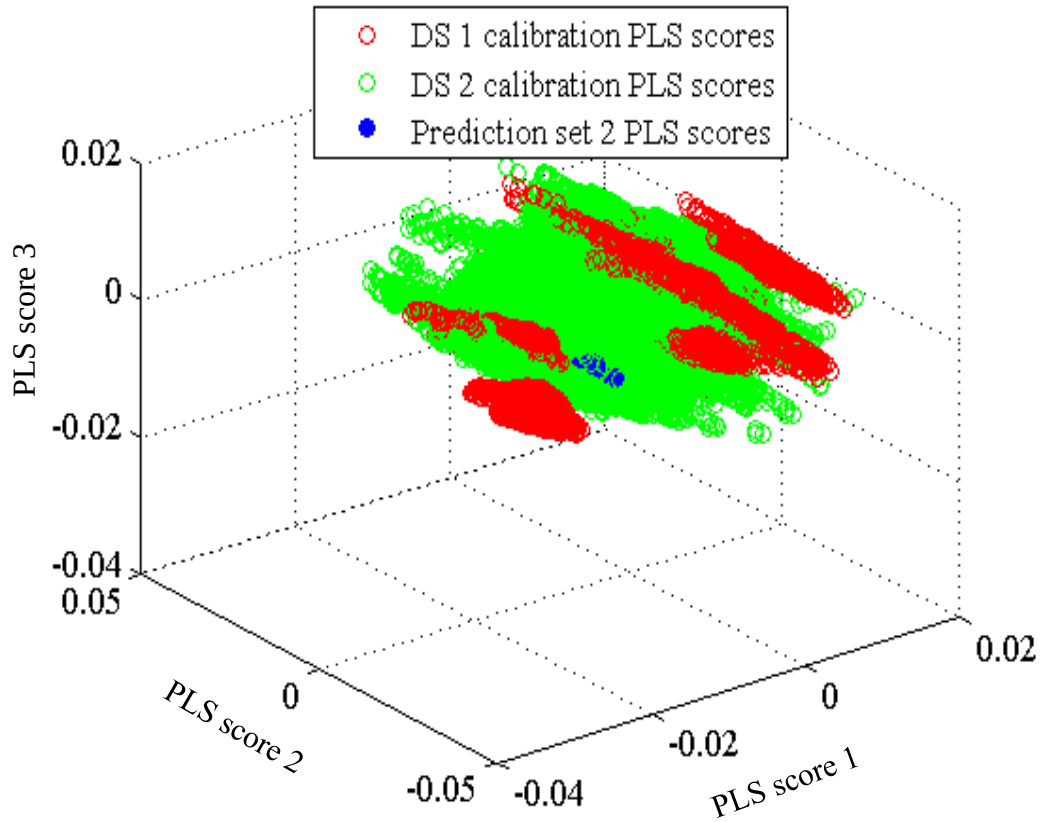


Figure 4.24 First three PLS scores plotted together for the calibration data set and prediction set 2. Red, green, and blue symbols denote the calibration data from DS 1, calibration data from DS 2, and data from prediction set 2, respectively. The two data sets are completely overlapped, indicating the calibration data provides an effective representation of the prediction data.

Figure 4.25 shows the glucose concentration profile for prediction set 3 in which there were 62 non-alarms and 71 alarm patterns. The first and the third classifiers predicted no missed or false alarms. The second classifier predicted one missed alarm and no false alarms Applying the alarm decision rule for the three replicate classifiers gave no missed or false alarms. This result is particularly impressive when the length of time between the collection of the calibration and prediction data is considered. The plot of discriminant scores is provided in Figure 4.26.

As shown in Figure 4.27, the second, third, and fourth PLS scores for prediction set 3 reveal clear separation between the alarm and non-alarm patterns. Figure 4.28 plots the discriminant scores with respect to the corresponding differential glucose concentrations. As observed with prediction sets 1 and 2, a clear relationship between the discriminant scores and differential concentrations is observed.

The first three PLS scores for the calibration set and prediction set 3 are plotted in Figure 4.29. As observed previously with the first two prediction sets, the calibration and prediction patterns cluster together in the same data space.

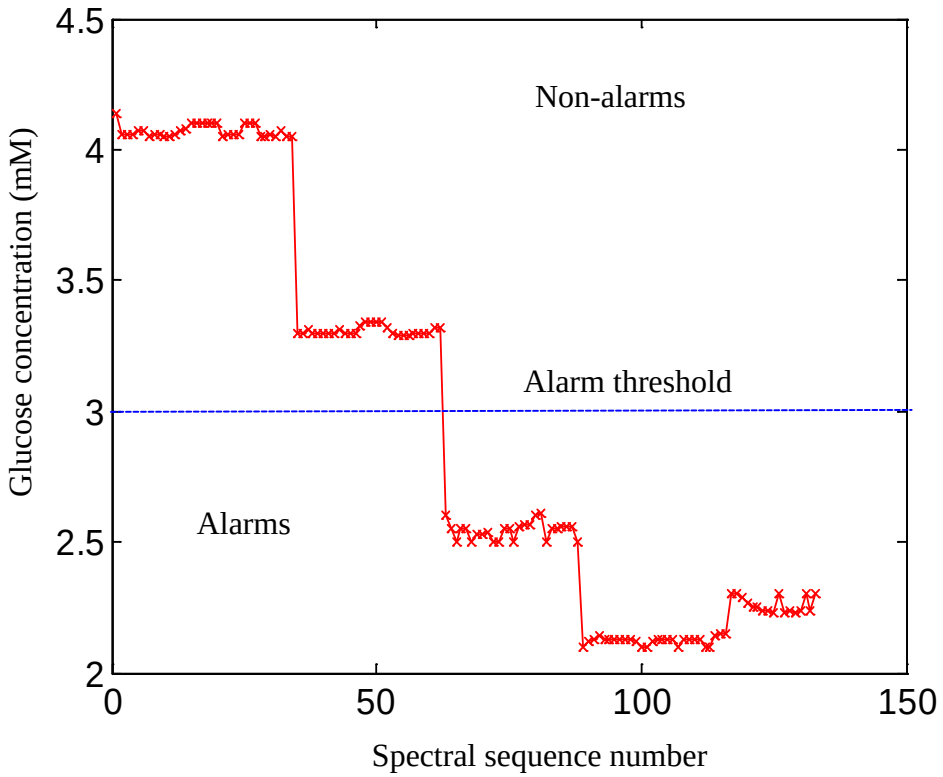


Figure 4.25 Glucose concentration profile for prediction set 3. The alarm threshold was 3.0 mM, and there were 71 and 62 alarm and non-alarm patterns, respectively.

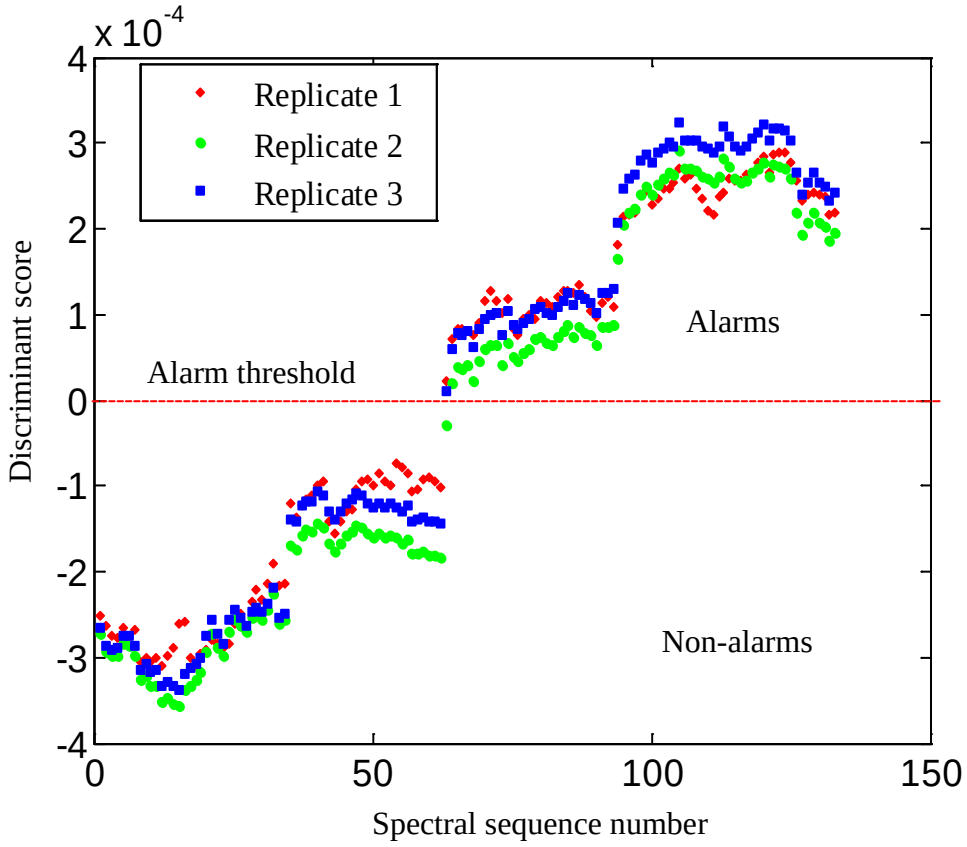


Figure 4.26 Discriminant score plot for prediction set 3. The red, green, and blue symbols denote the discriminant scores for the three replicate classifiers. There was a missed alarm in the second replicate classifier. The other classifiers had no missed or false alarms.

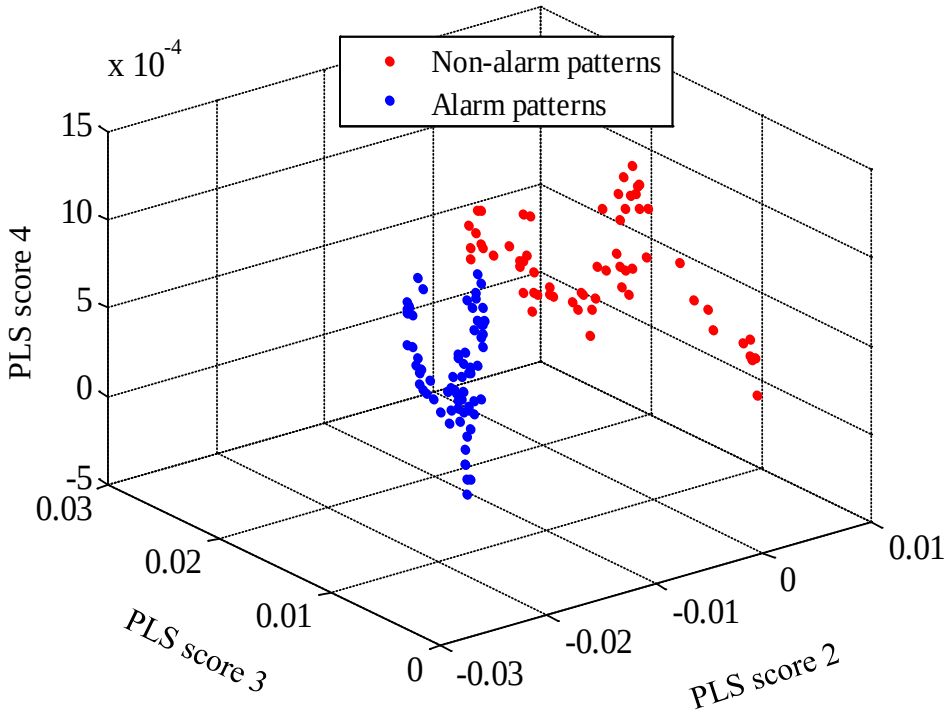


Figure 4.27 The second, third, and fourth PLS scores from prediction set 3 are plotted. Clear separation is noted between alarm (blue) and non-alarm (red) patterns.

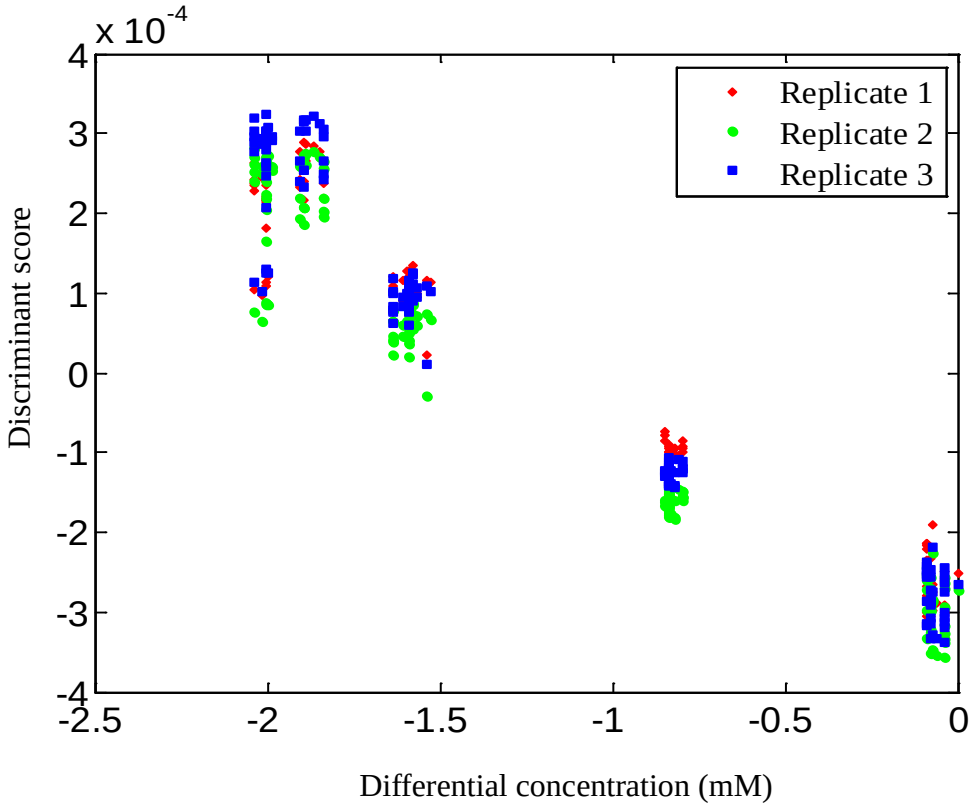


Figure 4.28 Discriminant scores are plotted with respect to differential glucose concentrations for prediction set 3. Red, green, and blue symbols denote the discriminant scores produced by the three replicate classifiers. A clear relationship between discriminant scores and differential concentrations is noted.

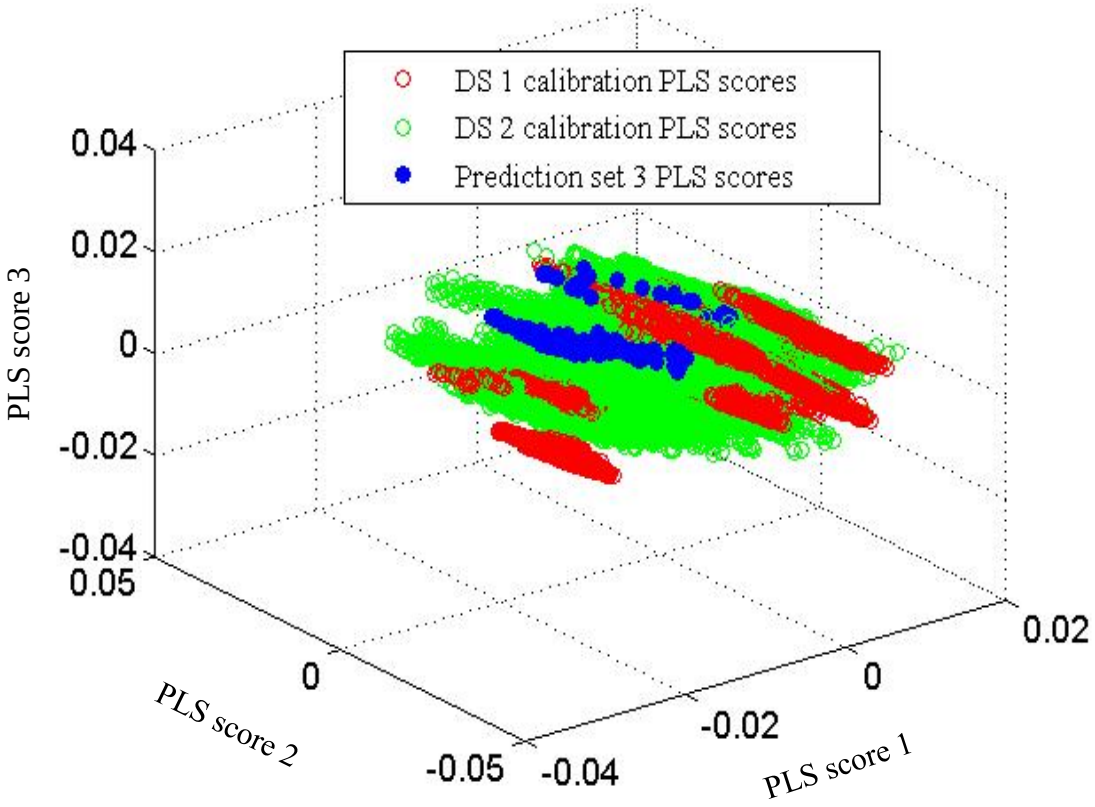


Figure 4.29 First three PLS scores plotted for the calibration data set and prediction set 3. Red, green, and blue symbols denote calibration patterns from DS1, calibration patterns from DS2, and prediction set 3, respectively. Clear overlap of the patterns is noted. This verifies that the calibration and prediction data are consistent.

Conclusions

In this chapter, the nocturnal hypoglycemic alarm algorithm presented in Chapter 3 was tested with an *in vitro* simulation study that served to mimic glucose excursions similar to those that occur in the human body. A synthetic sample matrix was constructed from urea, lactate, and triacetin in phosphate buffer to provide a challenge to the ability to extract glucose information selectively from NIR spectra in the combination region.

This study provided a first test of one of the key components of the alarm algorithm, the use of differential spectra computed relative to a glucose-containing reference spectrum. Within the calibration data, the calculation of all combinations of spectral ratios within time blocks served to expand the concentration data space. Further, the calculation of differential spectra served to simplify the resulting absorbance spectra by removing constant features of the spectral background.

The successful use of the PLDA method to implement the alarm decision provided verification that a pattern classification approach can be employed to identify concentration levels within NIR spectra. The iterative nature of the training of the classifiers raises the possibility that the optimization may become trapped in local maxima. This was illustrated in the anomalous results obtained with one of the replicate classifiers trained for use with prediction set 2. The need for replicating the training and the benefits to using multiple classifiers in making the alarm decision was made clear through this example.

The developed alarm algorithm was tested with three external prediction sets. The results obtained were very promising. No missed or false alarms were observed for any of the prediction sets. The robustness of the methodology was also tested by collecting

prediction set 3 two years and six months later than the last day of the calibration data. The result for that prediction set (i.e., no missed or false alarms) clearly demonstrated the excellent robustness of the methodology to changes in instrumental characteristics with time. The use of differential spectra computed relative to a same-day reference was considered to be a key component of the observed robustness of the methodology.

CHAPTER 5

STUDIES TO PREPARE SKIN TISSUE PHANTOMS TO SIMULATE *IN VIVO* NEAR-INFRARED SPECTRAL BACKGROUNDS

Introduction

In Chapter 4, the proposed nocturnal hypoglycemic alarm algorithm was tested *in vitro* with relatively simple sample matrixes. The work focused on the use of significant spectral interferences to glucose such as lactate, urea, and lipids (modeled by triacetin), but the measurements were made in a simple phosphate buffer solution. No attempt was made to incorporate the solid components of tissue into the simulation. However, the non-aqueous components in skin tissue can be the source of significant spectral variation in actual noninvasive glucose measurements made through tissue. For example, proteins and fat present in the tissue can have a major effect on the NIR spectrum.

The main absorber in skin tissue in the combination region of the NIR is water. The second most abundant absorber is protein. Two major types of protein present in skin tissue are collagen and keratin. Collagen is a type of elastic protein that is prominent in the dermal layer of the skin and keratin is a rigid protein that is rich in the outer layer of the skin, the epidermis. As shown in Figures 5.1 and 5.2, the absorption features associated with these two protein types can be seen in both the 4400-4200 cm^{-1} region and near 4600 cm^{-1}. Absorption features due to the combination of C-H bond stretching and bending in the long alkyl chains of fat-type molecules can also be observed between 4400 and 4200 cm^{-1}.

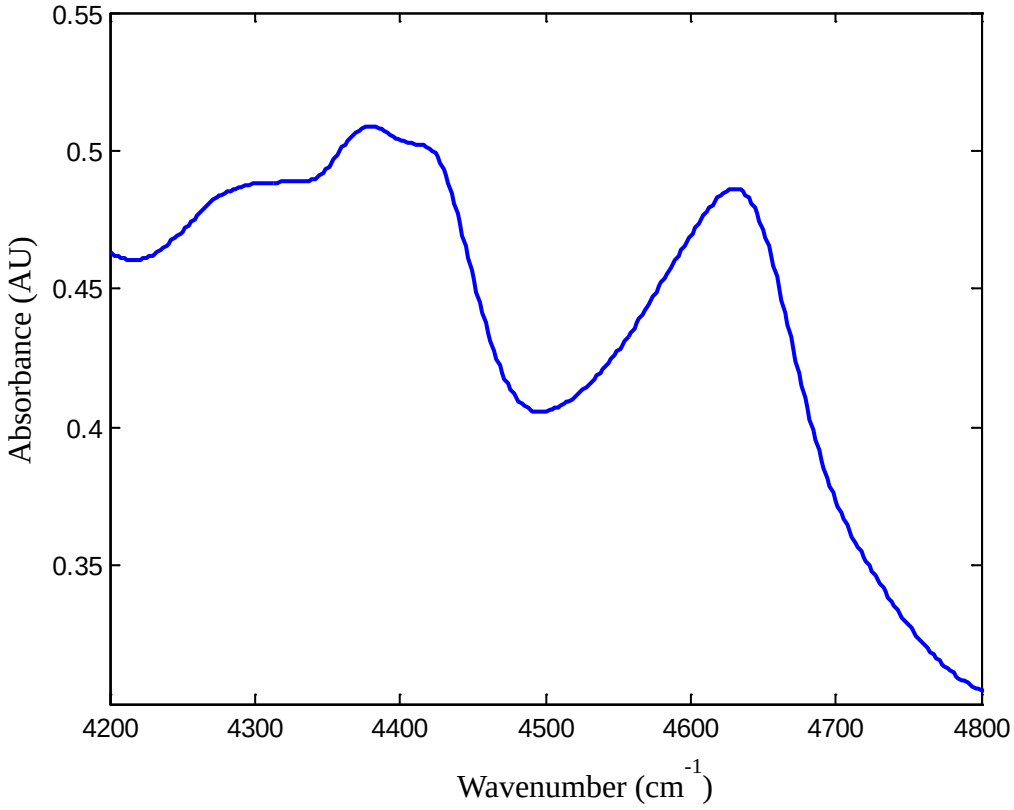


Figure 5.1 Absorbance spectrum of keratin film relative to an open-beam air background. Protein absorption features can be seen at 4400-4200 cm^{-1} and in the region of 4600 cm^{-1}.

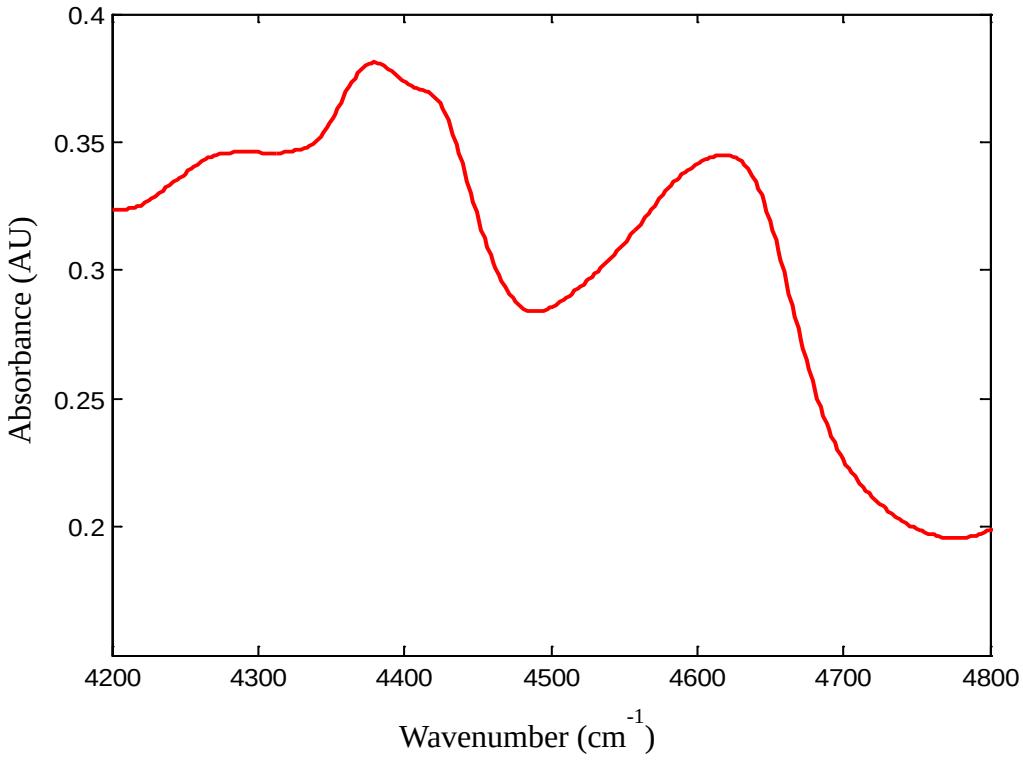


Figure 5.2 Absorbance spectrum of gelatin (hyrdrolyzed collagen) film relative to an open-beam air background. Protein absorption features can be seen at 4400-4200 cm^{-1} and in the region of 4600 cm^{-1}.

While the selection of a vascular region of the skin tissue where the fat component is relatively low can minimize the C-H signals arising from fat, any noninvasive measurement of tissue will be subject to interference from skin proteins such as keratin and collagen. Moreover, interference in the region near 4400 cm^{-1} can have a

direct effect on a glucose measurement because of spectral overlap with the glucose C-H combination band located there.

The objective of the research discussed in this chapter was to prepare tissue phantoms composed of two main proteins that exist in human skin tissue, keratin and collagen. Once prepared, these skin tissue phantoms will be incorporated into dynamic studies such as those described in Chapter 4 to simulate the skin tissue background in noninvasive NIR measurements. In this chapter, the preparation of these phantoms is explored and parameters such as reproducibility, stability, and thickness are examined. The ability to simulate the spectra of human tissue is also investigated. In Chapter 6, *in vitro* simulation studies employing the phantoms are described.

Experimental Methods

Preparation of Protein Films

This research was based on the construction of films of keratin and collagen that could be inserted into the optical path of a NIR transmission measurement. The basic approach was to coat each protein onto an optically transparent window by dissolving the protein with an appropriate solvent, placing the resulting solution into a sample well containing the window, and then allowing the solvent to evaporate. By repeating this procedure, protein films of desired thicknesses could be constructed.

The sample well compartments used for the preparation of the films were assembled as described below. A sapphire window (25 × 0.5 mm, Meller Optics, Providence, RI) was inserted into the bottom of a metal lens holder equipped with a threaded retaining ring (Model SM1L05, Thorlabs, Inc., Newton, NJ). Four to five Teflon

spacers (25 × 1.1 mm, McCarthy Scientific, Fallbrook, CA) were inserted tightly on the top of the sapphire window to make a sample well as shown in Figure 5.3. These spacers were tightened with the retaining ring to prevent leakage of solution. As shown in Figure 5.3, the outside of the lens holder was labeled at 14 equally spaced positions (1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E).

Protein solutions were prepared by dissolving the required amounts of the protein powder in 88 % (v/v) formic acid (ACS reagent, Fisher Scientific, Fair Lawn, NJ). For the preparation of the gelatin (hydrolyzed collagen) solution used to model collagen in the dermis, 0.75 g type-A gelatin powder form porcine skin (Product No.G8-150, Sigma-Aldrich, Inc., St. Louis, MO) was dissolved in 30.00 mL formic acid while stirring for approximately 45 min. For the preparation of the keratin solution, 2.00 g keratin powder (Product No.K3030, Spectrum Chemical Mfg. Corp., Gardena, CA) was dissolved in 10.00 mL formic acid while stirring for approximately 45 min.

Two types of protein films were prepared for the experiments described here: (1) films with approximately uniform thickness and (2) films with a wedge-shaped cross-section that had varying thickness around the circumference of the window. The need for films with variable thickness was motivated by a desire to incorporate variations in skin thickness and composition into subsequent *in vitro* simulation experiments such as those described in Chapter 4. Even for a single individual, changes in the exact site of the noninvasive measurement, changes in skin hydration, or the effects of applying pressure to the tissue can all cause a change in the effective composition of the tissue sample within the optical path of the spectrometer.

For the films with uniform thickness, approximately 25 drops of protein solution

were applied to the sample well and the sample holder was placed in a horizontal orientation into a glass desiccator equipped with drierite (W.A. Hammond Drierite Co, Ltd., Xenia, OH). For the preparation of the films with variable thickness, the sample holder was placed at an approximately 30° angle in the desiccator and the solutions were applied in different amounts to the sides of the holder.

Once the solutions were added, the sample holders were placed in the desiccator and allowed to dry. After 4-5 days in the desiccator, the same amounts of the protein solutions were applied to make the films thicker until the desired thickness was acquired.

Collection of Spectra of Protein Films

Spectra of the protein films were collected in transmission mode with a Nicolet 6700 FT spectrometer (Thermo Fisher Scientific, Inc., Waltham, MA) equipped with a CaF_2 beamsplitter and liquid-nitrogen-cooled InSb detector. A 50-W tungsten-halogen light bulb (Gilway Technical Lamp, Peabody, MA) equipped with an integrated, gold-coated reflector was used as the light source and directed into the spectrometer through the emission port. The source was powered at 4.96 V with an E3633A 200-W DC power supply (Agilent Technology, Van Nuys, CA). A metal neutral density screen (Screen C, Thermo Fisher Scientific, Inc.) in combination with a 63% transmittance metal thin-film type neutral density filter (Rolyn Optics, Inc., Covina, CA) were placed in front of the sample to prevent detector saturation. A K-band optical interference filter (Barr Associates, Westford, MA) was placed in the optical path to limit the spectral range to 5000 to 4000 cm^{-1}.

The lens tube containing the protein-coated sapphire window was placed in the optical path of the spectrometer by use of a V-shaped slide mount holder as shown in Figure 5.3. A retaining screw was used to secure the lens tube to the sample holder. Adhesive heating pads (Omega Engineering, Inc., Stamford, CT) were applied to the sample holder and a variable-voltage AC transformer (Powerstat 116, Superior Electric Co., Bristol, CT) was employed to maintain the temperature of the protein film to the region of 37 °C. The temperature of the holder was monitored with a Type-T thermocouple and thermocouple meter (Omega Engineering, Inc.) and manual control of the transformer was used to maintain the temperature.

The experimental setup used for the data collection is shown in Figure 5.4. Spectra were collected at a resolution of 8 cm^{-1} with 128 co-added scans. Interferograms were sampled at every zero-crossing of the HeNe reference laser, producing a maximum spectral frequency of 15,798.25 cm^{-1}. Each spectrum required 2 minutes for collection. Background air spectra were collected before the collection of the film spectra and were used for the calculation of spectra in absorbance units (AU). Collected inteferograms were Fourier processed to single-beam spectra with Happ-Genzel apodization, Mertz phase correction, and one level of zero of filling.

Human Subject Data

Noninvasive NIR spectra of human skin tissue were provided by our collaborators in the Arnold research group at the University of Iowa. These human subject data were used to help evaluate the degree to which the spectra of the protein films were effective in modeling skin tissue.

Human skin tissue spectra were collected with a Nicolet 6700 FT spectrometer (Thermo Fisher Scientific Inc. Waltham, MA) configured in the same manner as described previously. A similar external light source was used.

Figure 5.3 Sample holder assembly used for the preparation and measurement of the protein films. Films were coated onto sapphire windows held within a metal lens tube. The lens tube was placed on a V-shaped holder attached to a standard slide-mount sampling accessory for use in placing the sample into the spectrometer. A retaining screw was used to secure the rotational position of the lens tube in the sample holder. Fourteen labeled positions allowed reproducible placement of the films on the holder.

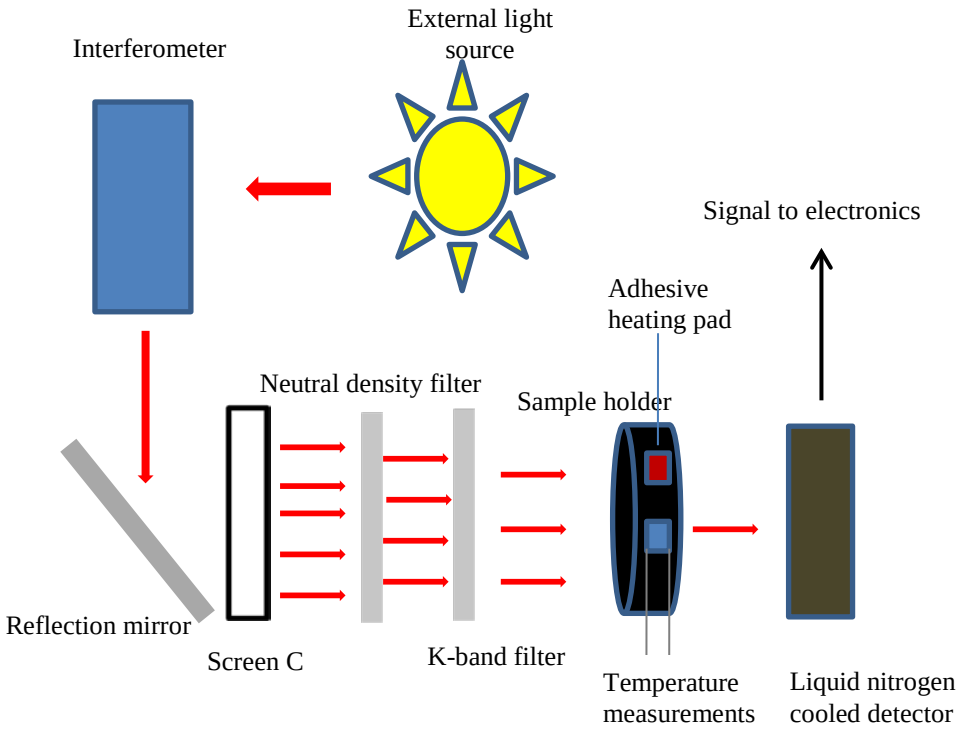


Figure 5.4 Experimental setup used for the collection of the spectra of the protein films. Light from the external source was directed through the interferometer before passing through two neutral density filters, a bandpass filter, and the sample. Transmitted light was directed onto the detector.

To bring the light from the spectrometer to the skin tissue and to collect the transmitted light, a custom sapphire-rod interface was used. A four-stage thermoelectrically cooled extended-wavelength InGaAs detector with a diameter of 2 mm and a 2.6 μm long wavelength cutoff (Judson Technologies, Montgomeryville, PA) was used to detect the transmitted light.

Figures 5.5 A and B show a photograph of the instrumentation and a detailed view of the human interface, respectively. A K-band interference filter placed in the optical path limited the spectral range to 5000 to 4000 cm^{-1}. Spectra were collected at a resolution of 8 cm^{-1} with 16 co-added scans. Interferogram points were sampled at each zero-crossing of the HeNe reference laser to yield a maximum spectral frequency of 15,798.25 cm^{-1}. Each spectrum required 8.5 seconds for collection. The interferograms were Fourier processed to single-beam spectra with triangular apodization, Mertz phase correction, and one level of zero-filling. The resulting spectral point spacing was 1.93 cm^{-1}.

The skin tissue spectra used were collected from volunteers who were adult males and females from 18 to 65 years old including people with and without diabetes. Different ethnic groups were also represented. All protocols used were approved by the Institutional Review Board at the University of Iowa.

Volunteers fasted overnight with food restriction for 10 hours. Before spectra were collected, the person's weight and blood pressure were recorded along with information such as age, sex, and ethnicity. The interface was applied to a fold of skin on the back of the hand and the spectral data were acquired while the participant sat beside the spectrometer. Spectra were collected repeatedly and continuously without movement

for a period of 90 minutes. Air reference spectra were also collected before and after each set of skin spectra.

In total, 48 subjects participated in this study. For the work described here, subjects 4-9 only were used in the evaluation of the ability to model spectra of human tissue with the spectra of the protein films. Selection of these subjects was done on the basis of low values of computed spectral noise levels. The noise calculation was performed as described in Chapter 3.

Computations

All computations were performed with code written in the Matlab development environment (Version 7.4, The MathWorks, Natick, MA) implemented on a Dell Precision 670 workstation (Dell Computer Corp., Austin, TX) operating under Red Hat Linux (Version 5.3, Red Hat, Inc., Raleigh, NC).

Results and Discussion

Characterization of Human Subject Spectra

For the subsequent data analysis, every five consecutive single-beam spectra of each of the human subjects were averaged to increase the signal-to-noise ratio. For subjects 4-9, this calculation yielded 127 averaged spectra for each subject, and an overall total of 762 spectra for analysis. The short-term noise for these signal averaged spectra was evaluated across the wavenumber region of 4300-4500 cm^{-1} for consecutive pairs of spectra. The RMS noise values calculated from a third-order polynomial fit as discussed in Chapter 3 were within the range of 30-41 μAU.

(A)

(B)

Figure 5.5 Experimental setup used for the collection of noninvasive NIR spectra of human tissue.[142] A. Overview of the instrumentation. B. Sapphire rod interface.

The mean signal-averaged single-beam spectra for these human subjects are shown in Figure 5.6 A. The average air spectrum collected on the corresponding day was used as the reference in the calculation of skin spectra in AU. Figure 5.6 B shows the computed mean absorbance spectra of human subjects 4-9. It can be observed that the spectral shapes for these human subjects are similar.

As expected, water is the most significant contributor to the overall spectrum, producing a parabolic-shaped response in the absorbance spectrum. This shape originates from the tails of the two large water absorption bands centered near 5200 and 3300 cm^{-1}. The absorbance peak at 4600 cm^{-1} is due to N-H combination bands arising from proteins in the skin tissue, mainly collagen and keratin. The features between 4250 and 4350 cm^{-1} are produced by C-H combination bands in the proteins, as well as from the small amount of fat present. The measurement site at the top of the hand has little fat. The significant baseline variations in the absorbance spectra are due to different magnitudes of scattering among the subjects and variation in the skin thickness placed between the sapphire rods of the interface. The variations in the skin tissue components among these human subjects lead to slight variations in the relative peak heights.

Modeling Spectra of Human Subjects

Noninvasive NIR spectra of human skin can be fitted by MLR (Chapter 3) to a linear function that incorporates the additive features of absorption due to water, fat, keratin, and collagen (gelatin).

This function takes the following form:

$$A_{skin} = \beta_w A_{water} + \beta_f A_{fat} + \beta_g A_{gelatin} + \beta_k A_{keratin} + e \qquad (5.1)$$

where the terms A_{skin}, A_{water}, A_{fat}, $A_{gelatin}$, and $A_{keratin}$ represent the absorbance contributions corresponding to skin, water, fat, gelatin, and keratin. The corresponding β_i terms represent regression coefficients for the fit and e denotes the residual (error) from the fit. The absorbance terms in Eq. 5.1 are all wavelength dependent.

Additional baseline correction terms (e.g., constant offset or linear slope) can also be added to Eq. 5.1 to account for variation such as scattering effects that cannot be explained by the spectra of the chemical constituents.

The magnitudes of the regression coefficients can be computed by applying MLR to a set of pure-component spectra that represent the different skin constituents. The regression coefficients approximate the amount of each component in the skin tissue matrix relative to that of the corresponding sample that was used to generate the pure-component spectrum. For example, if a spectrum of the keratin film produces a value of 0.5, the film is two times thicker than the corresponding amount of keratin present in the tissue. This assumes that the spectral response of the pure-component spectrum matches that in the tissue.

The objective of the research described here was to prepare protein films that gave regression coefficients close to unity. If produced, these films could be placed into the optical path of the spectrometer and thereby simulate the presence of the corresponding material as it exists in human skin tissue.

The relative thicknesses of the prepared films were calculated with the modeling approach described by Eq. 5.1. Separate fits were performed with the constant-thickness and variable-thickness gelatin and keratin films.

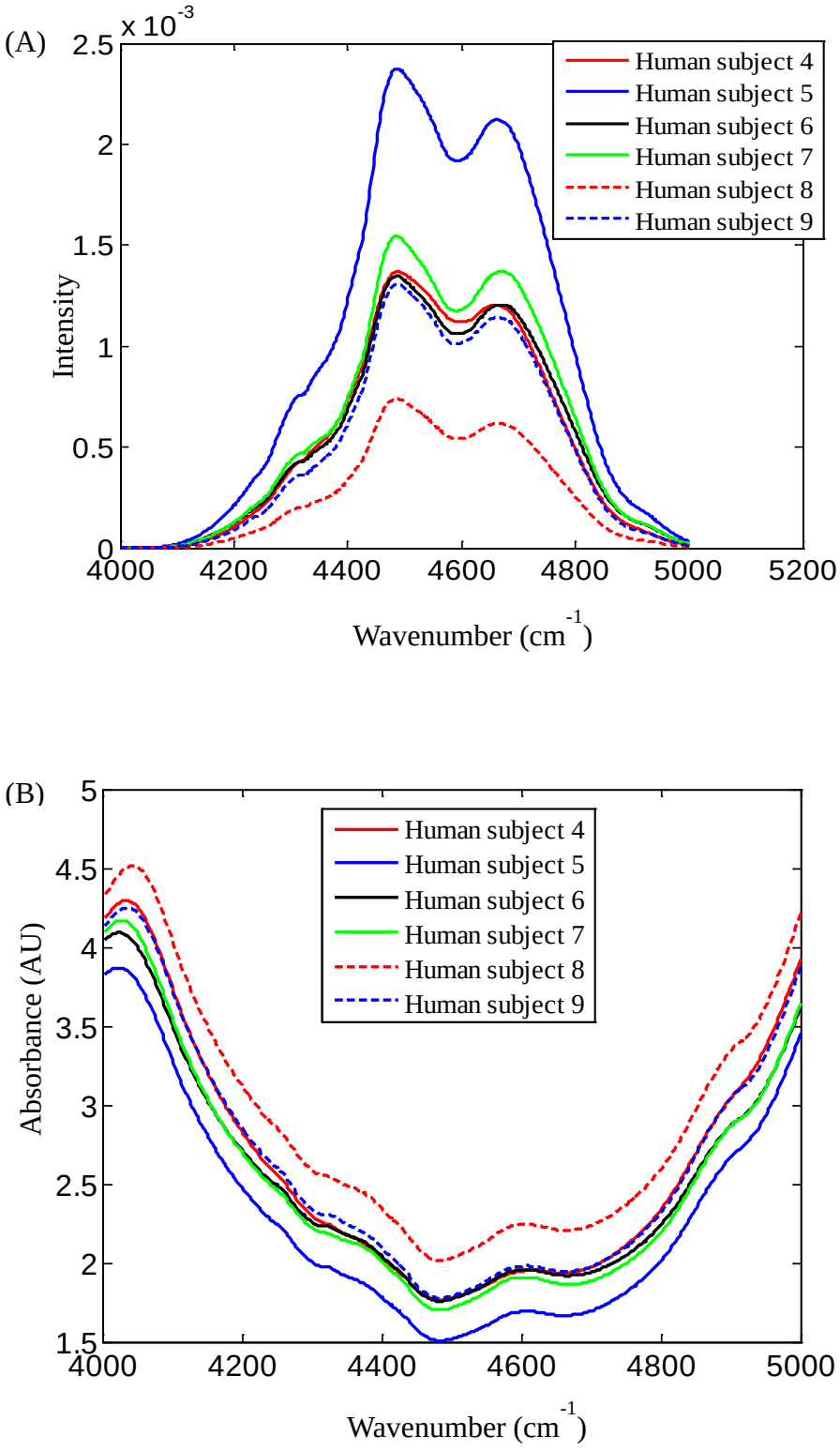


Figure 5.6 Mean signal-averaged spectra for subjects 4-9. A. Single-beam spectra. B. Absorbance spectra.

The regression used a series of four pure-component spectra including water, fat, gelatin film, and keratin film to fit the skin absorbance spectra. The water spectrum, A_{water}, in Eq. 5.1 was collected from a sealed 1-mm thick transmission cell. The fat spectrum, A_{fat}, was collected from a 1-mm thick sample of bovine fat. These two pure-component spectra were obtained from our collaborators in the Arnold research group. Figure 5.7 plots these absorbance spectra. The addition of constant offset and linear baseline correction terms to Eq. 5.1 was also investigated but these terms were not found to be helpful in improving the fits.

Table 5.1 presents the results obtained from the use of Eq. 5.1 to fit the 762 combined spectra from human subjects 4-9. Results are presented for fits using both the constant-thickness and variable thickness films. Also included in the table are the values of R^2 and the standard error of estimate (SEE) that describe the quality of each fit. The R^2 expresses the fraction of the variance in the measured human subject spectrum that is explained by the regression model, while the SEE is the square root of the mean squared error between the actual and fitted absorbance values (i.e., the mean squared residual). The regression coefficients for the gelatin and keratin films that were prepared at an approximately constant thickness were 0.6 and 1.4, respectively. The spectra of these films were displayed previously in Figures 5.1 and 5.2. The regression fit had a value of R^2 of 0.97 and the standard error of estimate (SEE) for the fit was 0.04 AU.

Figure 5.8 plots example spectra from human subjects 4, 5, and 6 along with the corresponding spectra predicted by the regression equation based on the constant-thickness films. The predicted spectra match the overall shapes of the experimentally

measured spectra, although there appear to be mismatches with the actual spectra in the protein absorption region near 4600 cm^{-1}.

Table 5.1 Results of fitting human subject spectra to pure components

Film Type	Films Used		Regression Coefficient ± Standard Error				R^2	SEE[a]
	Gelatin	Keratin	Gelatin	Water	Fat	Keratin		
Constant	Constant	Constant	0.6 ± 0.2	1.05 ± 0.01	-0.038 ± 0.005	1.4 ± 0.2	0.969	0.04
Variable	Position 1	Position 5	2.0 ± 0.3	0.944± 0.006	0.006 ± 0.003	1.3 ± 0.4	0.987	0.03
Variable	Position 1	Position A	1.7 ± 0.3	0.951 ± 0.006	0.007 ± 0.003	2.7 ± 0.6	0.987	0.03
Variable	Position C	Position 5	2.1 ± 0.4	0.969 ± 0.006	-0.005 ± 0.003	1.9 ± 0.3	0.987	0.03
Variable	Position C	Position A	1.9 ± 0.4	0.974 ± 0.006	-0.002 ± 0.003	3.6 ± 0.5	0.987	0.03

[a]Standard error of estimate in AU describing the quality of the fitted regression equation.

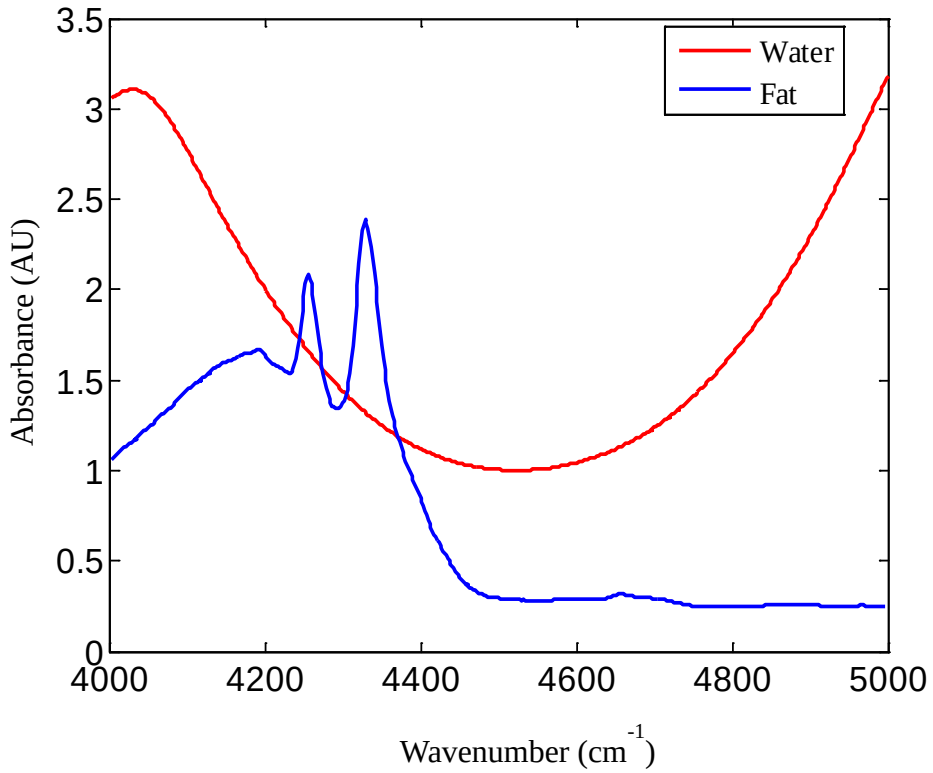


Figure 5.7 Pure-component spectra of water and fat used in fitting the regression model based on Eq. 5.1.

For the variable-thickness films, two positions on each of the keratin (5 and A) and gelatin (1 and C) films were used as examples of minimum and maximum thickness, and the regression equations were computed with all four combinations of the film positions. The spectra collected from the variable-thickness films are shown in Figures 5.9 and 5.10.

When the spectra collected with position 1 of the gelatin film and position 5 of the keratin film were used for the regression fit, the corresponding regression coefficients for gelatin and keratin were 2.0 and 1.3, respectively. The value of R^2 for the corresponding fit was 0.99. Example spectra from subjects 4, 5, and 6 are plotted in Figure 5.11 along with the predicted spectra resulting from the fit. As observed previously in Figure 5.8, while the overall fit is good, there are mismatches in the protein absorption region near 4600 cm^{-1}.

When the spectra collected from position 1 of the gelatin film and position A of the keratin film were used, the regression coefficients were 1.7 and 2.7 for gelatin and keratin, respectively. Thus, two separate estimates of position 1 of the gelatin film produced regression coefficients of 2.0 and 1.7, or an average value of 1.8 ± 0.3. The error associated with this reported mean is the pooled value computed from the standard errors of the coefficients in the two regression equations. These values were deemed consistent enough to say with confidence that position 1 of the variable-thickness film was approximately twice the effective thickness of collagen-like protein in an average *in vivo* spectrum collected with the interface shown in Figure 5.5.

Through a similar use of the data in Table 5.1, the average regression coefficient for position C in the gelatin film was 2.0 ± 0.4. The corresponding values for the keratin

film at positions 5 and A were 1.6 ± 0.4 and 3.2 ± 0.6, respectively. The errors reported here are again the pooled values obtained from the standard errors in the regression coefficients.

Evaluation of Sample Placement

The reproducibility of the placement of the sample holder in the sample compartment is very important during the data collection. If spectra are to be compared across multiple days, it must be possible to place the protein films in the optical path in such a way that the contribution of the film to the overall measured spectrum is reproducible.

To study the effect of the film placement in the sample compartment, the constant-thickness films were used. The sample holder was placed in the sample compartment and spectra were collected. The sample holder was then removed from the sample compartment and replaced again in the same position. Additional spectra were then collected. This procedure was repeated for four trials conducted over a total time of 24 minutes.

The absorbance spectra produced in this study are shown in Figures 5.12 A and B for gelatin and keratin, respectively. The absorbance spectra showed deviations of approximately 0.59 % for gelatin and 1.10% for keratin. These values were computed as the average % relative standard deviation computed across the four trials over the 4200 to 4800 cm^{-1} range.

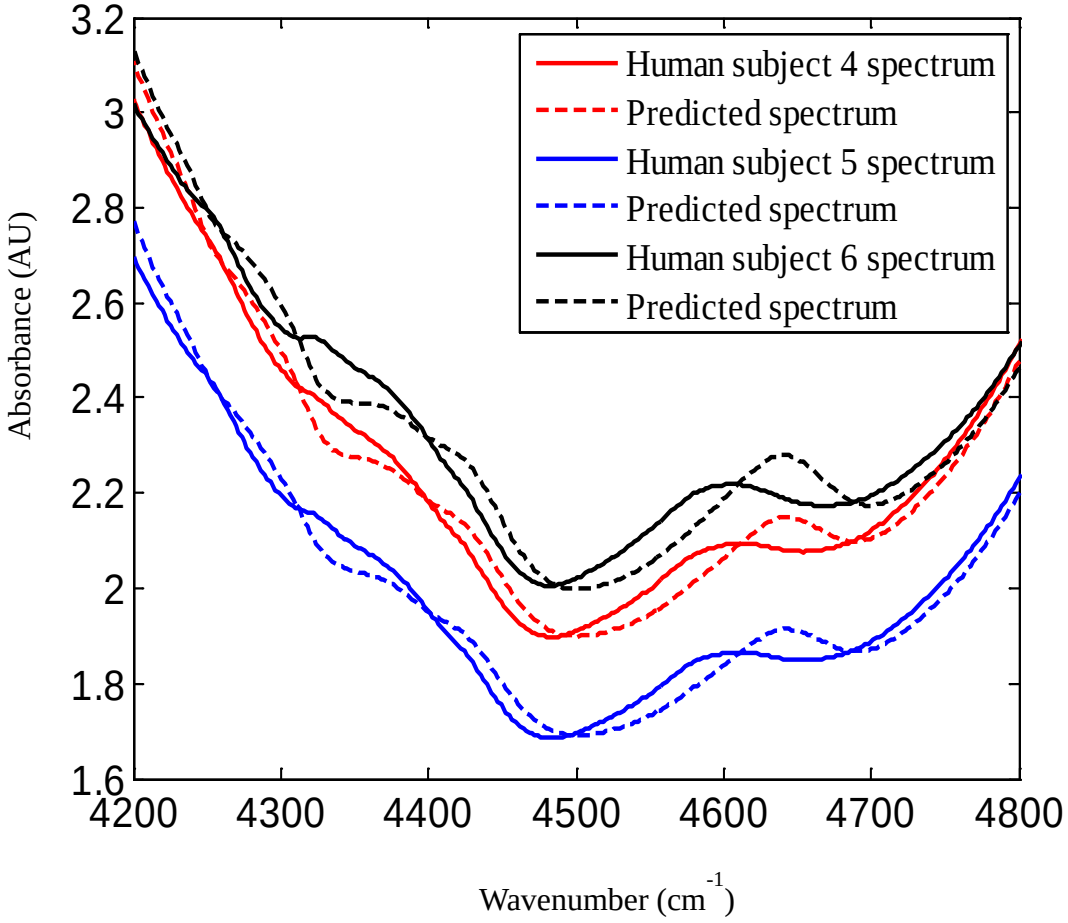


Figure 5.8 Human subject absorbance spectra (solid lines) from subjects 4, 5, and 6 plotted together with the predicted spectra (dashed lines) computed from the linear regression fit based on Eq. 5.1. Spectra of the constant thickness films were used in fitting the regression equation.

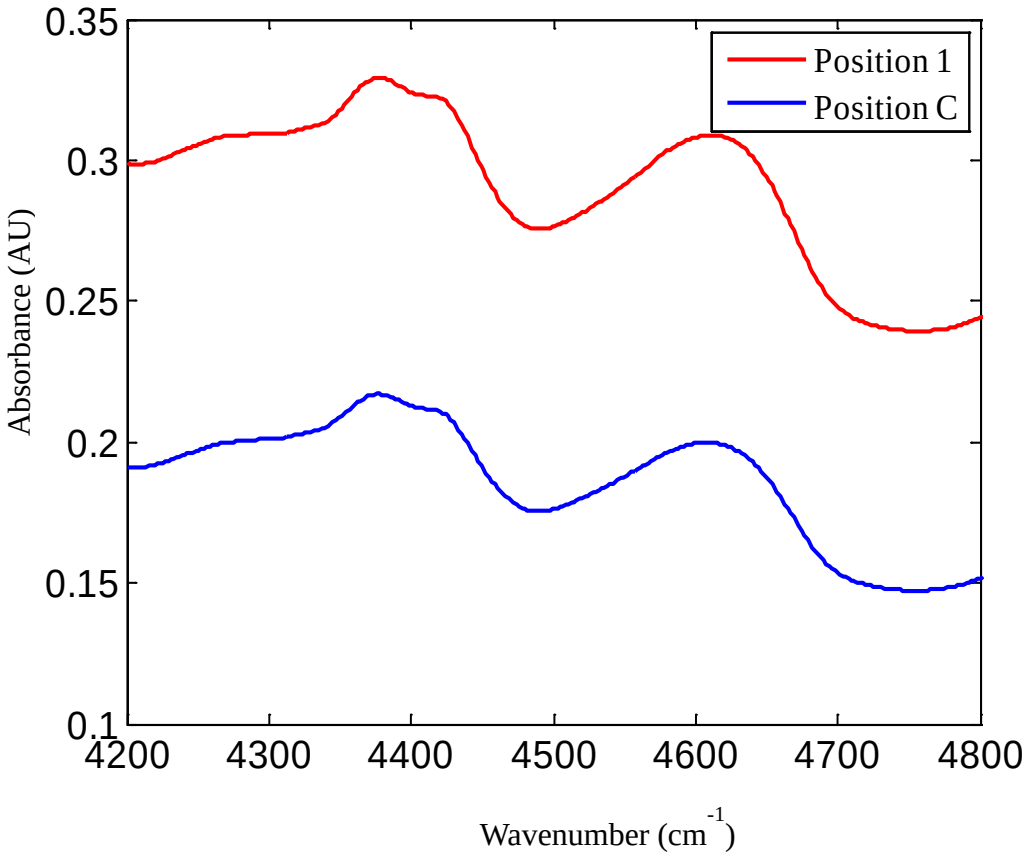


Figure 5.9 Absorbance spectra of the variable-thickness gelatin film at positions 1 and C are displayed. Protein absorption features can be seen at 4400-4200 cm^{-1} and in the region of 4600 cm^{-1}. These spectra represent the minimum and maximum thickness regions of the films.

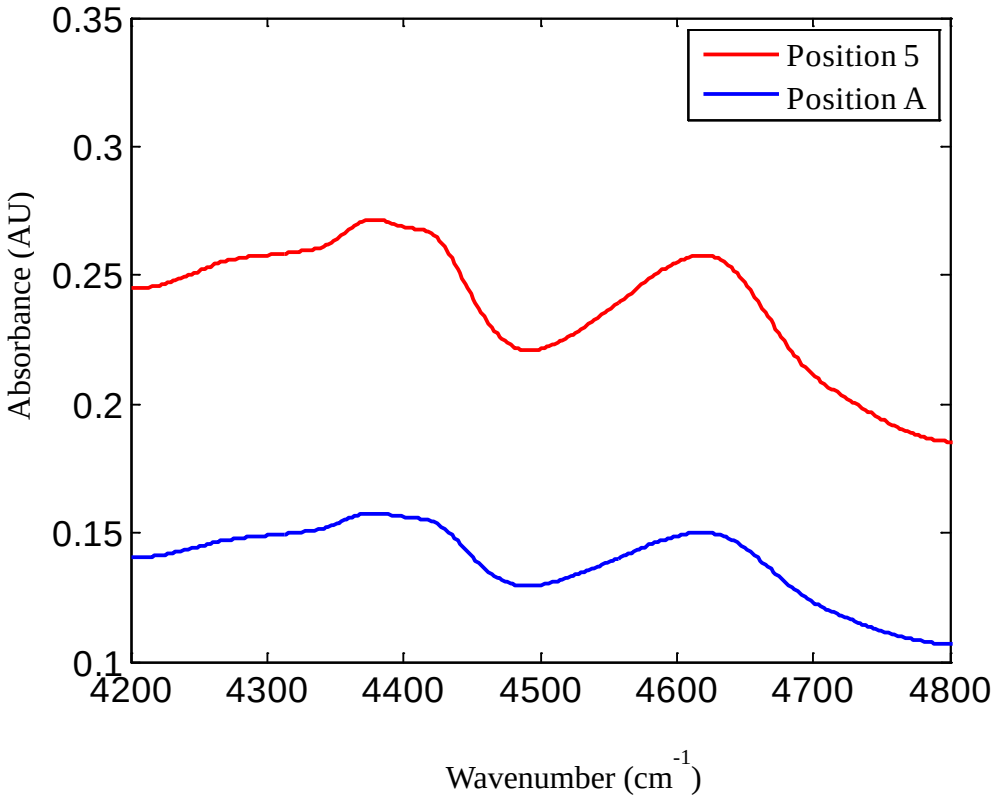


Figure 5.10 Absorbance spectra of the variable-thickness keratin film at positions 5 and A are displayed. Protein absorption features can be seen at 4400-4200 cm^{-1} and in the region of 4600 cm^{-1}. These spectra represent the minimum and maximum thickness regions of the films.

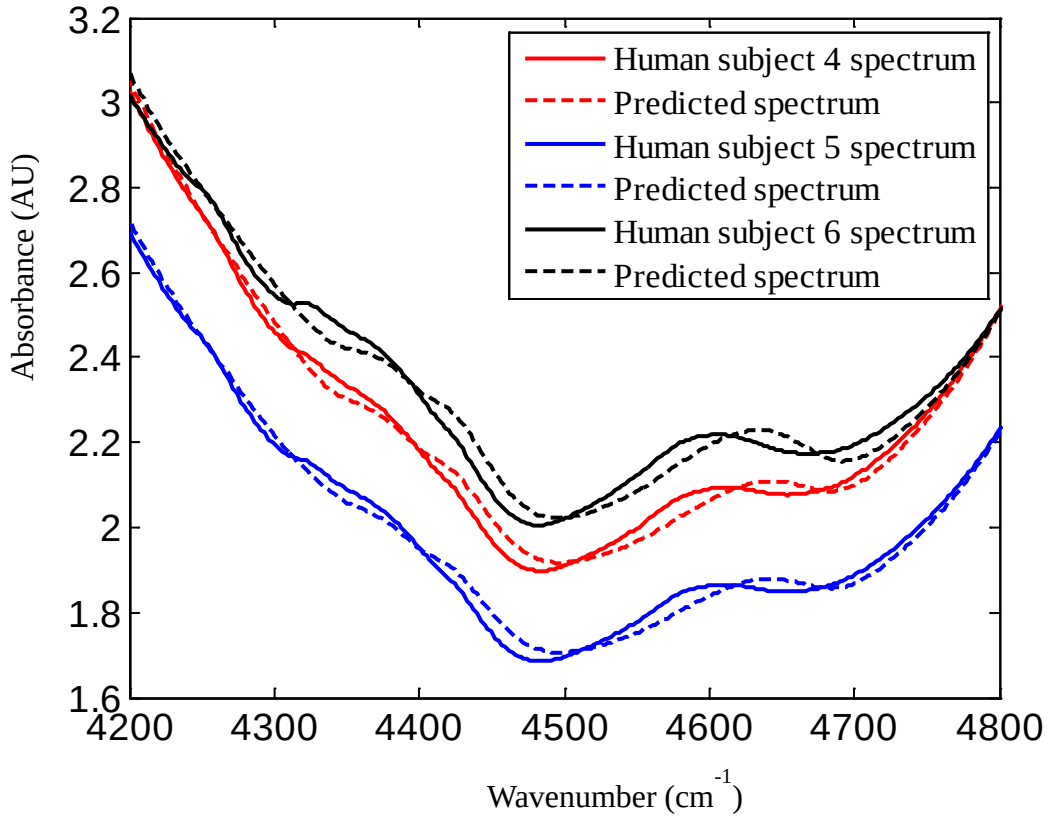


Figure 5.11 Absorbance spectra from human subjects 4, 5, and 6 plotted together with the corresponding predicted spectra obtained from the regression fit based on Eq. 5.1. Human subject spectra were fit to pure-component water and fat spectra together with spectra collected from the variable-thickness keratin and gelatin films. Positions 1 and 5, respectively, of the gelatin and keratin films were used.

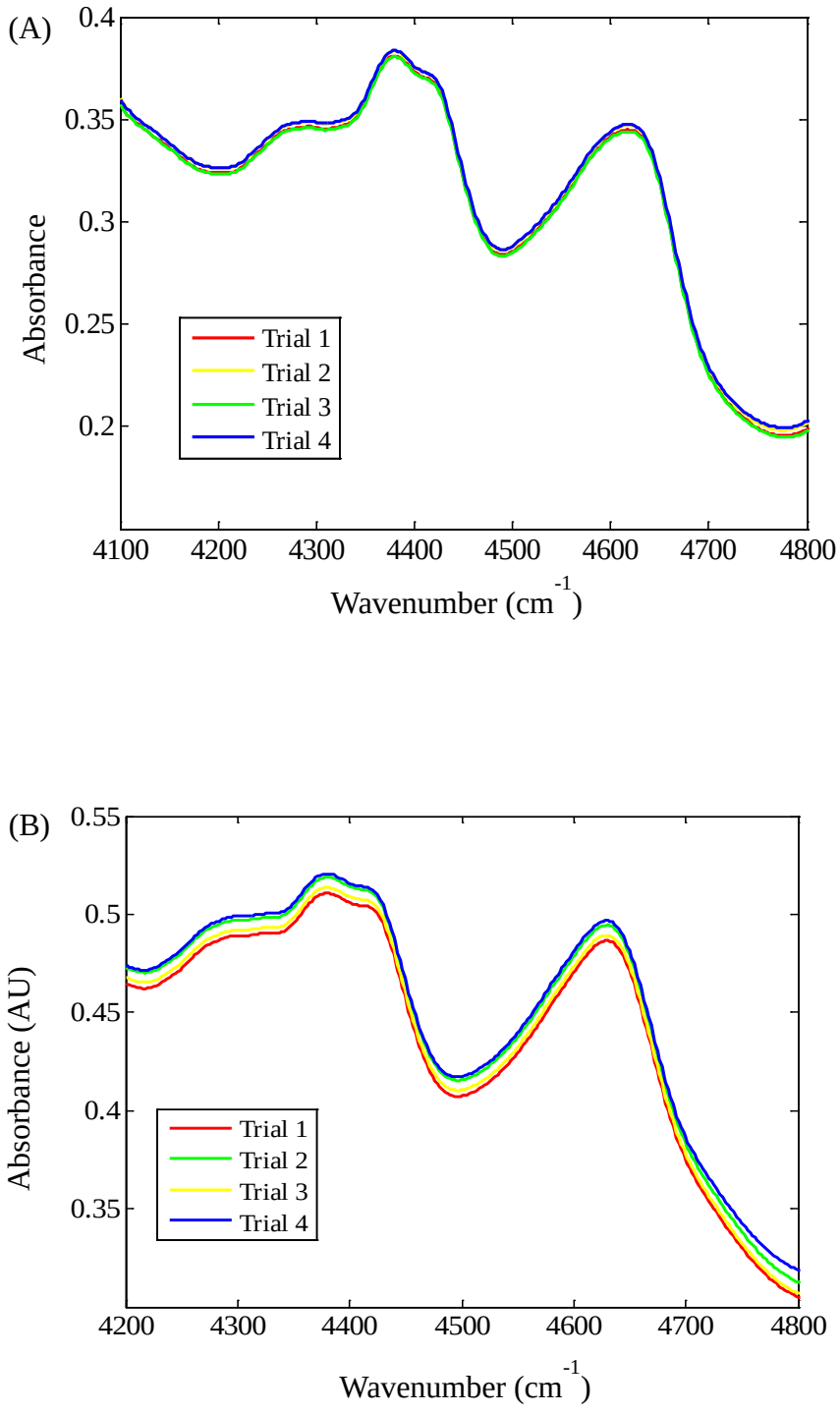


Figure 5.12 Study of the reproducibility of placement of the gelatin (A) and keratin (B) films in the sample compartment of the spectrometer. Absorbance values were computed relative to a single open-beam air background collected at the start of the experiment.

Evaluation of Stability of Prepared Films

The short-term stability of the prepared films was also assessed using the constant-thickness films. Absorbance spectra for the gelatin film at the same position were collected over four different days that spanned 14 days. Figure 5.13 shows the collected spectra. The average % relative standard deviation computed across the 4200 to 4800 cm^{-1} range was 1.96 %.

The same study was also performed with the keratin film. Spectra were acquired during four different days that spanned 13 days. Figure 5.14 shows the resulting absorbance spectra. The average % relative standard deviation computed across the 4200 to 4800 cm^{-1} range was 0.56 %.

Evaluation of Exposure of Films to Source Energy

Simulation studies of the type described in Chapter 4 consist of individual blocks of spectra collected over 2-3 hours. It was a concern that continuous exposure of the protein films to the heating effects produced by the incident source energy could cause degradation in the protein structures and hence spectral variation. This was studied by collecting spectra of the prepared films continuously for two hours.

Spectra of the constant-thickness gelatin film were collected for two hours without removing the sample holder from the spectrometer. Each mean spectrum collected at 20-minute increments was converted to absorbance units relative to an open-beam air background collected at the start of the experiment The six resulting spectra are shown in Figure 5.15. The average % relative standard deviation computed across the 4200 to 4800 cm^{-1} range was 0.57 %.

A similar experiment was performed over 1.5 hours for the constant-thickness keratin film. Each mean spectrum collected at 20-minute increments (10 minute interval for the spectra collected from 81-90 minutes) was converted to absorbance units relative to an air background recorded at the beginning of the experiment. The five resulting spectra are shown in Figure 5.16. The average % relative standard deviation computed across the 4200 to 4800 cm^{-1} range was 0.26%.

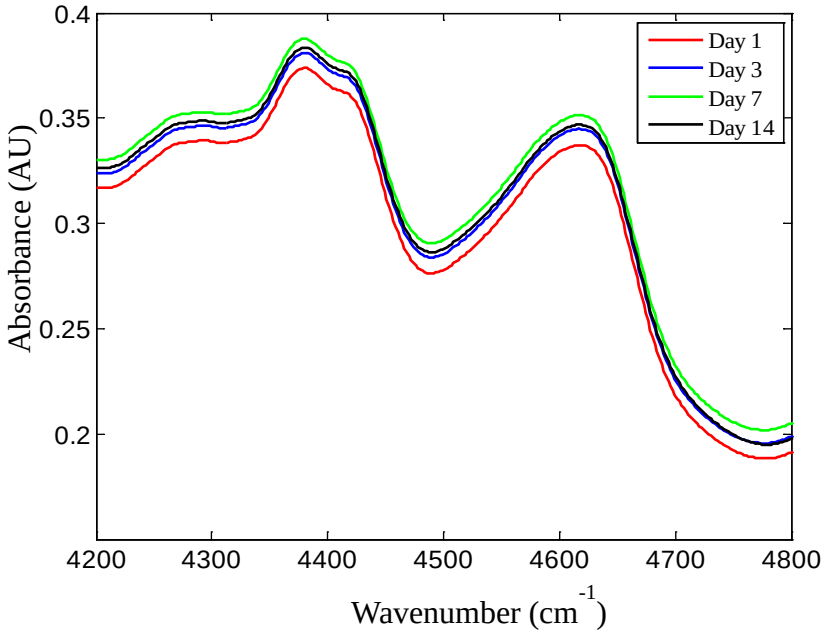


Figure 5.13 Study of the short-term of consistency of the prepared gelatin film. Spectra were collected over four different days that spanned 14 days.

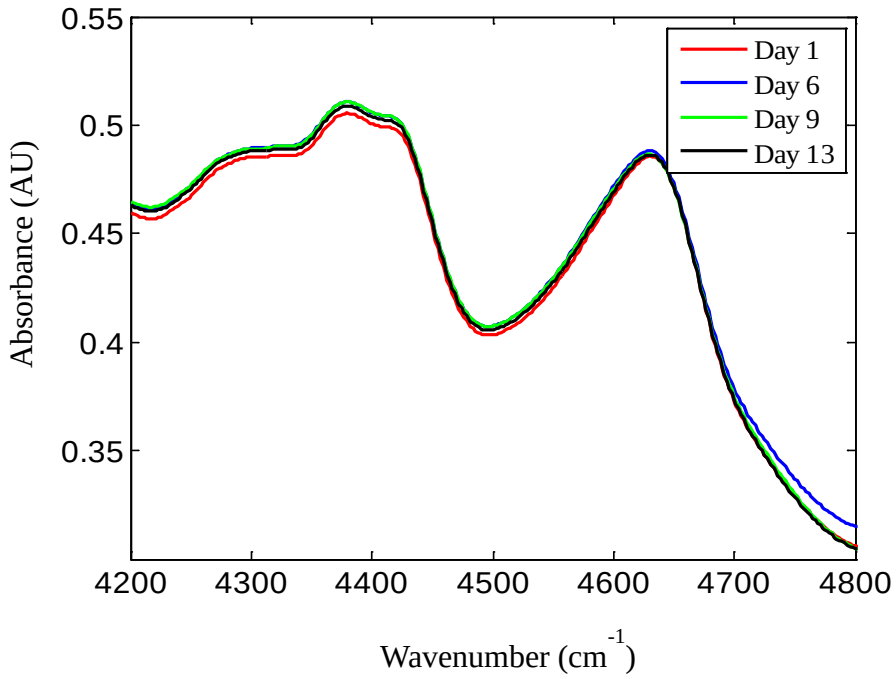


Figure 5.14 Study of the short-term of consistency of the prepared keratin film. Spectra were collected over four different days that spanned 13 days.

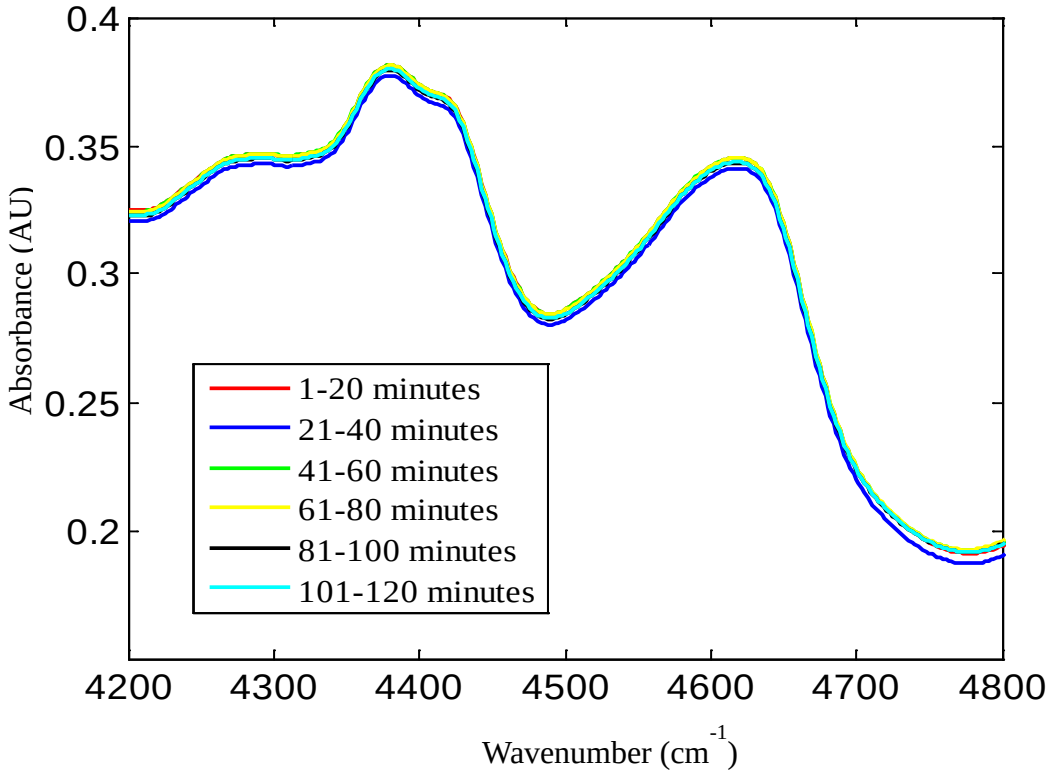


Figure 5.15 Study of the effect of the duration of the data collection on the prepared gelatin film. Spectra were collected continuously for two hours.

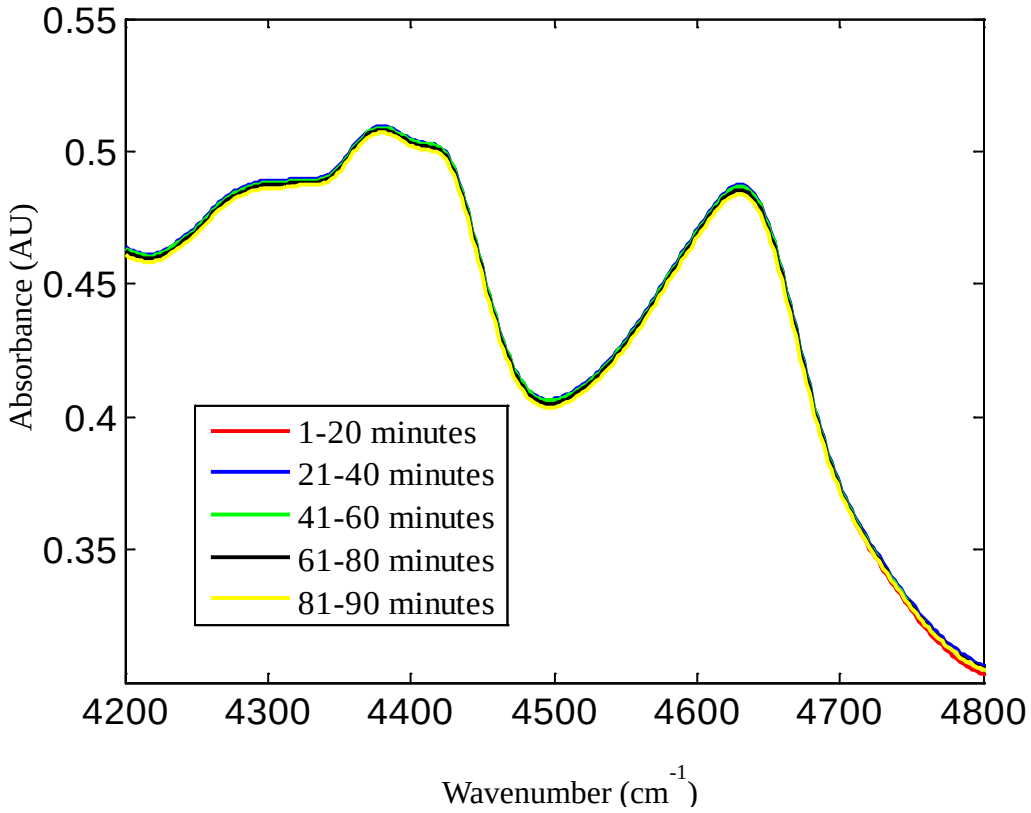


Figure 5.16 Study of the effect of the duration of the data collection on the prepared keratin film. Spectra were collected continuously for 1.5 hours.

Conclusions

This chapter discussed an approach to prepare a tissue phantom composed of two protein components present in human skin tissue, keratin and collagen. The relative thicknesses of the prepared films were evaluated using a linear regression fit to *in vivo* NIR spectra collected from human subjects. Values of R^2 for the fitted regression equations ranged from 0.97 to 0.99.

On the basis of the regression results, the prepared gelatin and keratin films provided an attenuation in the range of that observed in tissue samples measured *in vivo* with the interface shown in Figure 5.5. While the spectral match between the films and protein features in the human subject data was not exact, the films were judged to be useful in adding tissue-like components to the NIR spectral background employed in *in vitro* simulations of the type described in Chapter 4.

In addition, the variable-thickness films produced through this work provide some ability to change the effective protein content of a simulated tissue sample in an experimentally simple way. This allows the simulation of changes in protein composition that might occur in *in vivo* measurements conducted across individuals or at different measurement sites on the same individual.

Three studies were performed to gain insight into the reproducibility of use of the films. The spectral variation induced by removing and replacing the holder containing the films into the sample compartment of the spectrometer was evaluated. Across four trials, this study yielded an average % relative standard deviation in absorbance of 0.59 % and 1.10 % for gelatin and keratin films, respectively, across the 4200 to 4800 cm^{-1} range. The short-term spectral reproducibility of the prepared films was also evaluated and the

absorbance spectra collected on four different days over approximately two weeks demonstrated good consistency as measured by % relative standard deviations in the range of 1.96 % and 0.56 % for gelatin and keratin films, respectively. Finally, the effects of continuous exposure to the heating effects of the light source were studied over two hours and yielded % relative standard deviations in absorbance in the range of 0.57 % and 0.26 % for gelatin and keratin films, respectively.

Work next turned to the use of the prepared films in helping to implement an improved *in vitro* simulation experiment for evaluating the capabilities of the proposed nocturnal hypoglycemic alarm algorithm. This research will be described in Chapter 6.

CHAPTER 6

IN VITRO SIMULATION STUDIES USING PROTEIN TISSUE PHANTOMS FOR THE DEVELOPMENT OF A NOCTURNAL HYPOGLYCEMIC ALARM ALGORITHM BASED ON NEAR-INFRARED SPECTROSCOPY

Introduction

The nocturnal hypoglycemic alarm algorithm discussed in Chapter 3 was tested with *in vitro* simulation studies described in Chapter 4. In these studies, the effects of some of the main components present in the blood that can potentially interfere with a noninvasive blood glucose analysis in the NIR region were examined. However, the study did not include a realistic representation of the spectral background that would be encountered in a true *in vivo* measurement of tissue.

As discussed in the previous chapter, the heterogeneity of the chemical composition of the tissue matrix is responsible for a significant spectral background arising from both scattering and absorption of propagating photons. The components present in the skin tissue and the relative amounts and their distribution have an impact on the NIR spectrum collected. As a result, any successful *in vitro* simulation study to test the nocturnal hypoglycemic alarm algorithm should incorporate these elements of skin tissue.

Chapter 5 discussed in detail a successful approach to prepare a skin tissue phantom composed of two main skin tissue proteins, keratin and gelatin. In this chapter, the prepared tissue phantoms are utilized in *in vitro* simulation studies to test the hypoglycemic alarm algorithm. The study was made more complex by adding variable

levels of urea to the sample matrix, one of the main spectral interferences in the NIR analysis of glucose.

Two types of skin tissue phantoms were used for the multiple-day data collections described in this chapter. For the first study, the same orientation of the prepared skin tissue phantom was used in each day of the data collection. This simulated a scenario in which the skin tissue matrix does not change during the entire time span of the noninvasive data collection (i.e., during both the calibration and prediction phases of the data collection). While admittedly unrealistic as it applies to a true *in vivo* measurement, this scenario allowed a baseline performance level for the alarm algorithm to be established.

The second study attempted to simulate a more realistic scenario in which the skin composition changes during the time span of the data collection. This work was motivated by one of the observed problems associated with noninvasive glucose sensing: the heterogeneous nature of the skin tissue from location to location and the deformation of the tissue during the actual spectral acquisition when an interface such as that displayed in Figure 5.5 is used. These changes in the skin composition present in the optical path of the measurement can lead to variation in the effective path length and can also provide a challenge to any calibration procedure that assumes the spectral background has been adequately captured during the collection of the calibration data. To address this scenario, the variable-thickness protein films described in Chapter 5 were employed in the *in vitro* simulation studies to allow the incorporation of variation in tissue composition into the testing of the alarm algorithm.

Experimental Methods

The results presented in this chapter derive from two dynamic studies conducted in the presence of the skin tissue phantoms prepared in the laboratory as discussed in Chapter 5. The first study was a multiple-day data collection and in each day of the data collection, the constant-thickness tissue phantom was placed in the optical path in the same orientation. The second study was also a multiple-day data collection; however, the orientations of the variable-thickness films were changed during the data collection to simulate the presence of different tissue compositions.

Apparatus and Reagents

The chemical system used for both studies included three components, α-D-glucose, urea and 0.1 M, pH 7.4 phosphate buffer. The reagents used and preparation of the stock solutions of the aqueous components were as described previously in Chapter 4. To simulate glucose excursions in the body, the three stock solutions were mixed in different ratios using three peristaltic pumps (Rabbit-Plus and Dynamax Models, Rainin Instrument Co., Woburn, MA). The pump setup was the same as shown previously in Figure 4.4. A pump calibration was performed at the beginning of each day of data collection to correct for any differences in the theoretical pump speeds or deficiencies in the tubing. The total pump speed was maintained at a value of 10 rpm. The basic settings for the peristaltic pumps were the same as discussed in Chapter 4 and will not be discussed in detail here. Each stock solution was stationed in a water bath and regulated at ~55 °C to keep the flowing liquid temperature at 36.6-37.3°C.

As described in Chapter 4, the solution exiting the in-line mixer was flowed through a 20 mm-diameter circular aperture transmission cell (Model 118-3, Wilmad Glass, Buena, NJ) placed in a Nicolet 6700 FT spectrometer (Nicolet Analytical Instruments, Madison, WI). The spectrometer utilized a CaF_2 beamsplitter and liquid nitrogen-cooled InSb detector. There was a slight modification to the spectrometer setup described previously in Chapter 4 as shown in Figure 6.1. A 50-W tungsten-halogen light bulb (Gilway Technical Lamp, Peabody, MA) with an integrated, gold-coated reflector was used as the light source and an E3633A 200-watt DC power supply (Agilent Technology, Van Nuys, CA) was used to maintain the light bulb at the desired voltage for the data collection. No screens or neutral density filters were used in the optical path. A K-band optical interference filter (Barr Associates, Westford, MA) was placed in the optical path to limit the spectral range to 5000-4000 cm^{-1}.

The protein films were placed in the optical path using the lens tube and V-shaped holder as shown in Figure 6.1 and described previously in Chapter 5 (Figure 5.3). The V-shaped holder was attached with an adhesive heating pad (Omega Engineering, Inc., Stamford, CT) to maintain the temperature of the protein film near 37 °C and a T-type thermocouple pasted to the holder and associated digital thermocouple meter (Omega Engineering, Inc.) were used to measure the temperature.

The sample cell employed sapphire windows (Meller Optics, Providence, RI) and was configured with a path length of 1.26 mm using Teflon spacers. The temperature of the samples exiting the sample cell was monitored with a T-type thermocouple probe (Omega Engineering, Inc.) inserted into a port in the vinyl tubing. An Omega Model 670

digital meter recorded the temperatures with a precision of ± 0.1 ° C. For the entire study, the temperature range of the flowing liquid was maintained in the range of 36.6-37.3 °C.

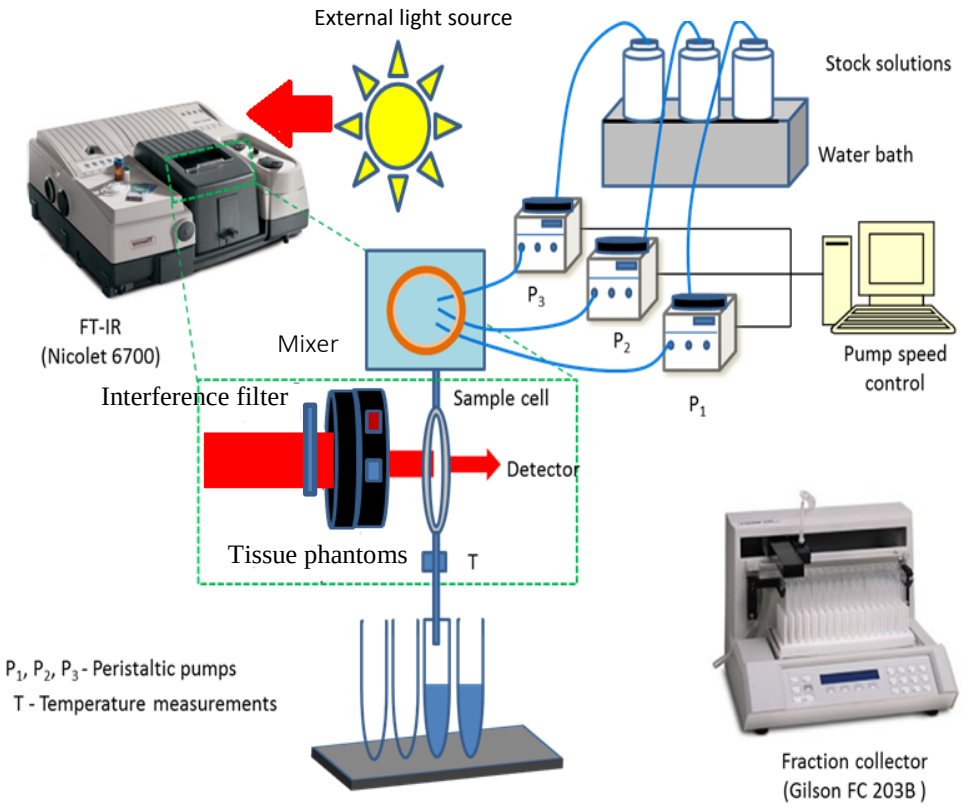


Figure 6.1 Schematic of experimental setup. Three peristaltic pumps and an in-line mixer were used to produce varying concentrations of glucose samples from the stock solutions. Samples flowed into the sample cell contained in the spectrometer, then were collected by the fraction collector. The sample holders containing the skin tissue phantoms were placed in front of the sample cell for the dynamic system. An external light source was used.

After the sample exited the sample cell and passed through additional insulated vinyl tubing (89.2 mm in length), fractions were continuously collected at a rate of 1 min/tube using a Gibson FC 203B fraction collector (Gilson, Inc., Middleton, WI). The glucose concentrations of each of the fractions were verified each day with a YSI Model 2300 STAT PLUS glucose analyzer (YSI Inc., Yellow Springs, OH) which had an instrumental error of ± 0.2 mM according to the YSI product specifications.

Procedures

Spectra for the liquid flowing through the sample cell were collected continuously as 64 co-added (~ 1 min) asymmetric scans consisting of 4097 points. The Fourier processing steps included one level of zero-filling, Happ-Genzel apodization, and Mertz phase correction, producing spectra with a point spacing of 1.93 cm^{-1}. This corresponds to 519 resolution elements over the range of 4000-5000 cm^{-1}.

The software used for the data collection and subsequent Fourier processing was Omnic (Version 7.1, Nicolet Analytical Instruments) operating on a Dell OptiPlex GX280 computer (Dell Computer Corp., Austin, TX) running under Windows 7 (Microsoft, Inc., Redmond, WA). The software recorded the time at the end of each spectrum. As discussed in Chapter 4, the delay time between the spectra collected in the sample cell and the fractions collected in the fraction collector were corrected before the data analysis.

Spectra corresponding to times in which the pumps were changing speeds were omitted from further analysis on the basis of uncertainty in the concentrations assigned to the spectra. The same methodology described in Chapter 4 based on the calculation of

95% confidence intervals about the reference concentrations was employed to identify these spectra.

Computations

After collection and Fourier processing, spectra were transferred from the computer controlling the spectrometer to a Dell Precision 670 workstation (Dell Computer Corp.) running under Red Hat Linux (Version 5.3, Red Hat, Inc., Raleigh, NC). All subsequent calculations were performed on this computer using the Matlab development environment (Version 7.4.0 (R2007a), The MathWorks, Inc., Natick, MA). Software for the calculation of piecewise linear discriminants used in-house software written in Fortran and compiled with the Intel Fortran Compiler for Linux (Version 10.0, Intel Corp., Santa Clara, CA).

Data Collection for Dynamic System 1

As discussed above, data for dynamic system (DS) 1 were collected over multiple days in the presence of a tissue phantom kept in the same orientation to the optical path for all the days. The tissue phantom was composed of a keratin film and a gelatin (hydrolyzed collagen) film prepared earlier. This phantom employed the constant-thickness films described in Chapter 5. In thickness estimates based on fitting the film spectra to *in vivo* spectra of tissue collected from six human subjects, the gelatin and keratin films produced regression coefficients of 0.6 and 1.4, respectively. These regression coefficients mean that the gelatin film should be 1/0.6 = 1.7 times thinner to match the average collagen thickness present in the skin tissue and the keratin film should

be 1.4 times thicker to match the average thickness of keratin present in the skin tissue. Thus, the tissue phantom employed in DS 1 was effectively higher in collagen and lower in keratin than actual human tissue.

The data collected during the first five days of data collection (runs 1-5) were used to build the calibration database and the last three days of data collection were used as prediction days to test the implementation of the alarm algorithm. The concentration values for each component per sample were assigned to minimize correlations between glucose and urea concentrations. For this study, the correlation coefficient between glucose and urea concentrations in the calibration data was -0.27. The correlation coefficients between glucose and urea concentrations in the three prediction sets were -0.46, -0.48 and -0.67, respectively.

Figure 6.2 summarizes the target concentrations for glucose and urea in the calibration data. Figure 6.3 is a similar plot that describes the target concentration profiles for the three prediction sets. For each day, the stock glucose solution concentrations were 10, 20, 10, 20, 10, 10, 10 and 10 mM, respectively. The stock concentrations for urea were 10, 20, 10, 40, 20, 5, 10, and 10 mM, respectively. The complete glucose concentration profile and the corresponding refined profile obtained after removal of spectra corresponding to changes in pump speed are given in Figures 6.4 A and B, respectively.

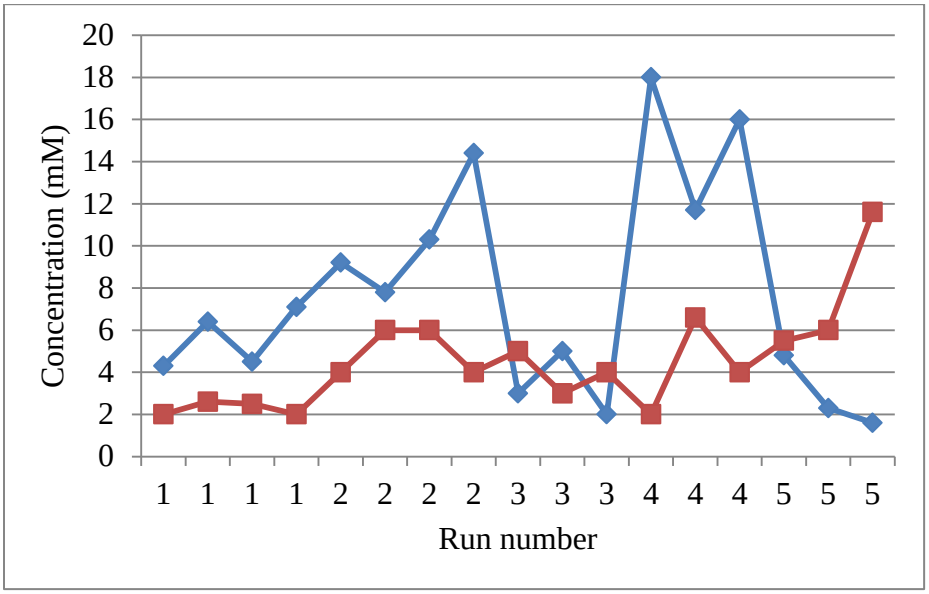


Figure 6.2 Target concentration profiles for the calibration data. Blue and red traces correspond to glucose and urea concentrations, respectively.

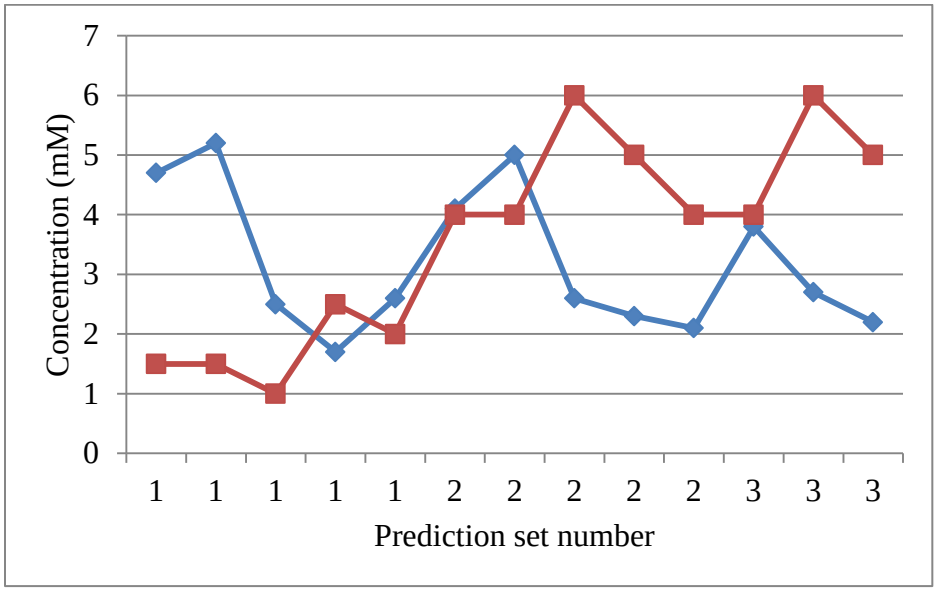


Figure 6.3 Target concentration profiles for the prediction data. Blue and red traces correspond to glucose and urea concentrations, respectively.

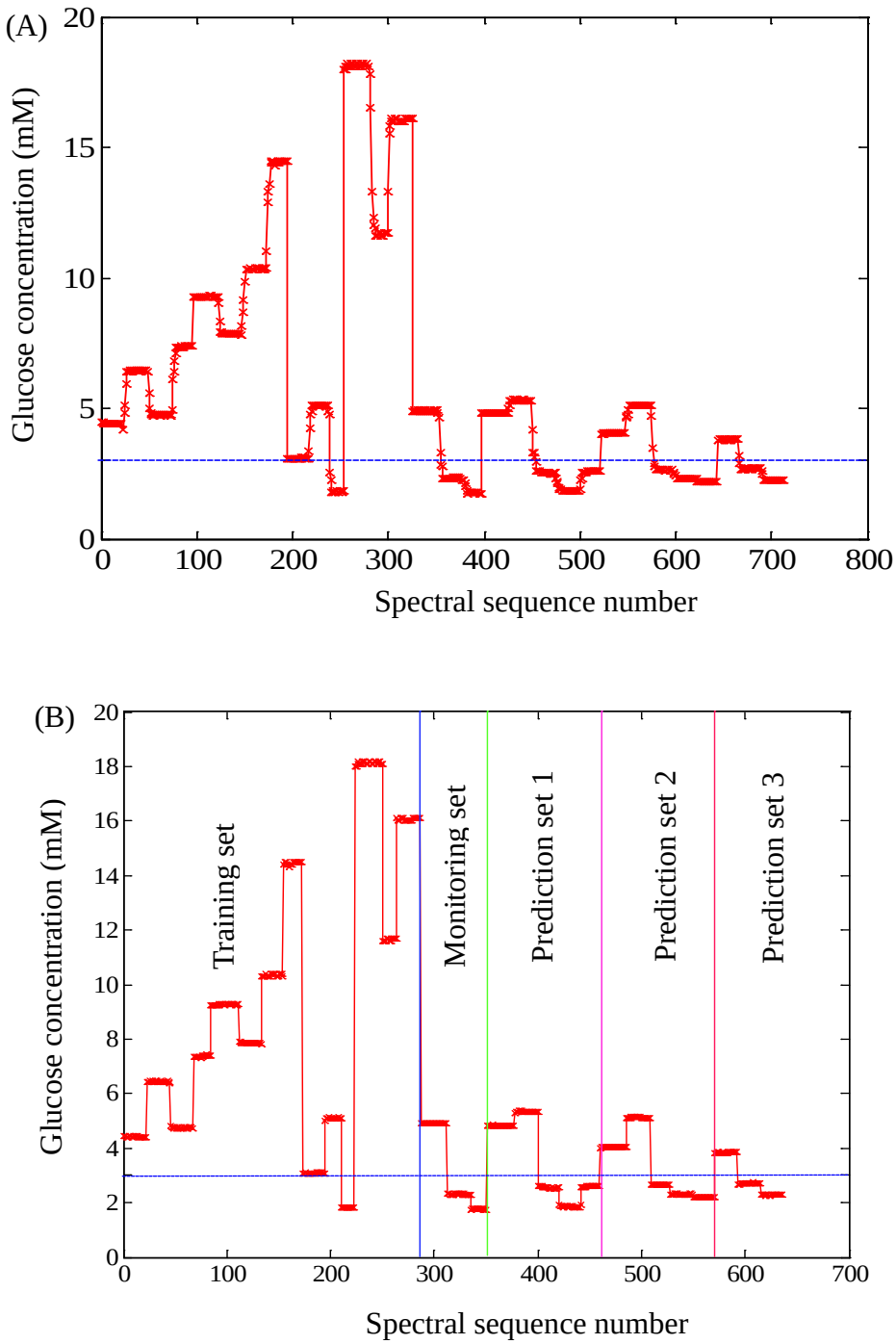


Figure 6.4 Concentration profiles for DS 1. A. Complete glucose concentration profile. B. Refined glucose concentration profile after removal of spectra collected at times when the pump speeds were changing. The labels in panel B denote the subdivision of the data into groups for calibration, calibration testing (monitoring), and external prediction. Horizontal lines in both panels denote the hypoglycemic alarm concentration of 3.0 mM used in this work.

Data Collection for Dynamic System 2

As discussed above, data for DS 2 were collected over multiple days in the presence of a variable-thickness tissue phantom and during each day of data collection the thickness settings were changed randomly by changing the orientations of the gelatin and keratin films. These changes in film thicknesses served to simulate the possible variations in the composition of skin that might occur if a human subject interface of the type shown in Figure 5.4 were used and different locations on the hand were measured.

Compared to DS 1, this study more closely simulated an actual noninvasive data collection through human tissue. As reported in Chapter 5, a linear regression fit with human subject spectra for the absorbance spectra collected from positions 1 and C of the variable-thickness gelatin film yielded average regression coefficients of 1.8 ± 0.3 and 2.0 ± 0.4, respectively. The corresponding absorbance spectra for these two positions were shown previously in Figure 5.9. The differences in absorbance as a result of the differences in film thickness are apparent. As discussed previously with respect to the constant-thickness films, values of the regression coefficients near 2.0 indicate the film was approximately half as thick as the corresponding effective thickness of collagen in tissue.

As reported in Chapter 5, a linear regression fit with human subject spectra for positions 5 and A of the keratin film yielded average regression coefficients of 1.6 ± 0.4 and 3.2 ± 0.6., respectively. The differences in the absorbance for these two positions were shown previously in Figure 5.10. These values indicate the keratin film was also generally thinner than the corresponding effective thickness of keratin in tissue.

Three different positions of the gelatin film, 1, 4, and C, and three different positions of the keratin film, 5, A and E, were used for this study. A 3×3 factorial design for these positions gave 9 total tissue phantom orientations. Using the labels on the film holder (Figure 5.3), the film positions were manually changed during the data collection to simulate variations in skin tissue thickness. Because spectra recorded during changes in pump speed were subsequently removed from the data analysis, the film positions were also changed during the times in which the pump speeds were changed.

Data for the study were collected over four different days that spanned a total of 18 days. Data collection occurred on days 1, 3, 12, and 18. Spectra from day 1 and the first part of day 3 were used to define the calibration database, while the data from the second part of day 3 and all of days 12 and 18 served as prediction sets.

The concentration values for each component per sample were assigned to minimize correlation between glucose and urea concentrations present. The correlation coefficients between glucose and urea concentrations in the calibration set and the three prediction sets were 0.19, 0.02, -0.33 and -0.40, respectively.

Figure 6.5 shows the desired concentrations for glucose and urea plotted with respect to the label of the film used for the particular concentration level. Figure 6.6 displays the corresponding plot for the three prediction sets. For each day, the concentrations of the stock glucose and urea solutions were both 20 mM. The complete glucose concentration profile and the refined glucose profile after removing uncertain concentrations corresponding to changes in pump speed are given in Figure 6.7 A and B, respectively.

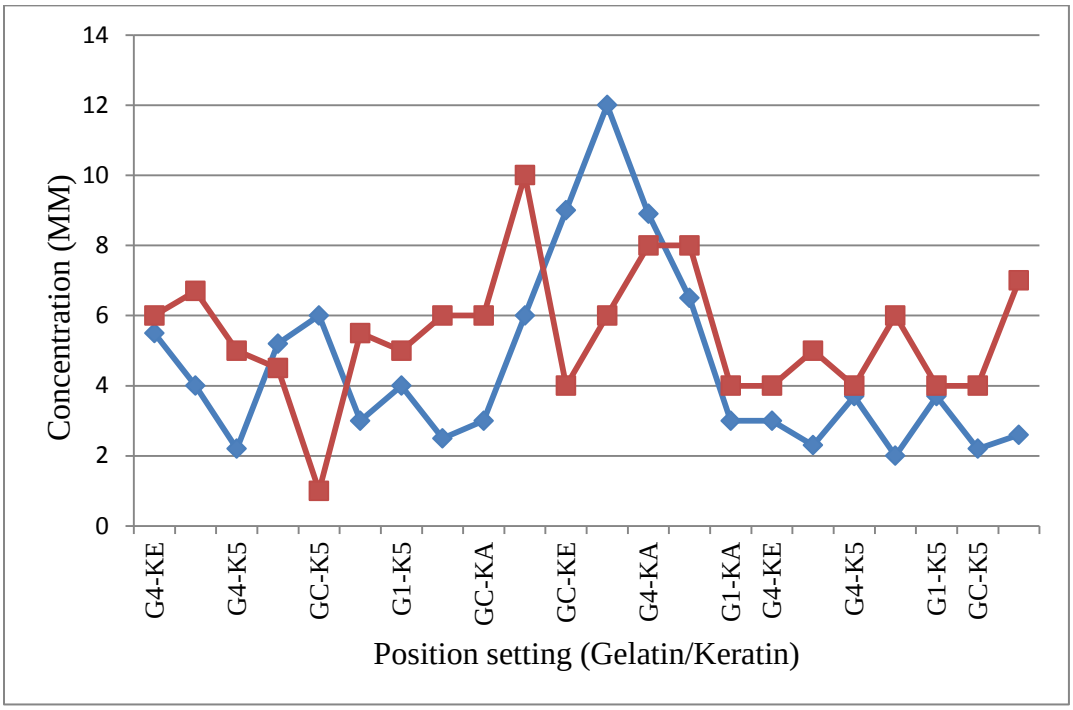


Figure 6.5 Target concentration profiles for the calibration data. Blue and red traces correspond to glucose and urea concentrations, respectively. The *x*-axis specifies the position settings for the gelatin (G) and keratin (K) films.

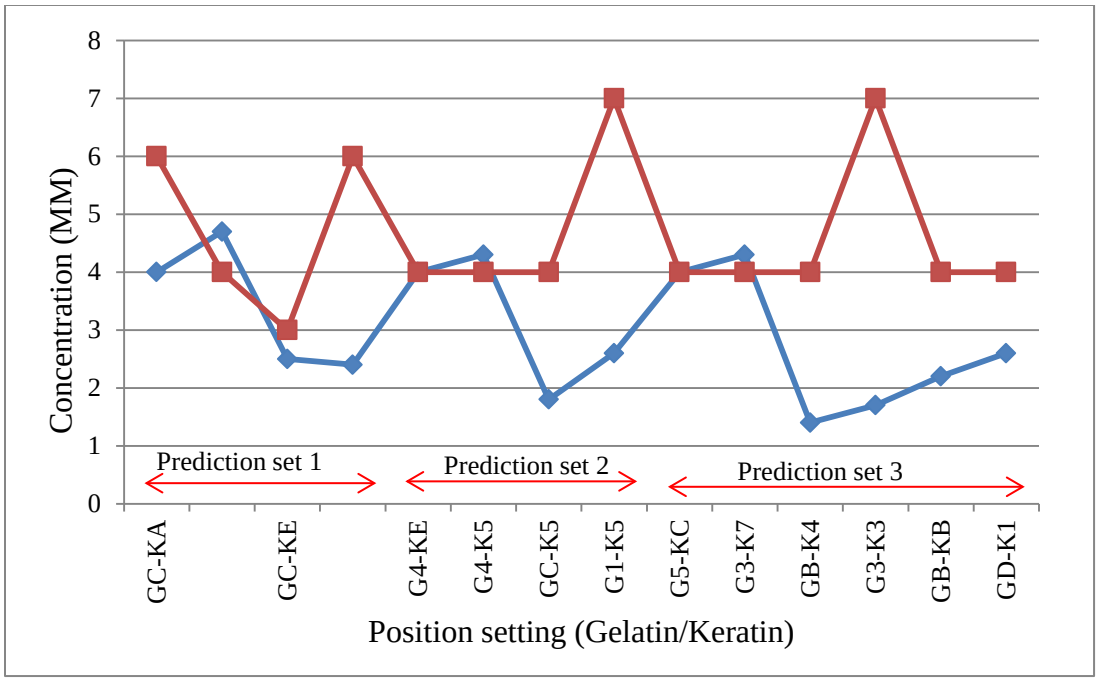


Figure 6.6 Target concentration profiles for the prediction data. Blue and red traces correspond to glucose and urea concentrations, respectively. The *x*-axis specifies the position settings for the gelatin (G) and keratin (K) films.

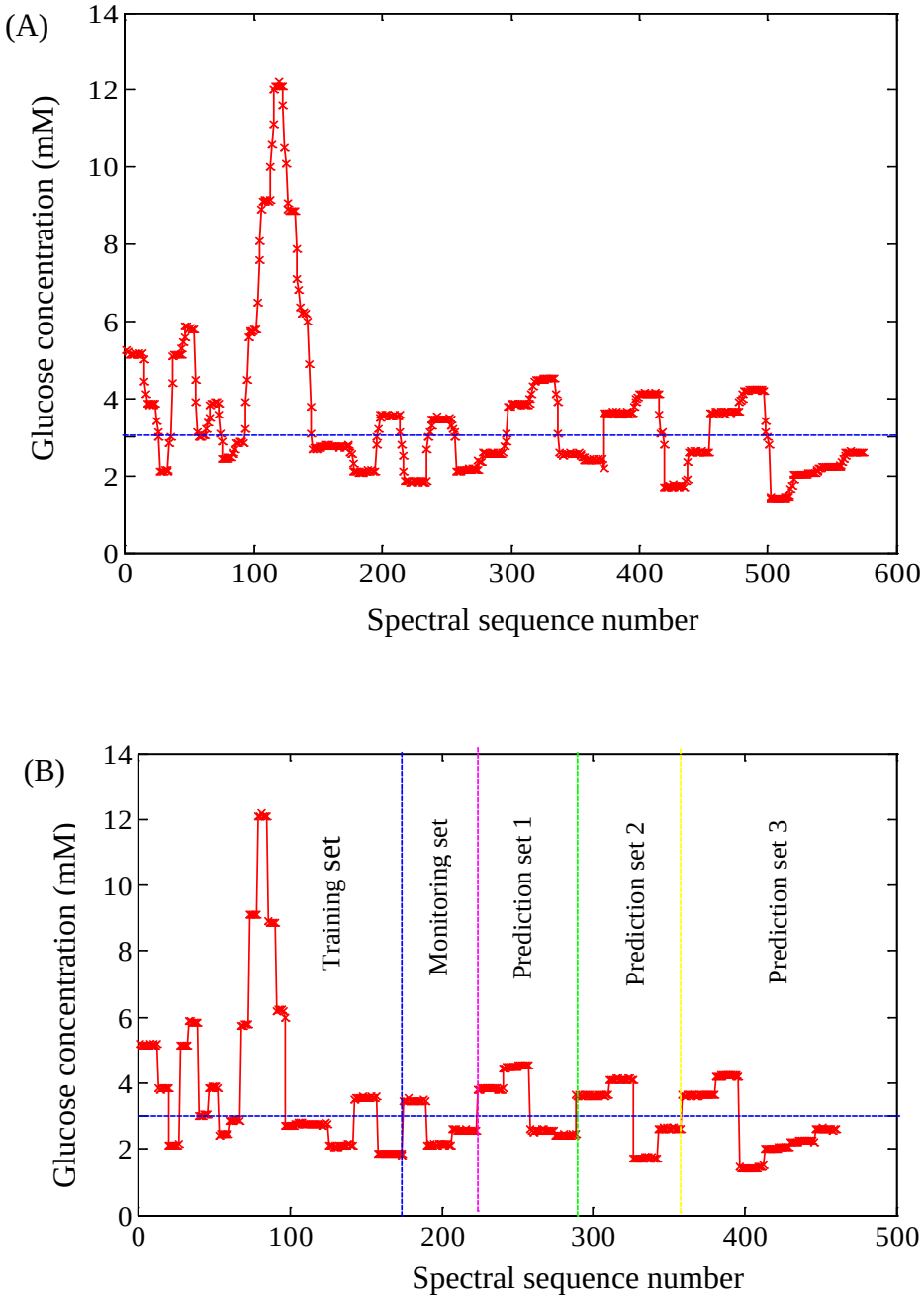


Figure 6.7 Glucose concentration profiles for DS 2. A. Complete glucose concentration profile. B. Refined glucose concentration profile after removal of spectra collected at times when the pump speeds were changing. The labels in panel B denote the subdivision of the data into groups for calibration, calibration testing (monitoring), and external prediction. Horizontal lines in both panels denote the hypoglycemic alarm concentration of 3.0 mM used in this work.

Results and Discussion

The data analysis protocols used with DS 1 and DS 2 mirror those described previously in Chapter 4. The analysis of DS 1 and DS 2 will be described separately below. Data analysis steps are presented below in a more streamlined manner than in Chapter 4 when the procedures were identical.

Analysis of Dynamic System 1

Data for DS 1 in the presence of the constant-thickness tissue phantom were collected over eight different days and consisted of 712 total single-beam spectra. The refined glucose concentration profile consisted of 635 spectra. All the refined spectra for the first five days were used as the calibration set while the data for the last three days were used as prediction sets.

Noise Analysis

For each concentration level, short-term noise was evaluated by computing 100 % lines from each pair of consecutive spectra. These 100 % lines were converted to AU, and the wavenumber region of 4300-4500 cm^{-1} was fitted to a third-order polynomial model. The RMS noise was then computed about the polynomial fit to obtain the intrinsic measurement noise. The RMS noise calculated for the spectra collected over eight different days at 64 scans ranged from 9.1-22.6 μAU. The noise calculation was described previously in Chapter 3. These noise values are higher than the corresponding noise estimates reported in Chapter 4 because of the increased attenuation of the incident source energy caused by the presence of the protein films.

Optimization of Calibration Parameters

The spectra collected on each day were put into separate blocks giving a total of five blocks for the complete calibration data set. Differential spectra were calculated by taking the ratios of all the combinations of single-beam spectra within a block. The PLS scores for these differential spectra were computed and were used as the patterns for the classification model.

For the optimization of wavenumber range and model dimensionality, the calibration set was again partitioned into a training set and a monitoring set. The complete glucose concentration profile which was partitioned into training, monitoring, and prediction sets is shown in Figure 6.4 B. The training data included four days of data collection over a total time span of 10 days (i.e., days 1-10). The monitoring data were collected on day 11. Prediction sets 1, 2, and 3 were collected on days 21, 34, and 46.

The mean single-beam spectra collected on the eight days of data collection are shown in Figure 6.8. As can be seen from the figure, the constant skin tissue phantom background during the data collection is further confirmed based on the similarities in the signal intensities for the eight days of data collection. In addition, it is clear that the light transmission is greatly attenuated at both the high- and low-frequency extremes with the greatest transmission in the center portion of the spectrum. Water is primarily responsible for attenuating the high-frequency and low-frequency radiation. The combination-region spectra possess an absorption feature around 4600 cm^{-1} and a sharp drop in transmitted light around 4400 cm^{-1}.

The training set consisted of 286 single-beam spectra and produced a total of 9774 differential spectra. Spectra with differential concentrations of 0.0 mM were not

used. The grid search used to determine the optimal spectral range and number of PLS latent variables was based on sliding a window of fixed spectral width in 50 cm^{-1} increments across the 4900-4100 cm^{-1} range. At each step, PLS models were constructed using 3-16 latent variables. The spectral widths studied were from 100 to 700 cm^{-1} in 50 cm^{-1}increments.

As described in Chapter 4, a grid search was used to identify the optimal spectral ranges and corresponding number of latent variables for use in building PLS models for differential glucose concentrations. As before, the grid search was based on a cross-validation calculation with spectra withheld in blocks of 10 %. An *F*-test at the 95% level was used to remove latent variables that did not significantly reduce the SECV. Table 6.1 summarizes the four optimal wavenumber ranges and the relevant latent variables obtained with the grid search. Figure 6.9 plots the SECV with respect to the number of latent variables for the 4650-4250 cm^{-1} range that was identified as optimal by the grid search. The top wavenumber ranges include the key glucose absorption features near 4400 and 4300 cm^{-1}. Even though several of the spectral ranges found 16 latent variables as optimal, it was decided not to extend the search range further. As shown in Figure 6.9, the SECV values are changing only very slowly by the time the model size reaches 16.

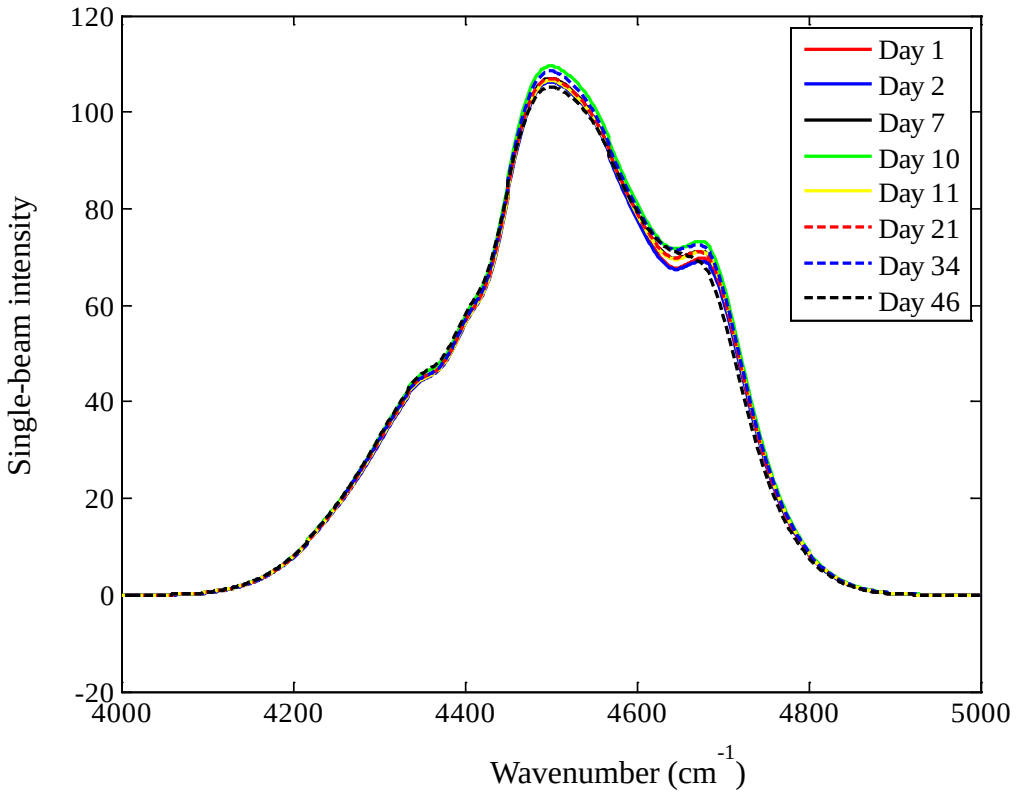


Figure 6.8 Single-beam spectra collected for DS 1 over eight different days in the presence of the constant-thickness skin tissue phantom.

Table 6.1 Results from grid search analysis of DS 1

Wavenumber (cm^{-1})	Latent variables	SECV (mM)
4650-4250	16	0.60_3
4650-4200	15	0.61_6
4650-4150	16	0.64_1
4750-4300	15	0.72_6

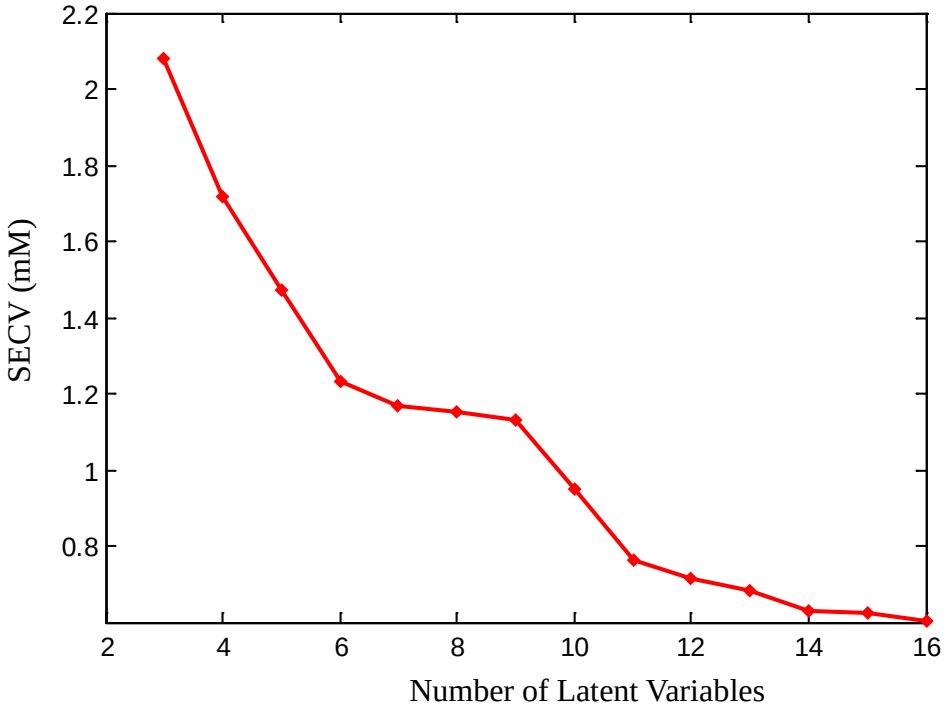


Figure 6.9 Values of SECV with respect to the number of latent variables for the optimal wavenumber range of 4650-4250 cm^{-1}.

Test of Monitoring Data

The grid search analysis was used to identify the best possible wavenumber ranges and latent variables for use in the construction of classification models with PLDA. The performance of the monitoring set was tested with PLDA for the top four wavenumber ranges resulting from the grid search analysis and with latent variables 9-14. This range of latent variables was selected on the basis of preliminary experiments performed before the full study was attempted.

The glucose concentration for the first spectrum in the monitoring set was 4.9 mM. The difference between the alarm threshold concentration, for this study 3.0 mM, and the above mentioned reference concentration gave the critical concentration, a value of -0.9 mM. Partitioning the differential calibration spectra on the basis of this critical concentration led to 4781 and 4993 alarm and non-alarm patterns, respectively.

Table 6.2 summarizes the percentage (average ± standard deviation) of missed and false alarms when three replicate classification models were built with PLDA. Because of the increased complexity of the spectral background, it was found that piecewise linear classifiers based on three discriminant functions were required. This is in contrast to the work described in Chapter 4 with the simpler sample matrix in which only a single discriminant function was required.

The three discriminant functions from each of the replicate classifiers were applied to classify 26 alarm and 38 non-alarm concentrations in the monitoring set. More than one of the wavenumber and latent variable combinations showed the minimum percentage of missed and false alarms. The data separation with nine latent variables was low and the second and third discriminants also did not meet the single-sided requirement

discussed in Chapter 3. Because of these reasons, the wavenumber range of 4750-4300 cm^{-1} and 10 latent variables were used as the optimal combination for the data prediction. This combination showed no missed or false alarms for the monitoring data set.

The selection of 10 latent variables was based on the criterion defined previously of choosing the minimum model dimensionality that achieved equivalent performance to the best results. In this case, an argument could be made for a higher number of latent variables on the basis of an increased number of patterns separated in the training set (e.g., see the results in Table 6.2 for the 4750-4300 cm^{-1} range with 14 latent variables). However, if the performance of the classification model in prediction is considered, adding four additional latent variables (i.e., going from 10 to 14) does not improve performance.

The glucose concentration profile for the monitoring set is given in Figure 6.10. The discriminant scores plot for the monitoring set for the optimal wavenumber and latent variable combination is given in Figure 6.11.

The wavenumber range of 4750-4300 cm^{-1} and 10 latent variables was chosen as the optimal combination for use with the three prediction sets. The training set and the monitoring set were combined to form an overall calibration database. A total of 11,655 differential spectra were computed from these data. As before, differential concentrations of 0.0 mM were not included. The region of 4750 to 4300 cm^{-1} in the differential spectra was decomposed into 10 PLS spectral loadings and loading weights. The corresponding PLS scores defined the pool of patterns used to implement the alarm algorithm with the three predictions sets.

Table 6.2 Average percentage of missed and false alarms for the monitoring set

Spectral range (cm^{-1})		Latent variables					
		9	10	11	12	13	14
4650-4250	A.F (%)[a] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.M (%)[b] ± S.D	68.4 ± 31.6	64.9 ± 24.5	24.6 ± 28.9	2.6 ± 4.6	7.0 ± 6.6	64.0 ± 34.8
	D.S (%)[c] ± S.D	49.1 ± 2.0	51.6 ± 0.5	56.3 ± 1.4	62.9 ± .4	64.5 ± 0.7	71.4 ± 0.7
4650-4200	A.F (%)[a] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.M (%)[b] ± S.D	52.6 ± 9.5	43.9 ± 17.9	20.2 ± 14.5	21.1 ± 4.6	1.8 ± 1.5	14.9 ± 14.5
	D.S (%)[c] ± S.D	40.0 ± 5.0	60.0 ± 5.6	55.3 ± 3.1	62.0 ± 5.6	77.8 ± 1.3	79.3 ± 1.9
4650-4150	A.F (%)[a] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.M (%)[b] ± S.D	80.7 ± 8.5	58.8 ± 21.3	35.9 ± 16.8	16.7 ± 15.4	27.2 ± 38.0	57.0 ± 5.5
	D.S (%)[c] ± S.D	41.8 ± 1.2	46.0 ± 5.7	55.6 ± 1.1	54.7 ± 4.6	57.7 ± 7.4	58.9 ± 8.1
4750-4300	A.F (%)[a] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.M (%)[b] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	5.3 ± 5.3	2.6 ± 2.6
	D.S (%)[c] ± S.D	57.9 ± 1.2	63.5 ± 1.3	64.2 ± 1	70.8 ± 1.3	74.6 ± 2.5	82.7 ± 3.4

[a]Average percentage of false alarms (AF) ± standard deviation.

[b]Average percentage of missed alarms (AM) ± standard deviation.

[c]Average percentage of separated alarm patterns with three discriminant functions (DS) ± standard deviation.

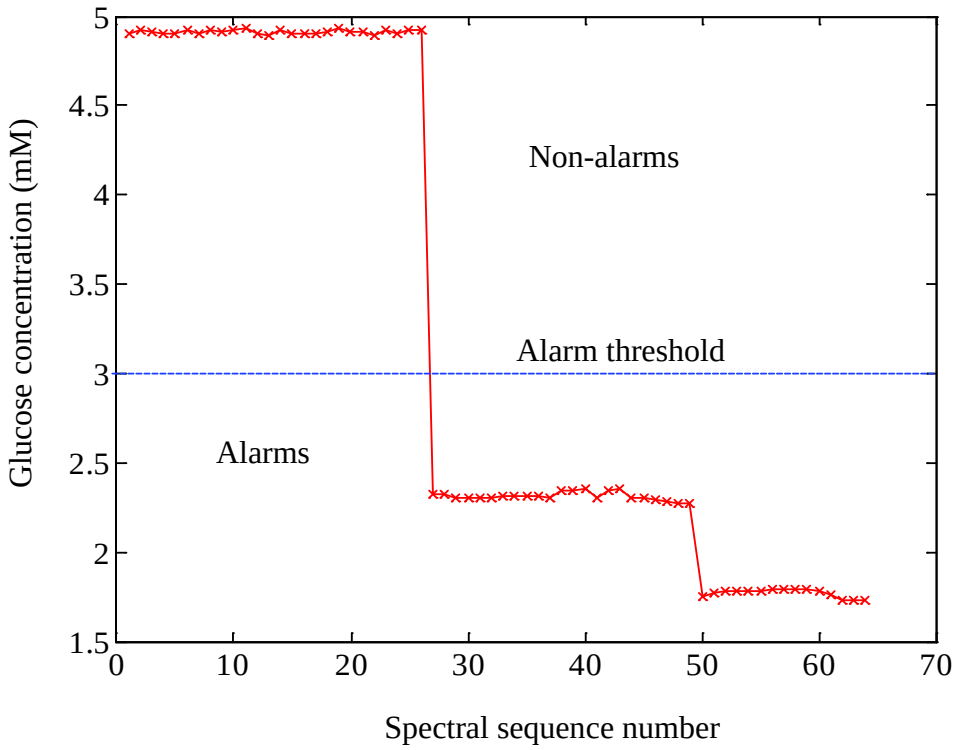


Figure 6.10 Glucose concentration profile for the monitoring set. The alarm threshold value was 3.0 mM. There were 26 and 38 non-alarm and alarm patterns in the monitoring set.

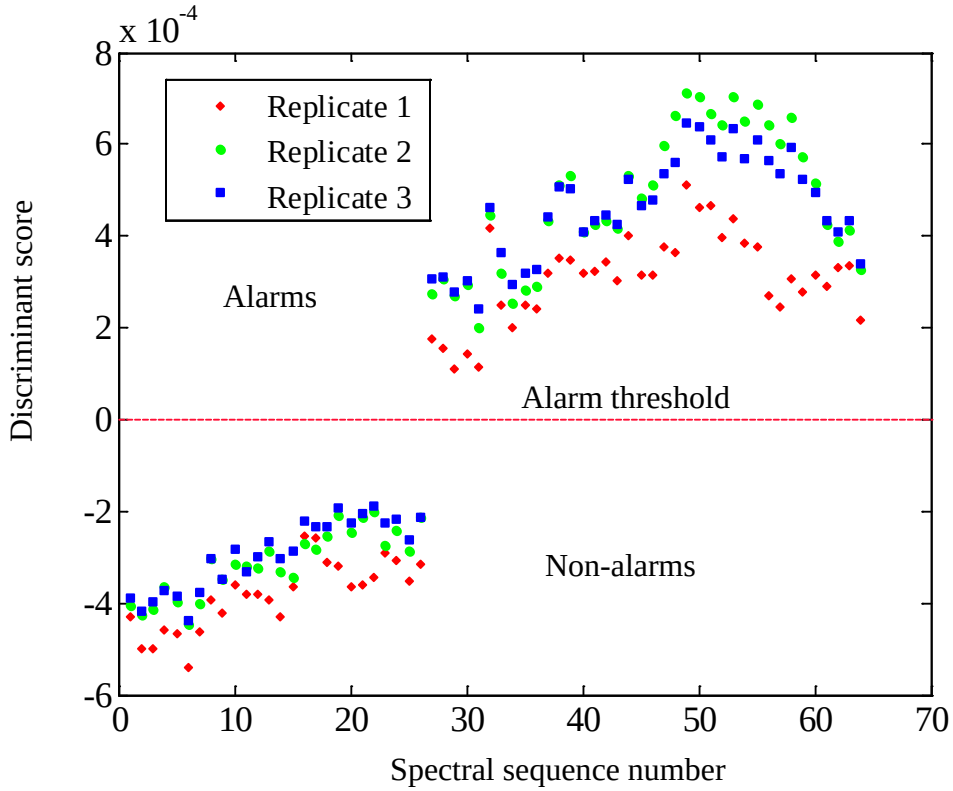


Figure 6.11 Discriminant score plot for the monitoring set. Discriminant scores are plotted for the three replicate classifiers. No missed or false alarms were observed.

Analysis of Prediction Sets

The data for the first prediction set (PS) were collected 10 days after the last calibration data collection day. The reference concentration for PS 1 was 4.8 mM and the critical concentration was -1.8 mM. The calibration PLS patterns were partitioned based on the critical concentration into 5795 alarm patterns and 5860 non-alarm patterns.

Three replicate classifiers were computed with the calibration PLS patterns and the first three discriminants from each of the three replicates were applied for the alarm classification. As discussed in Chapter 4, the decision rule used was that a pattern was placed in the alarm class if two of the three replicate classifiers produced a positive discriminant score (i.e., classified it as an alarm pattern).

Figure 6.12 shows the glucose concentration profile for PS 1 in which there were 59 alarm patterns and 50 non-alarm patterns. These alarm and non-alarm patterns were classified 100% correctly using the calibration discriminants giving no missed or false alarms. The corresponding discriminant score plot is shown in Figure 6.13. The developed alarm algorithm was able to classify the alarm and non-alarm patterns in PS 1 very effectively, even in the presence of the skin tissue phantom and the interfering component, urea.

The 4^{th}, 5^{th} and 6^{th} PLS scores for the prediction set were plotted together as shown in Figure 6.14 and clearly demonstrate the data separation between alarm and non-alarm patterns. The first four PLS scores of the prediction set were plotted together with the corresponding maximum and minimum calibration PLS scores as shown in Figure 6.15 A, B, C, and D. It is very clear that the prediction PLS scores are within the space defined by the calibration PLS scores, even though the prediction data were collected 10

days after the last calibration day. These plots further confirm the excellent data prediction.

Data for PS 2 were collected 23 days after the last calibration day. The critical concentration for PS 2 was -1.0 mM. The data partitioning based on this critical concentration produced 7706 alarm patterns and 3949 non-alarm patterns.

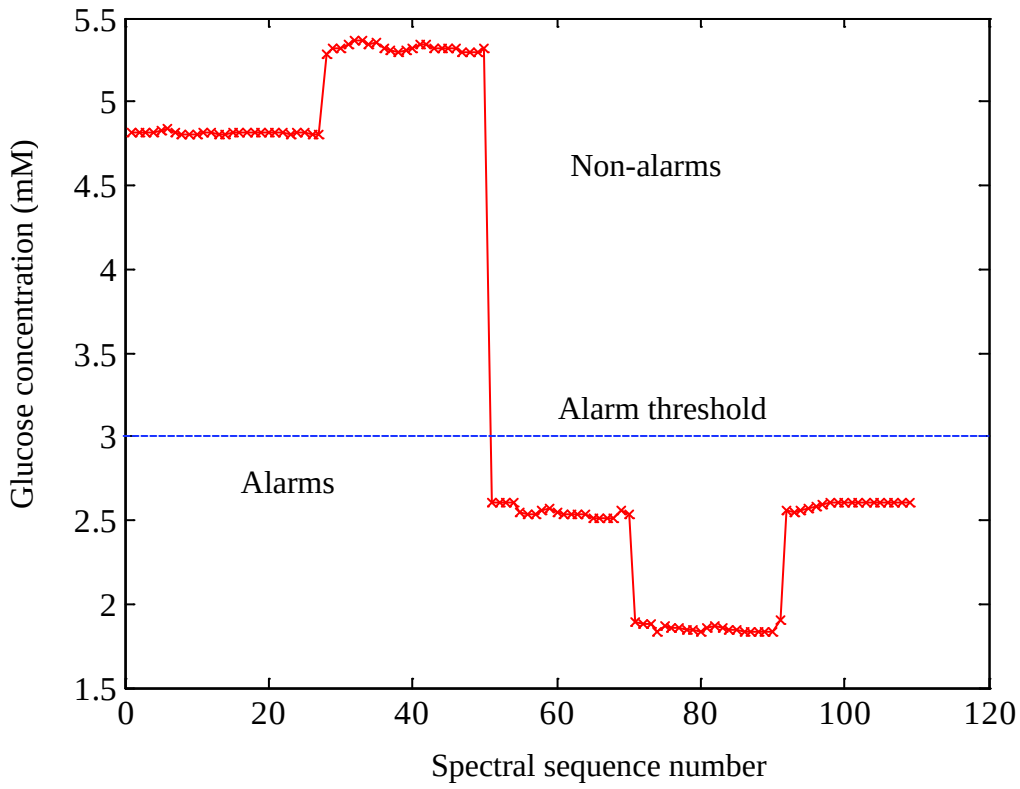


Figure 6.12 Glucose concentration profile for PS 1. The alarm threshold value was 3.0 mM. There were 59 alarm and 50 non-alarm patterns in this data set.

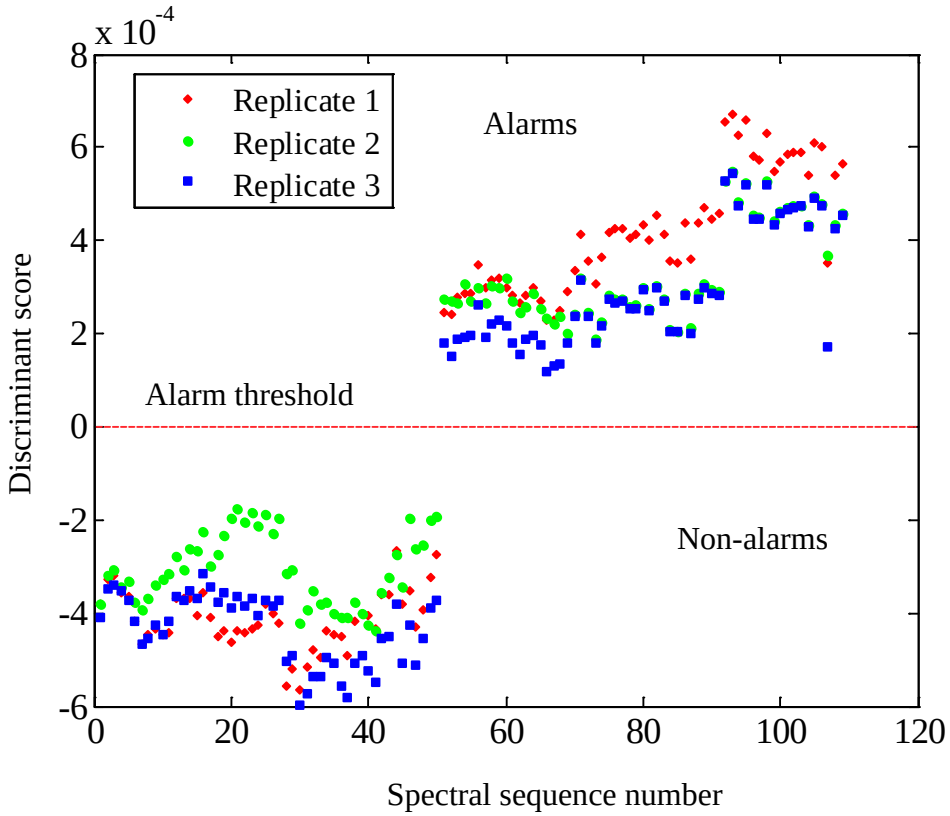


Figure 6.13 Discriminant score plot for PS 1. Results from the three replicate classifiers are shown. No missed or false alarms were observed.

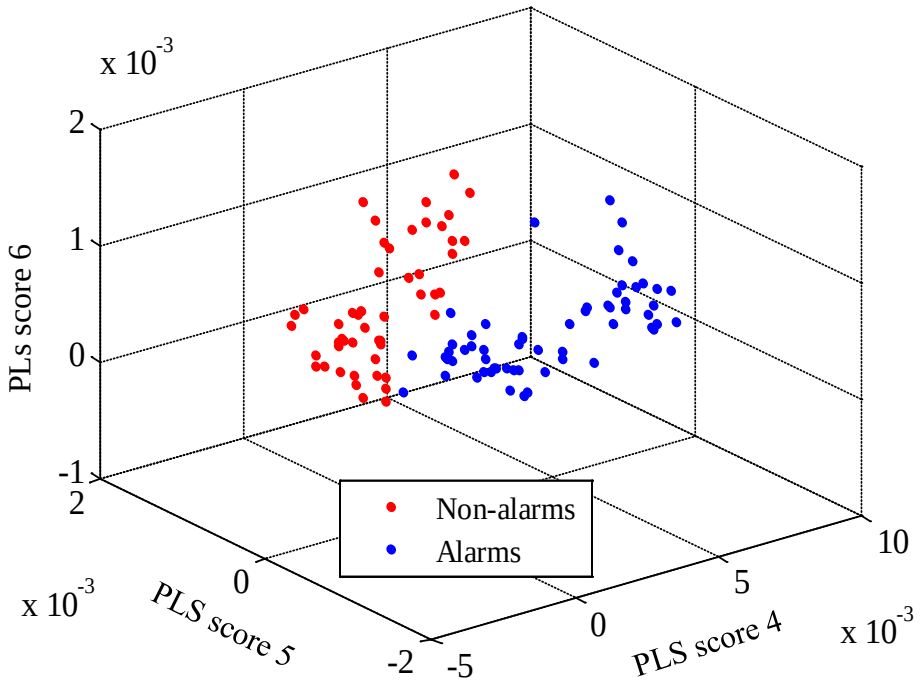


Figure 6.14 The 4th, 5th and 6th PLS scores plotted together for PS 1. Clear separation between the alarm and non-alarm patterns is observed.

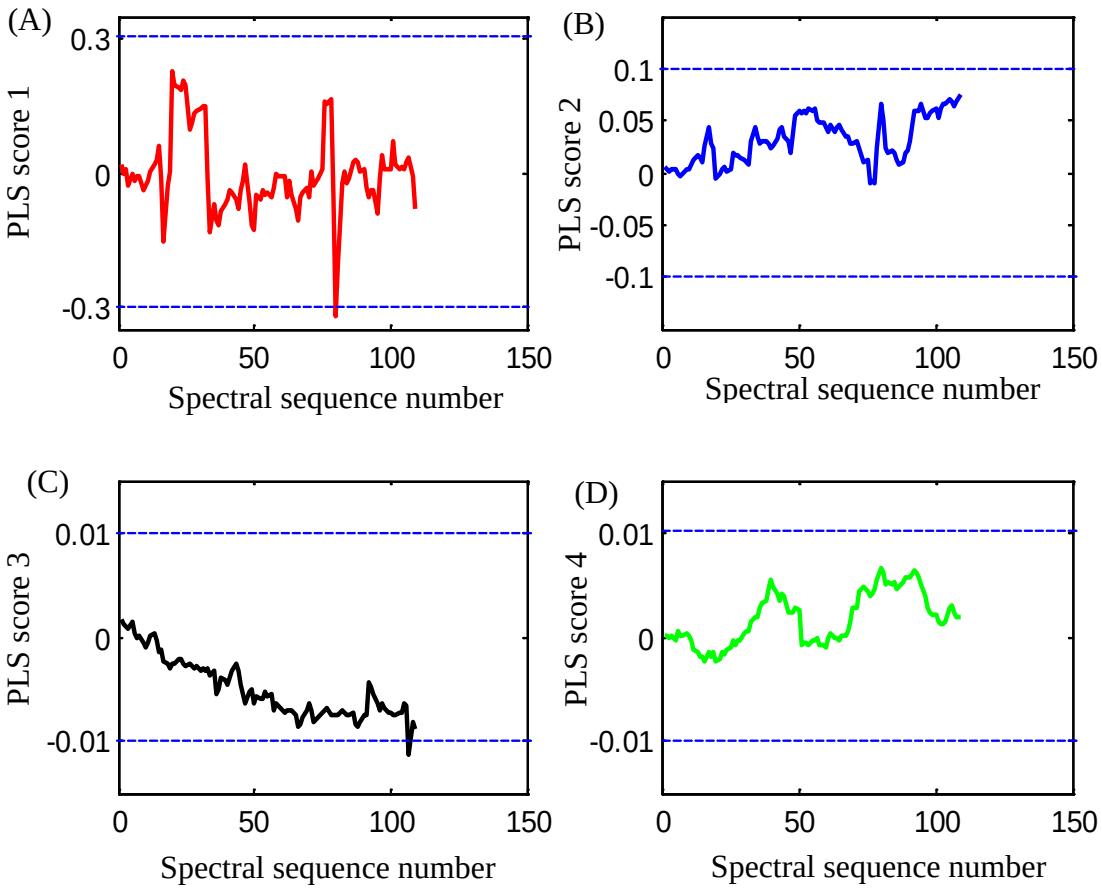


Figure 6.15 Plot of PLS scores with respect to spectral sequence number for PS 1. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as the dashed lines. Panels A, B, C, and D correspond to the scores along latent variables 1, 2, 3, and 4, respectively.

Figure 6.16 shows the glucose concentration profile for PS 2 in which there were 49 non-alarm patterns and 62 alarm patterns. The resulting discriminant score plot is shown in Figure 6.17. There were no missed or false alarms using the first replicate classifier for the data prediction. There were no missed alarms and 1 false alarm using the second replicate classifier. The third replicate classifier predicted no missed alarms and 12 false alarms. However, applying the alarm decision rule described previously, there were no missed alarms and one false alarm for the three replicate classifiers employed.

The third, fourth and fifth PLS scores for PS 2 were plotted together to see the differences in the regions in the data space occupied by the alarm and non-alarm patterns. The three PLS scores plotted together in Figure 6.18 clearly demonstrate the different regions for the alarm and non-alarm patterns. The first four PLS scores of PS 2 were plotted together with the corresponding maximum and minimum calibration PLS scores as shown in Figure 6.19. It is clear that the prediction PLS scores are within the space defined by the calibration PLS scores, even after a time lag of 23 days. These plots further confirm the excellent prediction results.

The data for PS 3 was collected 35 days after the last calibration day. The critical concentration was -0.8 mM, which partitioned the calibration database into 8139 alarm patterns and 3516 non-alarm patterns. Figure 6.20 shows the glucose concentration profile for PS 3 in which there were 22 non-alarms and 43 alarms.

There were no missed and one false alarm using the first replicate classifier for the data prediction. There were no missed alarms and three false alarms using the second replicate classifier. The third replicate classifier predicted no missed alarms and three false alarms. However, applying the alarm decision rule described previously, there were

no missed alarms and three false alarms for the three replicate classifiers employed. The discriminant score plot is shown in Figure 6.21.

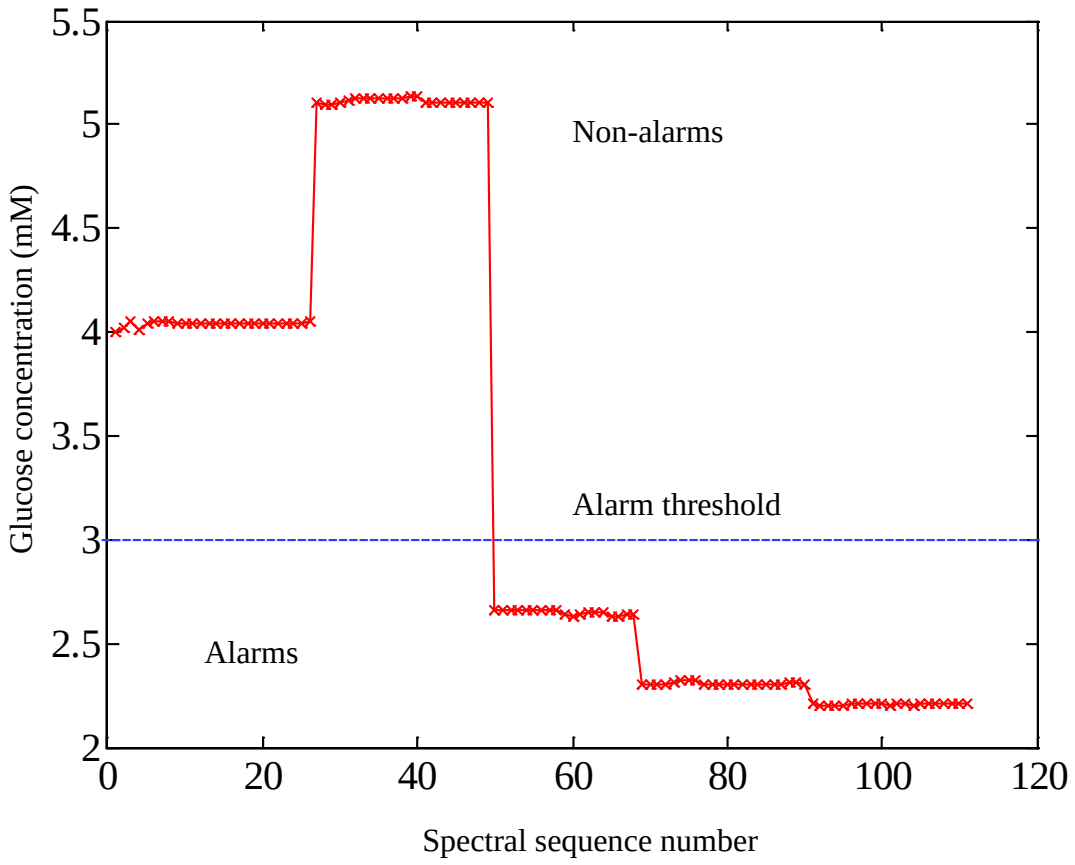


Figure 6.16 Glucose concentration profile for PS 2. The alarm threshold value was 3.0 mM. There were 62 alarm and 49 non-alarm patterns in this data set.

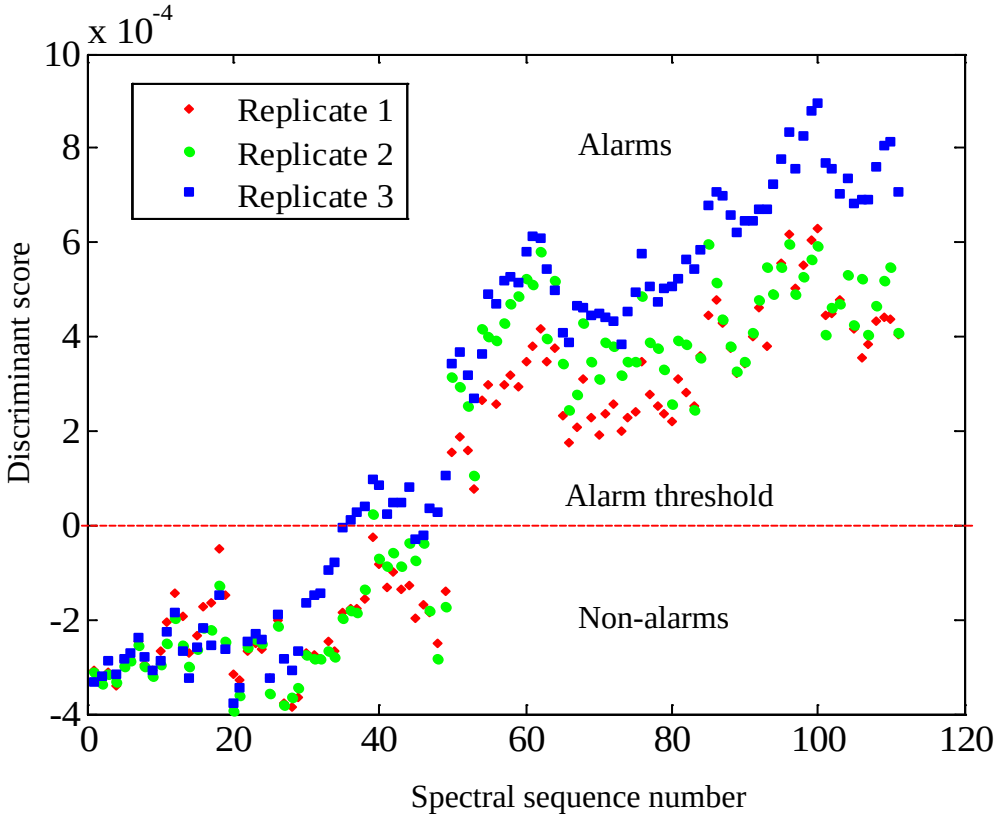


Figure 6.17 Discriminant score plot for PS 2. Results from the three replicate classifiers are shown. No missed alarms and one false alarm were observed when the classification rule was applied based on two out of three classifiers placing the pattern in the alarm class.

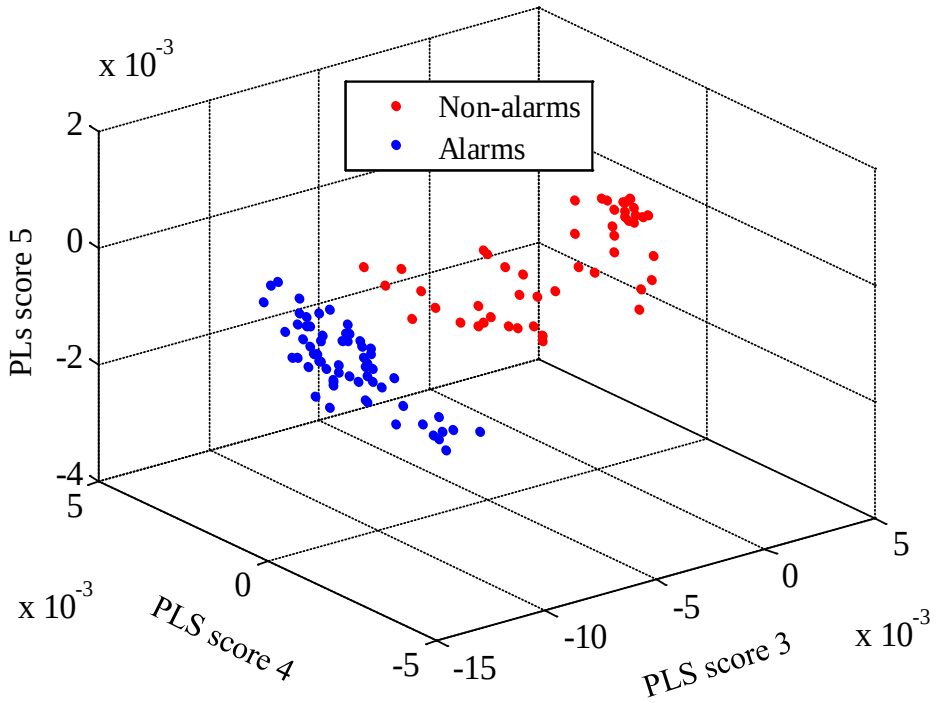


Figure 6.18 The 3rd, 4th and 5th PLS scores plotted together for PS 2. Clear separation between the alarm and non-alarm patterns is observed.

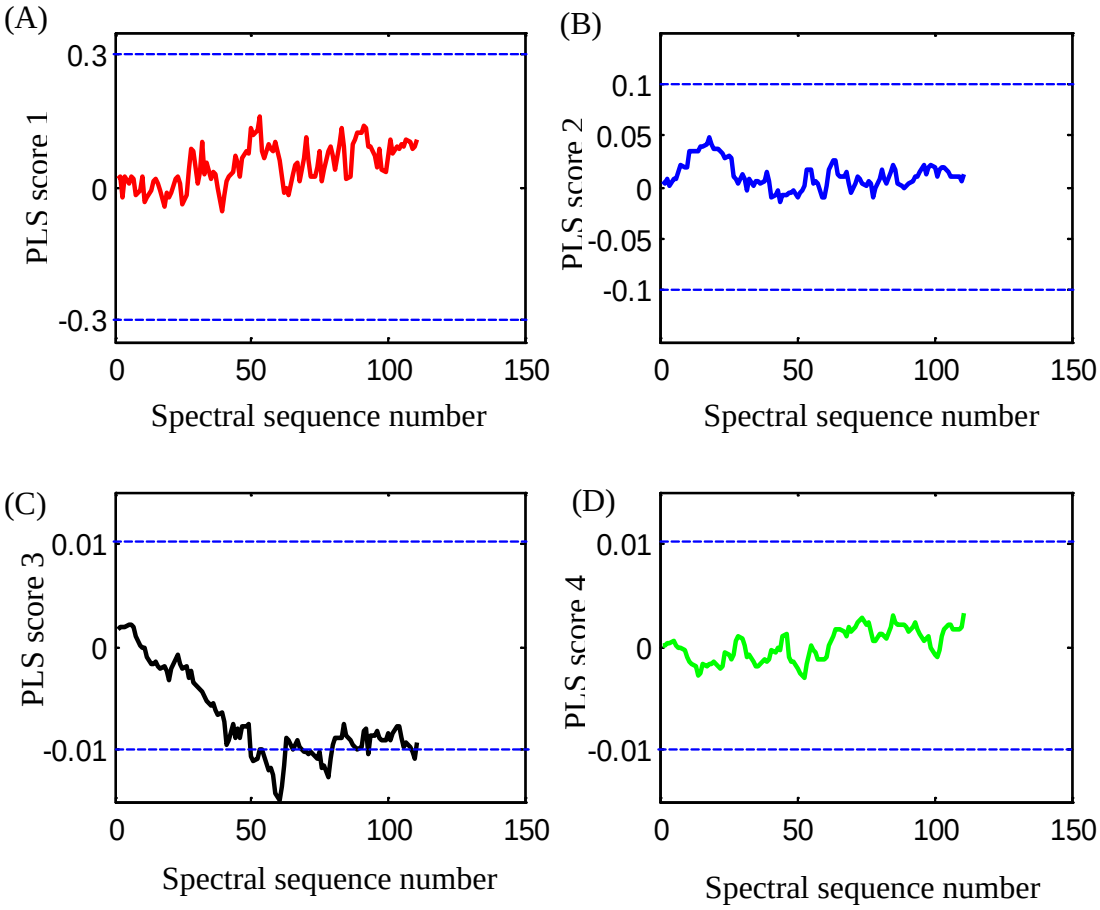


Figure 6.19 Plot of PLS scores with respect to spectral sequence number for PS 2. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as the dashed lines. Panels A, B, C, and D correspond to the scores along latent variables 1, 2, 3, and 4, respectively.

The second, third and fourth PLS scores for PS 3 were plotted together to see the differences in the regions of the data space occupied by the alarm and non-alarm patterns. The three PLS scores plotted together in Figure 6.22 clearly show the different regions for the alarm and non-alarm patterns. The first four PLS scores of PS 3 were plotted together with the corresponding maximum and minimum calibration PLS scores as shown in Figure 6.23. It is clear that the prediction PLS scores are within the space defined by the calibration PLS scores even after a lag of 35 days. These plots further confirm the excellent prediction results.

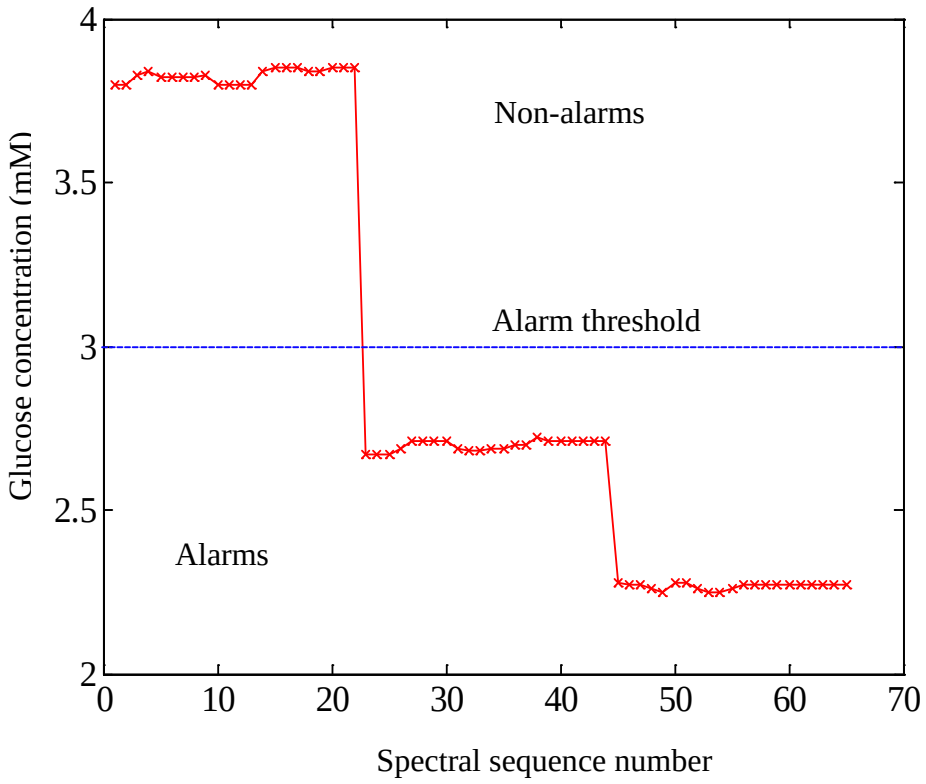


Figure 6.20 Glucose concentration profile for PS 3. The alarm threshold value was 3.0 mM. There were 43 alarm and 22 non-alarm patterns in this data set.

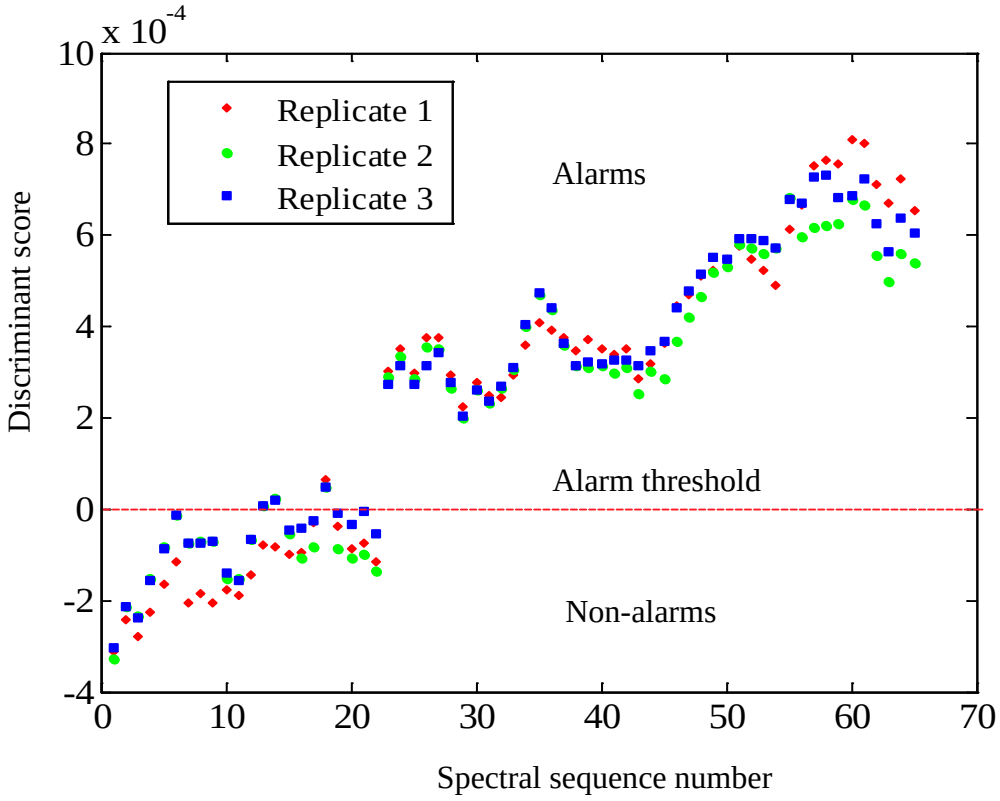


Figure 6.21 Discriminant score plot for PS 3. Results from the three replicate classifiers are shown. No missed alarms and three false alarms were observed when the classification rule was applied based on two out of three classifiers placing the pattern in the alarm class.

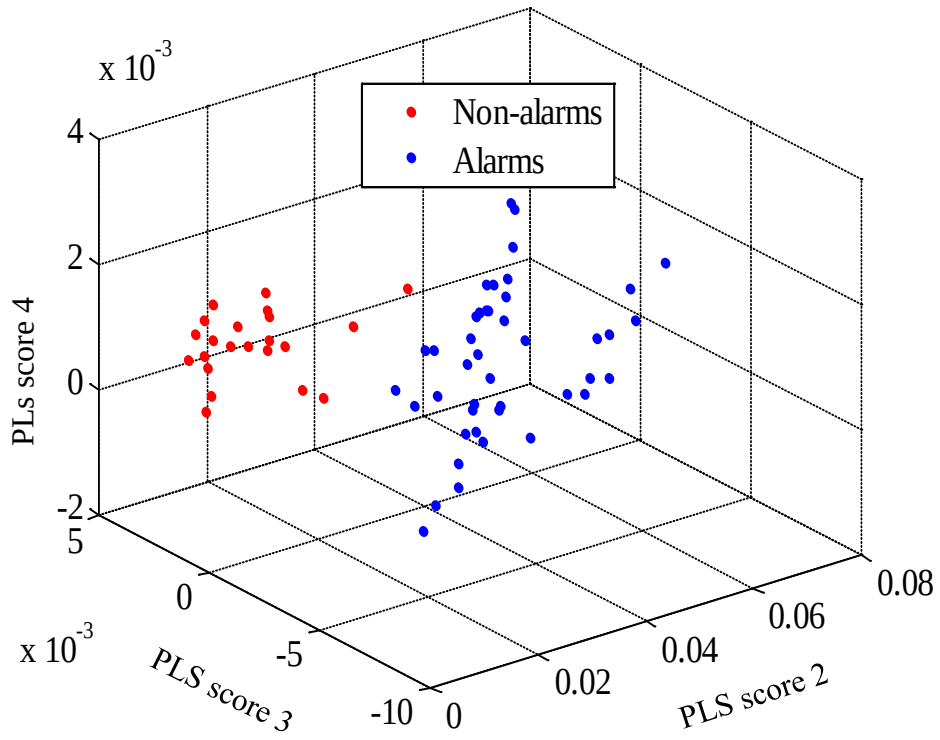


Figure 6.22 The 2nd, 3rd and 4th PLS scores plotted together for PS 3. Clear separation is noted between the data classes.

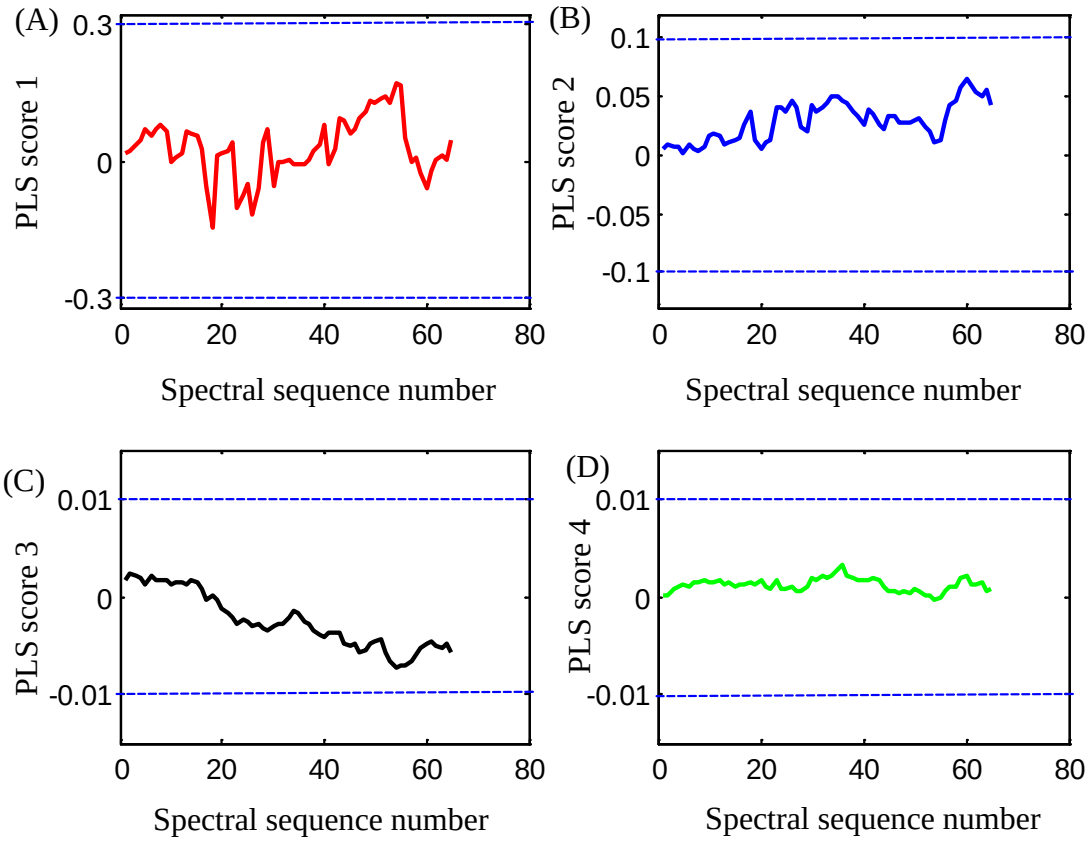


Figure 6.23 Plot of PLS scores with respect to spectral sequence number for PS 3. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as the dashed lines. Panels A, B, C, and D correspond to the scores along latent variables 1, 2, 3, and 4, respectively.

Analysis of Dynamic System 2

Data for the second dynamic study were collected in four sessions that spanned 18 days. The single-beam spectra collected over these four days are shown in Figure 6.24. As can be seen from the figure, the tissue phantom thickness variations during the data collection induce differences in the signal intensities. Compared to Figure 6.8, the absorbance feature around 4600 cm^{-1} and a sharp drop in transmitted light around 4400 cm^{-1} are not as intense since the films used for this study were comparatively thinner. Figure 6.25 is a principal component score plot generated from the single-beam spectra. Clusters of points in the principal component space correspond to the different combinations of gelatin and keratin film positions. Depending on the film orientation used for the spectral collection, the PC scores show different clustering, too.

Noise Analysis

A total of 574 single-beam spectra were collected over the four days. After removing spectra corresponding to changes in pump speed, 460 spectra remained. All the refined spectra for the first day and the first part of the second day were used as the calibration set while the data for the last part of the second day and all the data in the third and fourth days were used as the prediction sets. For each concentration level, the short term noise was evaluated as described previously. The RMS noise calculated for consecutive pairs of spectra ranged from 6.3- 22.3 μAU.

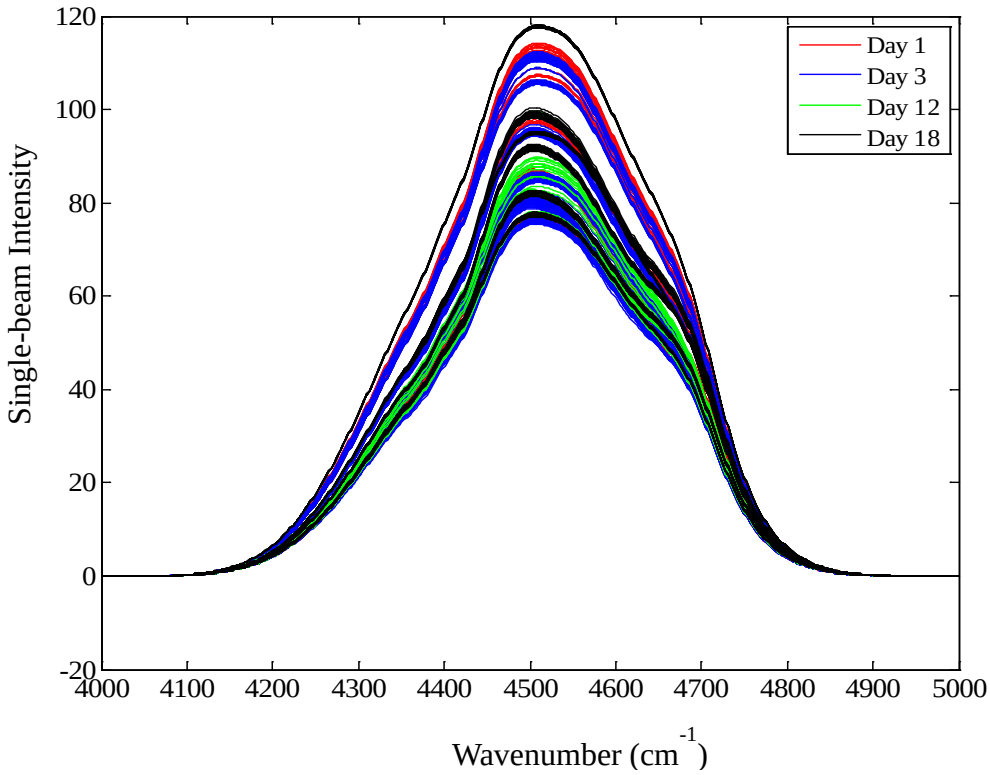


Figure 6.24 Single-beam spectra collected for DS 2. The skin tissue phantom had different orientations during the data collection.

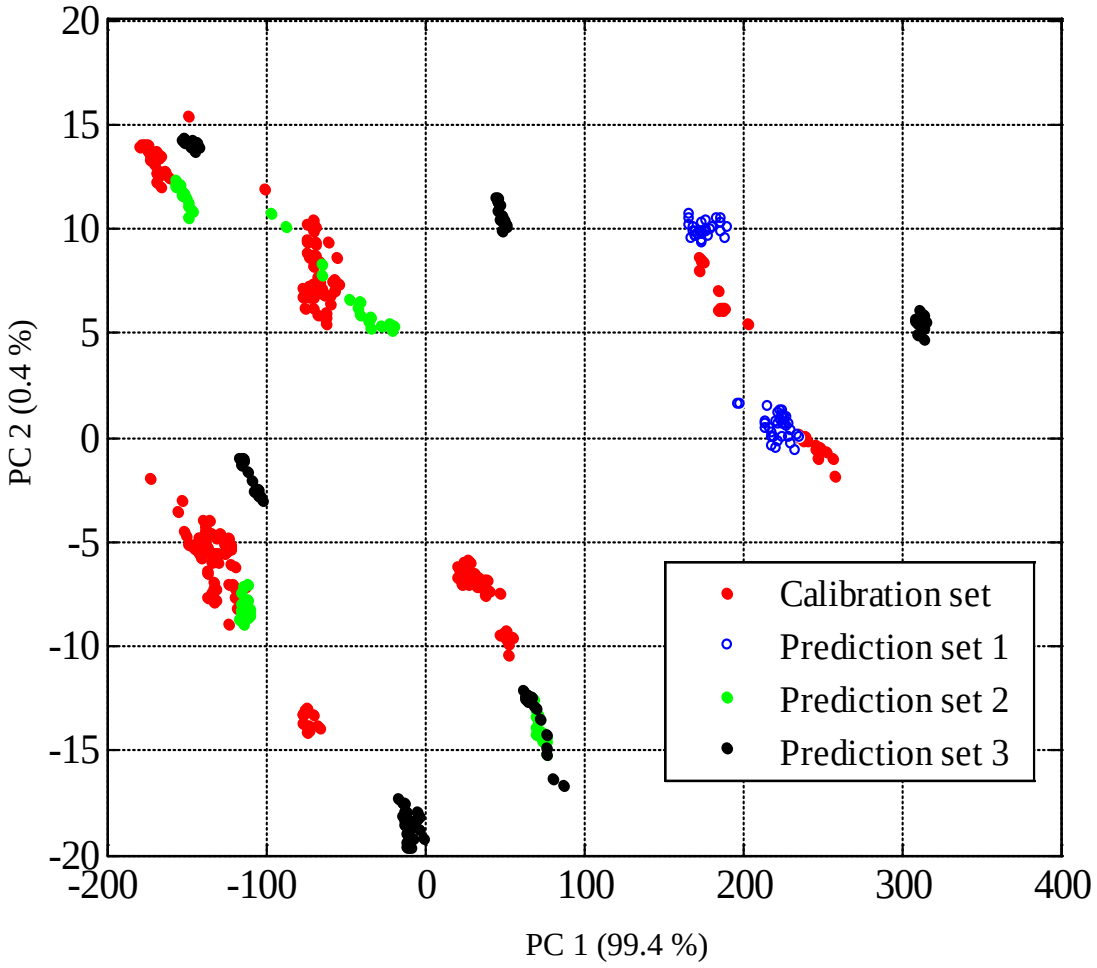


Figure 6.25 Principal component score plot computed from the single-beam spectra in DS2. Spectra were mean-centered before the calculation and the 4200 to 4900 cm^{-1} range was used. The first two principal components account for 99.8 % of the data variance. Clusters in the figure correspond to the different combinations of the gelatin and keratin films.

Optimization of Calibration Parameters

The spectra collected in each day were put into different blocks giving a total of two blocks for the calibration set. As before, differential spectra were calculated by taking the ratios of all the combinations of single-beam spectra within each block. The PLS scores for these differential spectra were computed and were used as the patterns for the classification model.

For the optimization of wavenumber and model dimensionality, the calibration set was again partitioned into a training set and a monitoring set. The complete glucose concentration profile which was partitioned into training, monitoring, and three prediction sets is shown in Figure 6.7 B. The training set consisted of 174 single-beam spectra collected over two sessions. The calculation of differential spectra produced 7548 spectra after removing spectra with differential concentrations of 0.0 mM.

The same procedures described previously were used with the calibration data to identify optimal combinations of the spectral range and number of PLS latent variables. Cross-validation results based on computed SECV values were again used as the basis for choosing the best model parameters. Table 6.3 summarizes the five optimal wavenumber ranges and the corresponding numbers of latent variables. The top wavenumber ranges encompass the important glucose C-H combination band at 4400 cm^{-1}.

The grid search analysis was used to identify the best wavenumber ranges and latent variables for use in the construction of classification models with PLDA. Figure 6.26 plots the SECV with respect to the number of latent variables for the 4650-4250 cm^{-1} range that was found to be optimal in the grid search. Comparison of this plot with Figure 6.9 reveals the greater difficulty of modeling the glucose concentration in DS 2

relative to DS 1. Possible extension of the grid search past 16 latent variables is also suggested by this plot.

Test of Monitoring Data

The performance of the monitoring set with PLDA was then tested with the five wavenumber ranges listed in Table 6.3. The work with PLDA examined each spectral range over 12-16 latent variables. This selection of the range of latent variables was made on the basis of preliminary trials with the data from DS 2.

The glucose concentration for the first spectrum in the monitoring set was 3.4 mM. This produced a critical concentration of -0.4 mM and led to 5966 and 1582 differential spectra, respectively, in the alarm and non-alarm data classes.

Table 6.3 Results from grid search analysis of DS 2

Wavenumber (cm^{-1})	Latent variables	SECV (mM)
4650-4300	16	0.88_2
4700-4300	16	0.90_2
4700-4250	16	0.95_3
4750-4250	16	0.99_5
4800-4250	16	1.03_6

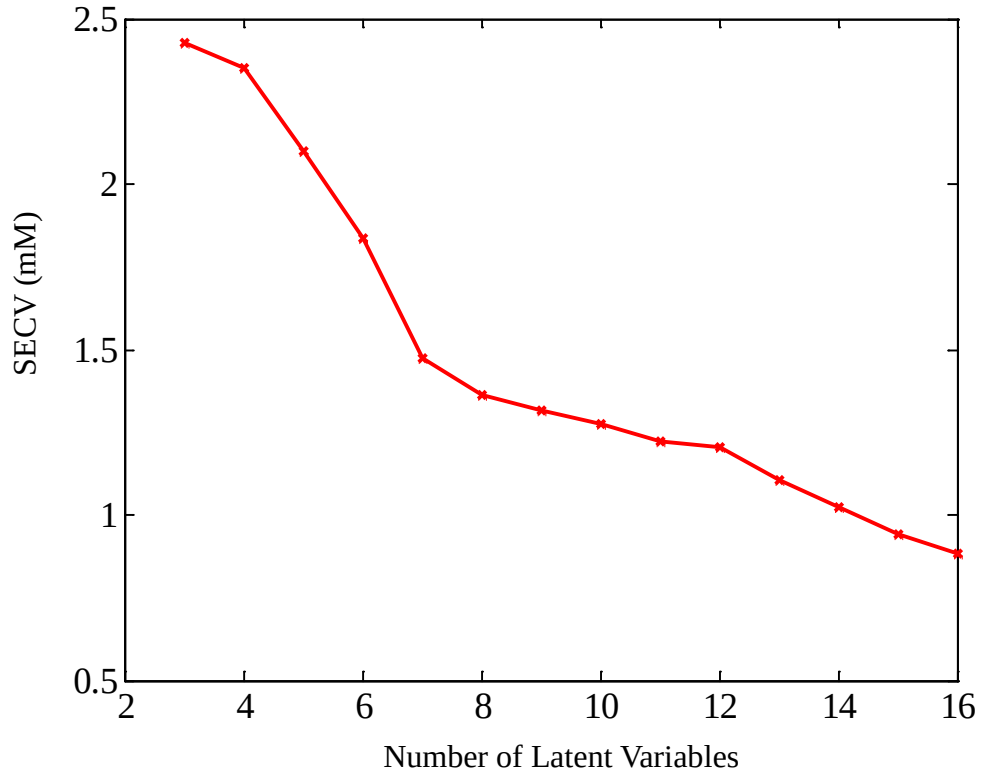


Figure 6.26 Plot of SECV with respect to the number of latent variables for the optimal wavenumber range of 4650-4300 cm^{-1}.

For each combination of spectral range and 12-16 latent variables, three replicate classifiers were computed based on three discriminant functions. Table 6.4 summarizes the percentage (average ± standard deviation) of missed and false alarms obtained when the computed classifiers were applied to the 34 alarm and 15 non-alarm patterns in the monitoring set

The wavenumber range of 4700-4300 cm^{-1} and 13 PLS factors were used as the optimal combination of parameters on the basis of the minimum number missed and false alarms. This combination showed no missed or false alarms for the monitoring set. The

glucose concentration profile for the monitoring set is given in Figure 6.27. The plot of discriminant scores for the optimal wavenumber and latent variable combination is given in Figure 6.28.

Table 6.4 Average percentage of missed and false alarms for the monitoring set

Spectral range (cm^{-1})		Latent variables				
		12	13	14	15	16
4650-4300	A.F (%)[b] ± S.D	0.0 ±0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.M (%)[a] ± S.D	66.7 ± 57.3	100.0 ± 0.0	100.0 ± 0.0	88.2 ± 5.9	87.3 ± 19.6
	D.S (%)[c] ± S.D	64.2 ± 3.2	67.3 ± 2.7	56.5 ± 8.0	68.6 ± 2.0	73.0 ± 4.7
4700-4300	A.F (%)[b] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.M (%)[a] ± S.D	21.5 ± 23.8	0.0 ± 0.0	73.5 ± 18.4	95.1 ± 8.5	56.9 ± 4.5
	D.S (%)[c] ± S.D	67.3 ± 3.3	69.1 ± 2.3	59.9 ± 5.0	67.8 ± 3.8	63.9 ± 1.0
4700-4250	A.F (%)[b] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.M (%)[a] ± S.D	32.4 ± 22.2	19.6 ± 14.5	18.6 ± 16.2	1.9 ± 1.7	35.3 ± 15.6
	D.S (%)[c] ± S.D	58.9 ± 5.7	64.6 ± 8.1	56.9 ± 13.9	65.8 ± 8.7	67.0 ± 4.2
4750-4250	A.F (%)[b] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.M (%)[a] ± S.D	93.1 ± 11.8	94.1 ± 10.1	97.1 ± 5.1	66.7 ± 19.6	67.7 ± 37.0
	D.S (%)[c] ± S.D	62.9 ± 2.5	62.2 ± 6.2	72.7 ± 6.3	67.2 ± 7.5	72.6 ± 4.0
4800-4250	A.F (%)[b] ± S.D	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	A.M (%)[a] ± S.D	20.6 ± 25.6	9.8 ± 10.3	20.6 ± 18.4	2.9 ± 2.9	3.9 ± 3.4
	D.S (%)[c] ± S.D	65.1 ± 0.9	69.8 ± 1.2	66.0 ± 2.3	70.9 ± 3.4	67.2 ± 2.4

[a]Average percentage of missed alarms (AM) ± standard deviation.

[b]Average percentage of false alarms (AF) ± standard deviation.

[c]Average percentage of separated alarm patterns with a single discriminant (DS) ± standard deviation

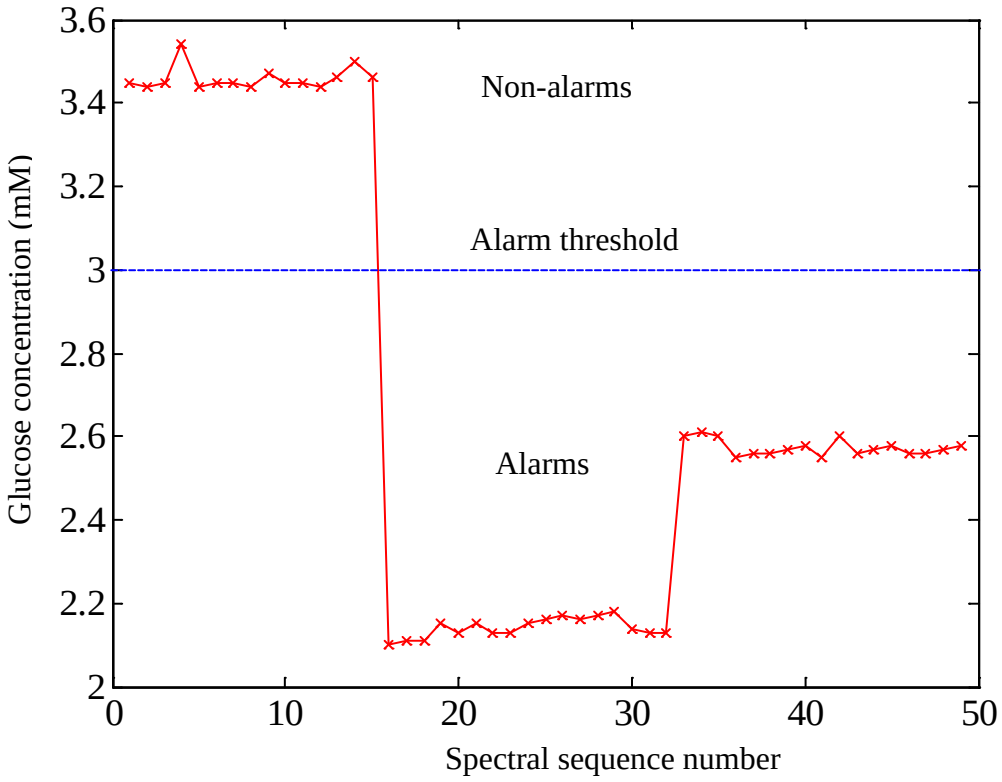


Figure 6.27 Glucose concentration profile for the monitoring set. The alarm threshold value was 3.0 mM. There were 15 non-alarms and 34 alarm patterns.

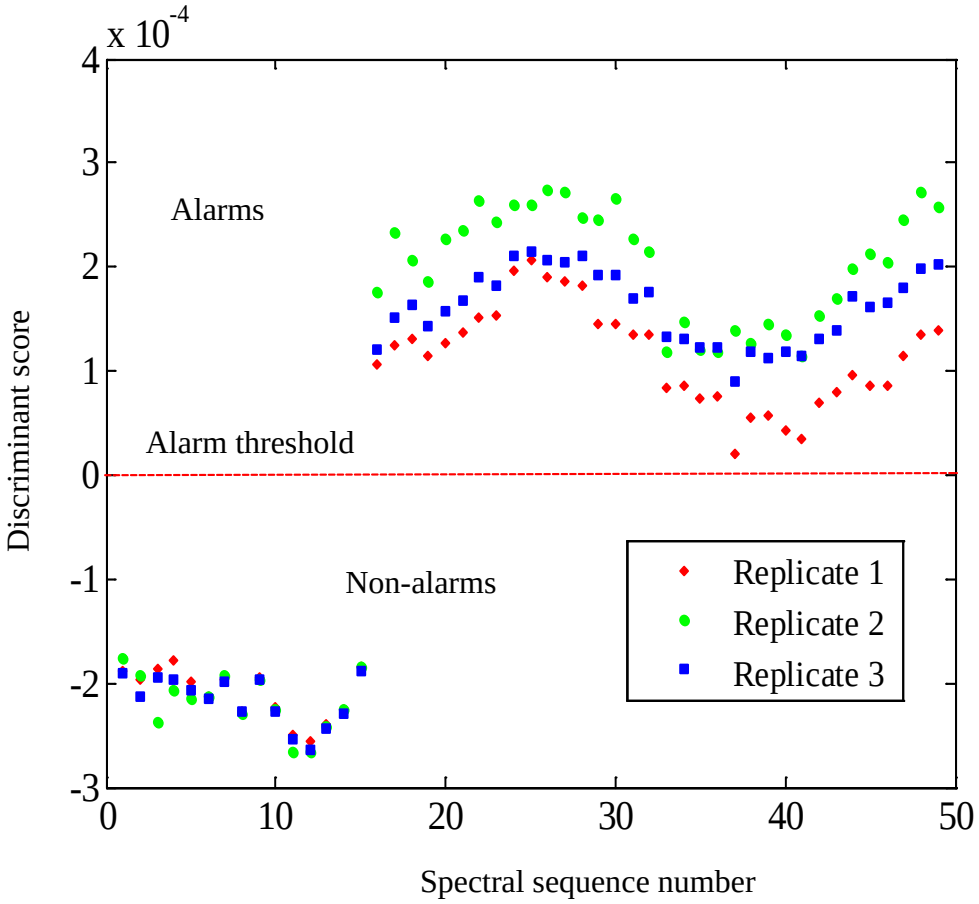


Figure 6.28 Discriminant score plot for the monitoring set. Results from the three replicate classifiers are shown. There were no missed or false alarms.

Analysis of Prediction Sets

The wavenumber range of 4700-4300 cm^{-1} and 13 latent variables were used as the optimal combination for application to the three prediction sets. The training set and the monitoring set were combined to form the full calibration set. Blocks were again defined on the basis of the day of data collection. A total of 12,173 differential spectra were obtained after removal of spectra with differential concentrations of 0.0 mM. The

optimal wavenumber range of the differential spectra was decomposed into 13 spectral loadings and corresponding loading weights. Patterns were defined on the basis of the 13 PLS scores. The data for the first prediction set of DS 2 (PS 4) were collected on the second day of data collection (day 3 out of the total of 18 days).Two different tissue phantom orientations were used in the collection of these prediction spectra. These orientations were also used in the collection of some of the calibration spectra. This simulates a scenario in which there is variation in the tissue composition but this variation has been captured in the calibration data.

The reference concentration for PS 4 was 3.8 mM and the critical concentration was -0.8 mM. The calibration PLS patterns were partitioned based on the critical concentration into 6558 alarm patterns and 5615 non-alarm patterns. Three replicate classifiers were computed with the calibration PLS patterns and the first three discriminants from each of the three replicates were applied to the classification of the patterns in PS 4.

Figure 6.29 shows the glucose concentration profile for PS 4 in which there were 31 alarm patterns and 34 non-alarm patterns. These alarm and non-alarm patterns were classified 100% correctly using the calibration discriminants giving no missed or false alarms. The corresponding discriminant score plot is shown in Figure 6.30. The developed alarm algorithm was able to classify the alarm and non-alarm patterns in PS 4 very effectively, even in the presence of the skin tissue phantom that employed different thicknesses during the data collection and in the presence of the urea interferent.

The 1^{st}, 2^{nd} and 3^{rd} PLS scores for the prediction set were plotted together as shown in Figure 6.31 and clearly demonstrate the data separation between the alarms and

non-alarms. For a successful data classification, the space explained by the prediction PLS scores should be within that of the corresponding calibration scores. The first three PLS scores for the calibration set and PS4 are plotted together in Figure 6.32. It can be observed that the calibration and prediction patterns occupy the same space

The first four PLS scores of PS 4 were plotted together with the corresponding maximum and minimum calibration PLS scores as shown in Figure 6.33. It is clear that the prediction PLS scores are within the space defined by the calibration scores. These plots confirm the excellent prediction results.

Data for the second prediction set (PS 5) were collected 9 days after the collection of the last calibration data. Five different orientations of the gelatin and keratin films were used during the data collection. All of these orientations were used in the collection of the calibration data.

The critical concentration for PS 5 was -0.6 mM. The data partitioning based on this critical concentration produced 8017 alarm patterns and 4156 non-alarm patterns. Figure 6.34 shows the glucose concentration profile for PS 5 in which there were 38 non-alarm and 32 alarm patterns. The resulting discriminant score plot is shown in Figure 6.35. There were 10 missed alarms and no false alarms using the first replicate classifier. The second replicate classifier produced one missed alarm and twelve false alarms, while the third classifier had seven missed alarms and no false alarms. Applying the committee decision rule described previously, there were seven overall missed alarms and one false alarm. This corresponds to 78 % correct detection of the alarm events and a false detection percentage of 2.6 %.

The third, fourth and fifth PLS scores for PS 5 were plotted together to see the differences in the regions in the data space occupied by the alarm and non-alarm patterns. The three PLS scores plotted together in Figure 6.36 show different regions for the alarm and non-alarm patterns.

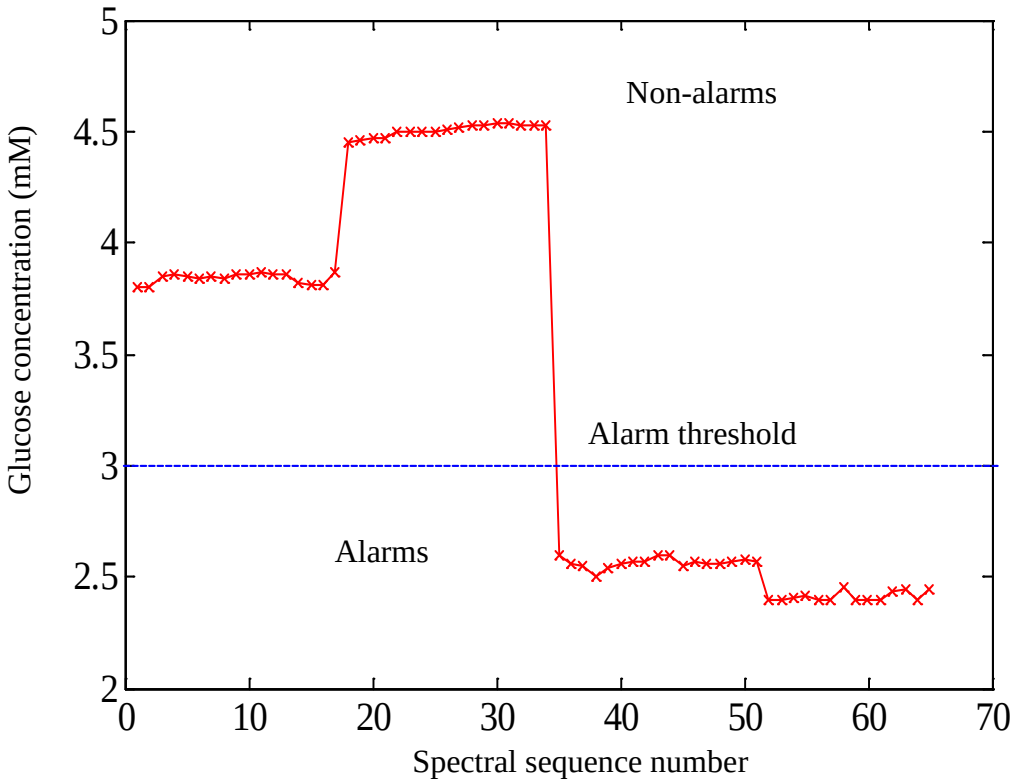


Figure 6.29 Glucose concentration profile for PS 4. The alarm threshold was 3.0 mM. There were 31 non-alarm and 34 alarm patterns, respectively.

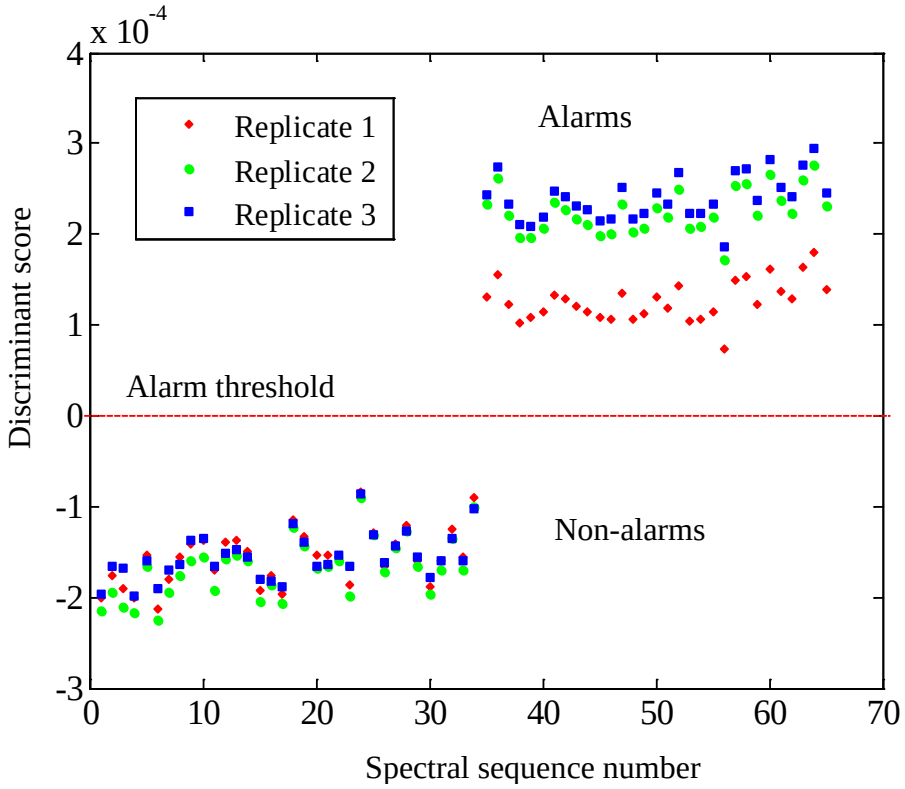


Figure 6.30 Discriminant score plot for PS 4. Results from the three replicate classifiers are shown. There were no missed or false alarms.

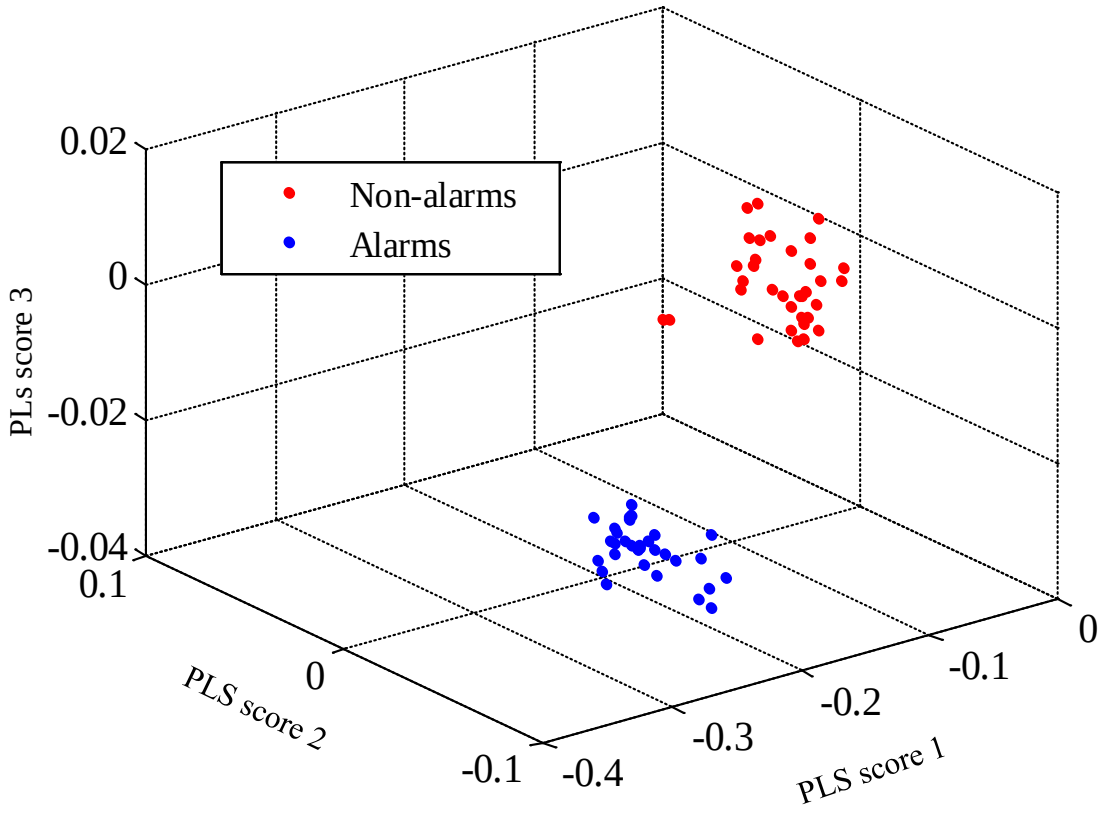


Figure 6.31 The 1st, 2nd and 3rd PLS scores plotted together for PS 4. Clear separation is noted between the alarm and non-alarm patterns.

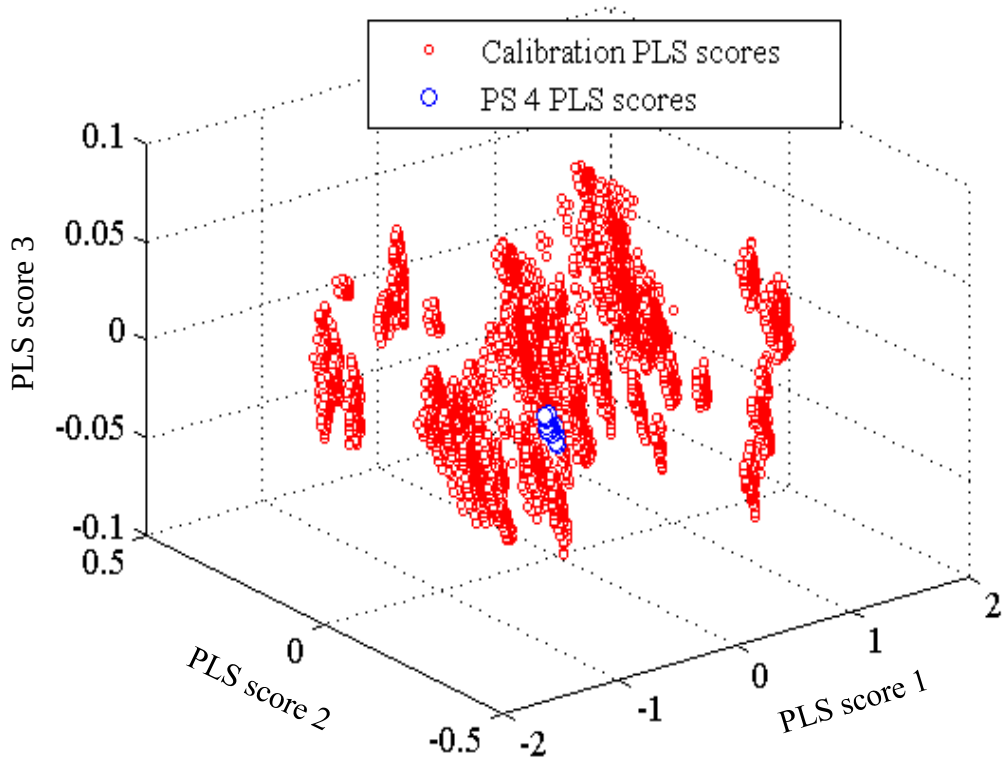


Figure 6.32 First three PLS scores plotted for the calibration data set and PS4. Clear overlap of the patterns is noted. This verifies that the calibration and prediction data are consistent.

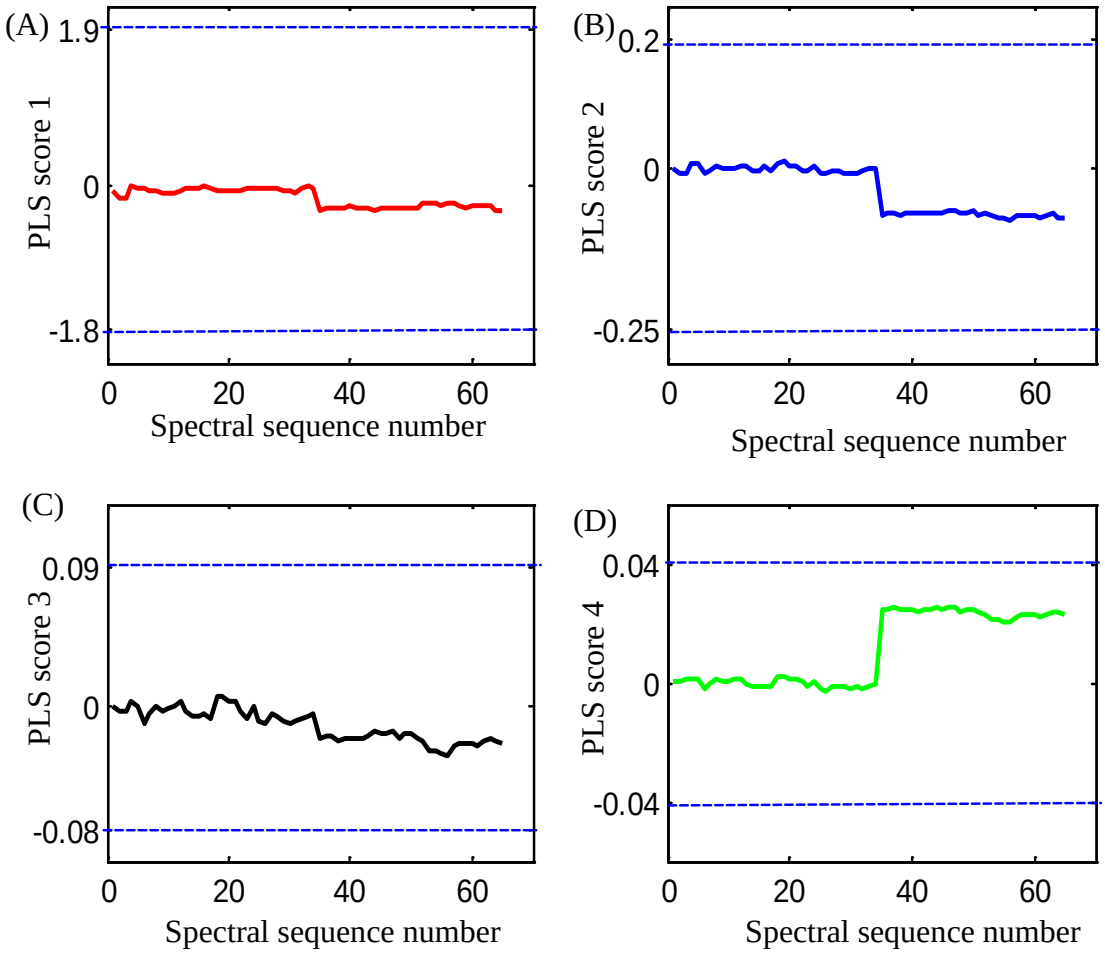


Figure 6.33 Plot of PLS scores with respect to spectral sequence number for PS 4. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as the dashed lines. Panels A, B, C, and D correspond to the scores along latent variables 1, 2, 3, and 4, respectively.

The first three PLS scores for the calibration set and PS5 are plotted together in Figure 6.37. It can be noted that the two data groups occupy the same space, indicating good agreement between the calibration and prediction data.

The first four PLS scores of the prediction set were plotted together with the corresponding maximum and minimum calibration PLS scores as shown in Figure 6.38. The PLS scores for the prediction data fall within the data space defined by the calibration.

The data in Figures 6.36, 6.37 and 6.38 suggest the classification results should be better than the actual results observed. The plot of discriminant scores in Figure 6.35 reveals that five of the seven missed detections occur in a contiguous range of spectra. These missed alarms correspond to spectra collected just after the film orientation was changed. There might have been a disturbance to the experimental setup because of this manual change. This was further confirmed by applying PCA to the set of single-beam spectra collected in this film orientation. The PC scores in Figure 6.39 clearly demonstrate the differences in the first set of spectra just after the film orientation was changed (sequence numbers 55-59 in the figure). This leads to a different pattern orientation for this group of spectra. Because the calibration data are used in the positioning of the discriminant boundaries, mismatches between the pattern orientations in the calibration and prediction data spaces can lead to imprecisely placed boundaries. This may be causing the lower classification performance with PS 5.

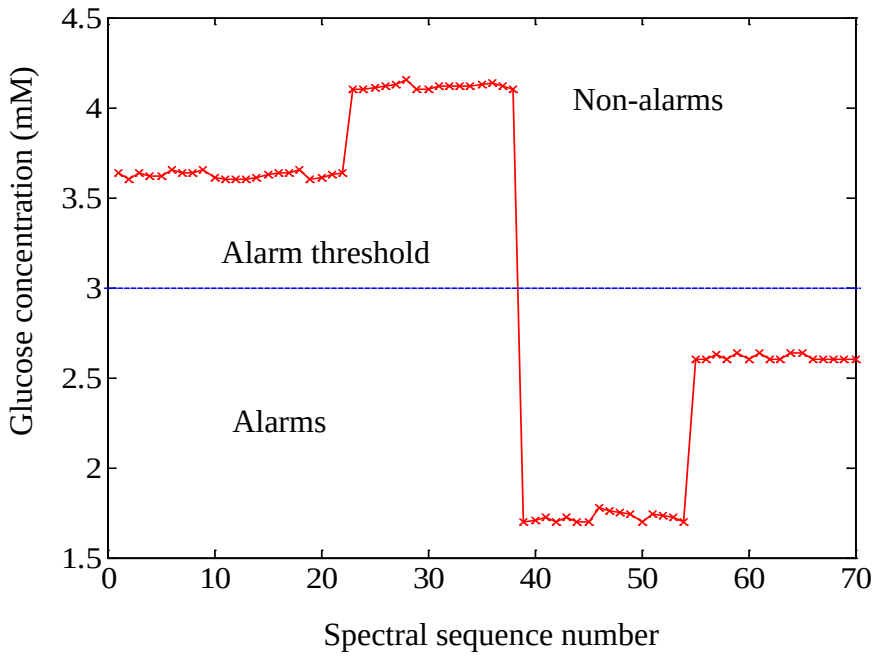


Figure 6.34 Glucose concentration profile for PS 5. The alarm threshold was 3.0 mM, and there were 64 non-alarm and 38 alarm patterns.

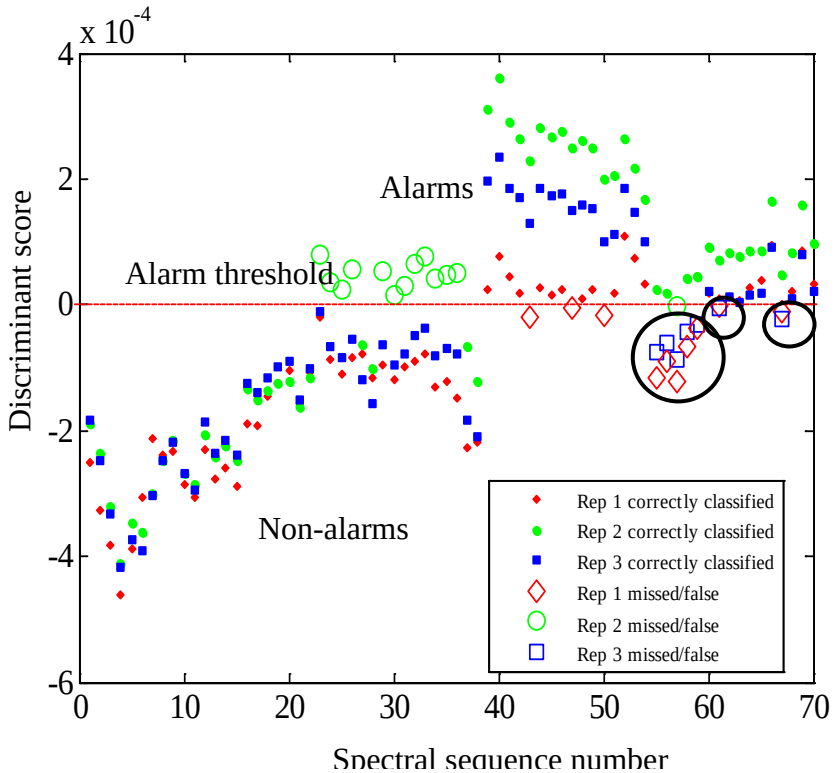


Figure 6.35 Discriminant score plot for PS 5. Discriminant scores are displayed for each of the three replicate classifiers. When the committee classification rule was applied, there were seven missed detections and one false alarm. The missed alarms that resulted from the application of the committee rule are circled in the figure.

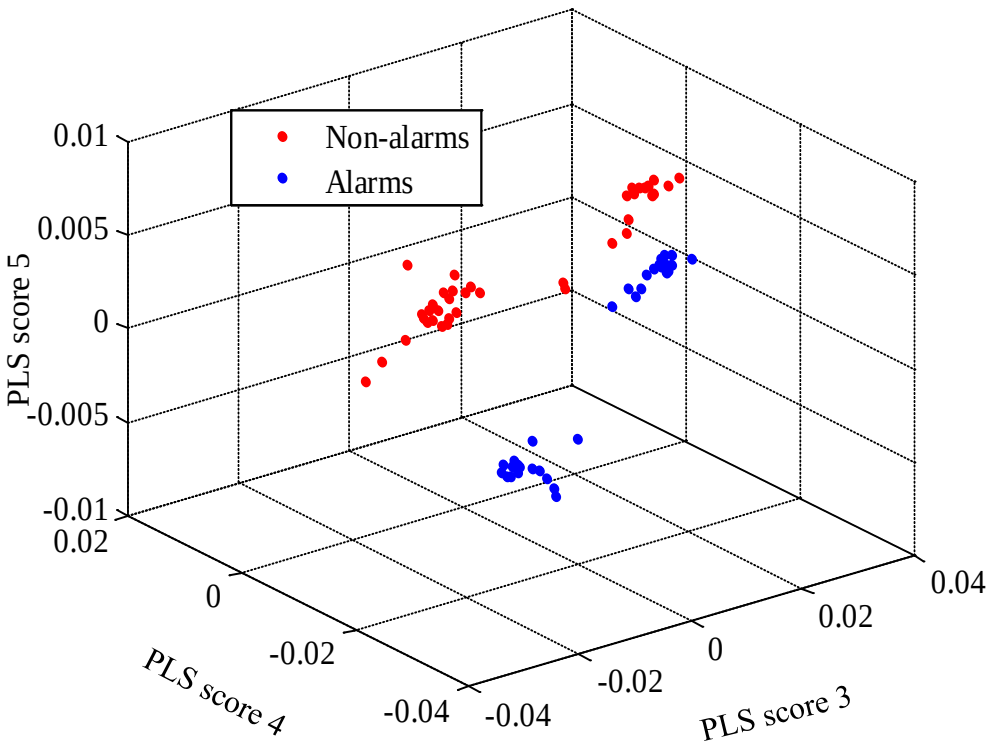


Figure 6.36 The 3rd, 4th and 5th PLS scores plotted together for PS 5. The alarm and non-alarm patterns are clearly separated.

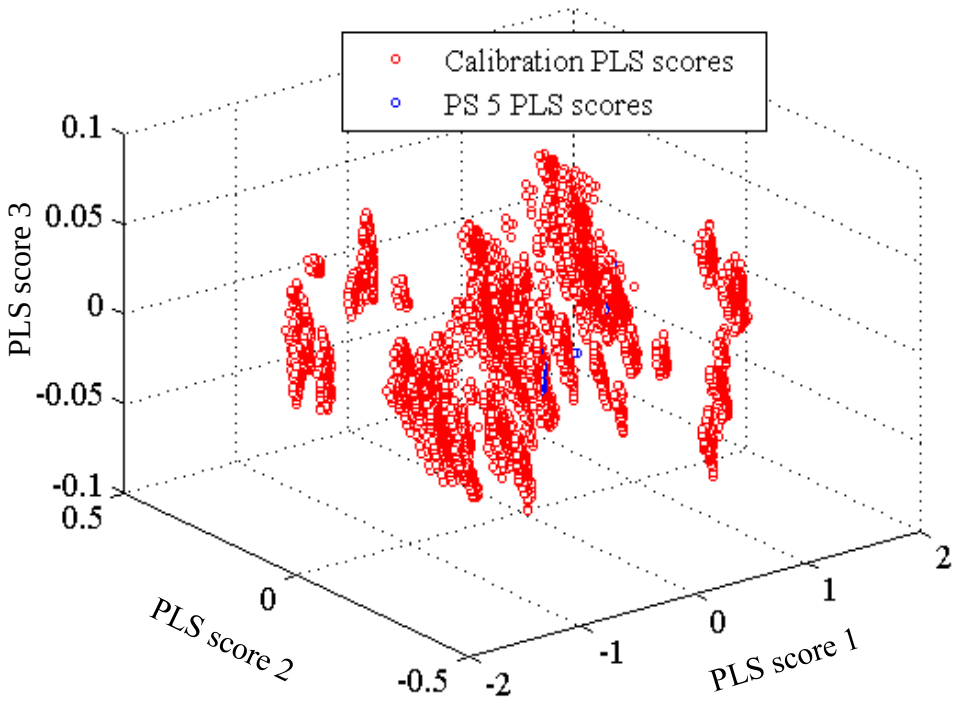


Figure 6.37 First three PLS scores plotted for the calibration data set and PS5. Clear overlap of the patterns is noted. This verifies that the calibration and prediction data are consistent.

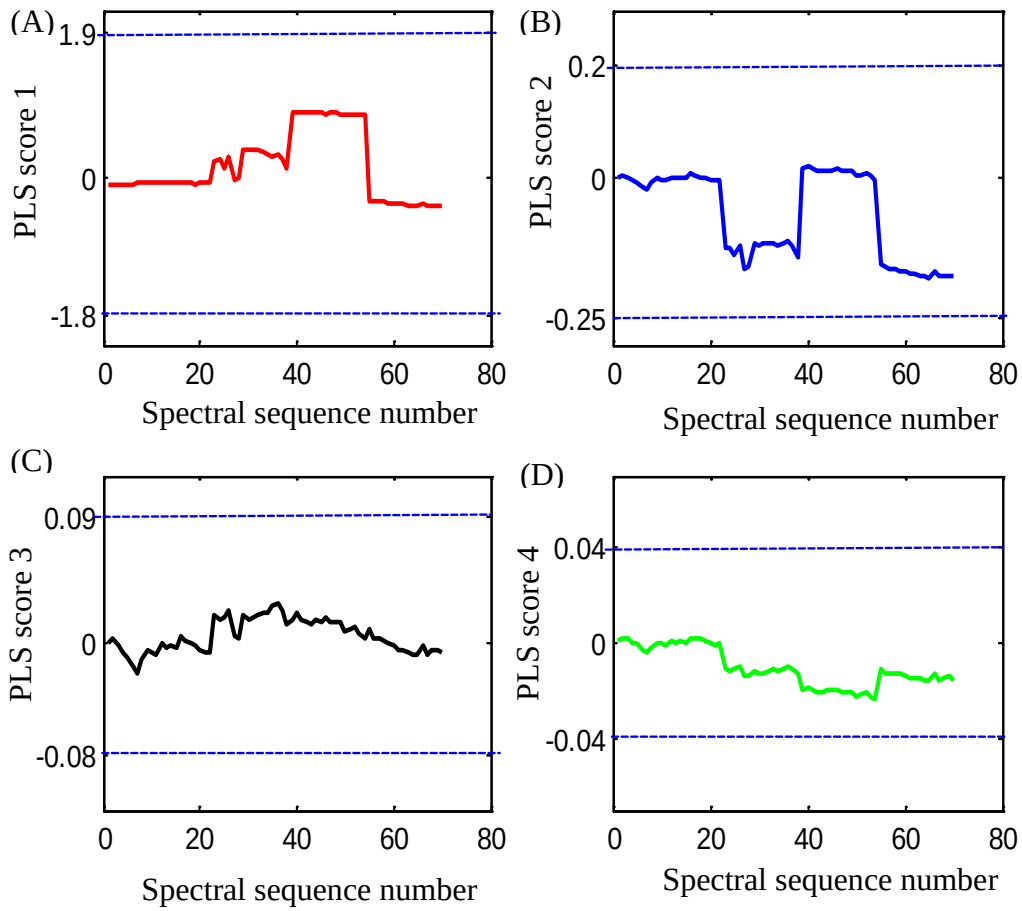


Figure 6.38 Plot of PLS scores with respect to spectral sequence number for PS 5. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as the dashed lines. Panels A, B, C, and D correspond to the scores along latent variables 1, 2, 3, and 4, respectively.

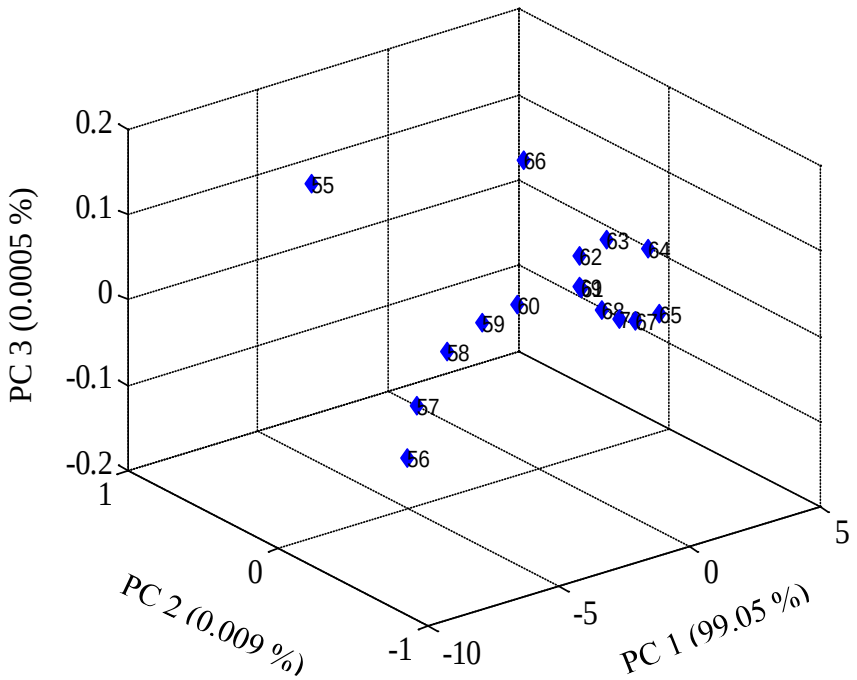


Figure 6.39 Principal component score plot computed from the last set of spectra collected in PS 5 (spectra 55-70).Spectra were mean-centered before the calculation and the 4200 to 4900 cm^{-1} range was used. The first three principal components account for 99.6 % of the data variance. The spectra collected immediately after the manual change of film orientation have a different variance from the rest (spectra 55-59).

The data for the third prediction set (PS 6) were collected on day 18, 15 days after the last calibration data. The tissue phantom orientations used with this prediction set were completely different from those used on the calibration days. This simulates a possible scenario in *in vivo* experiments in which a measurement site on the tissue is used whose spectral background has not been precisely captured during the collection of the calibration data.

The critical concentration for PS 6 was -0.6 mM which yielded 8017 alarm patterns and 4156 non-alarm patterns when the calibration database was partitioned. Figure 6.40 shows the glucose concentration profile for PS 6 in which there were 38 non-alarm and 64 alarm patterns. The resulting discriminant score plot for PS 6 is shown in Figure 6.41. There were no missed alarms and no false alarms using the first replicate classifier for the data prediction. There were no missed alarms and eight false alarms using the second replicate classifier. The third replicate classifier predicted no missed alarms and sixteen false alarms. When the committee classification rule was applied, there were no missed alarms and eight false alarms. This corresponded to successful detection of 100% of the alarm events with 21 % false alarms.

The first, second and third PLS scores for PS 6 are plotted together in Figure 6.42. As observed previously, there are different regions allocated to the alarm and non-alarm patterns. The calibration PLS scores are plotted together with the PS6 PLS scores in Figure 6.43. In this case, it can be seen that some PS 6 patterns have PLS scores that are outside the space defined by the calibration scores. This again illustrates that unpredictable results may be obtained when the calibration data do not adequately

encompass the prediction data. In Figure 6.44, the first four PLS scores are plotted together with the corresponding maximum and minimum calibration scores.

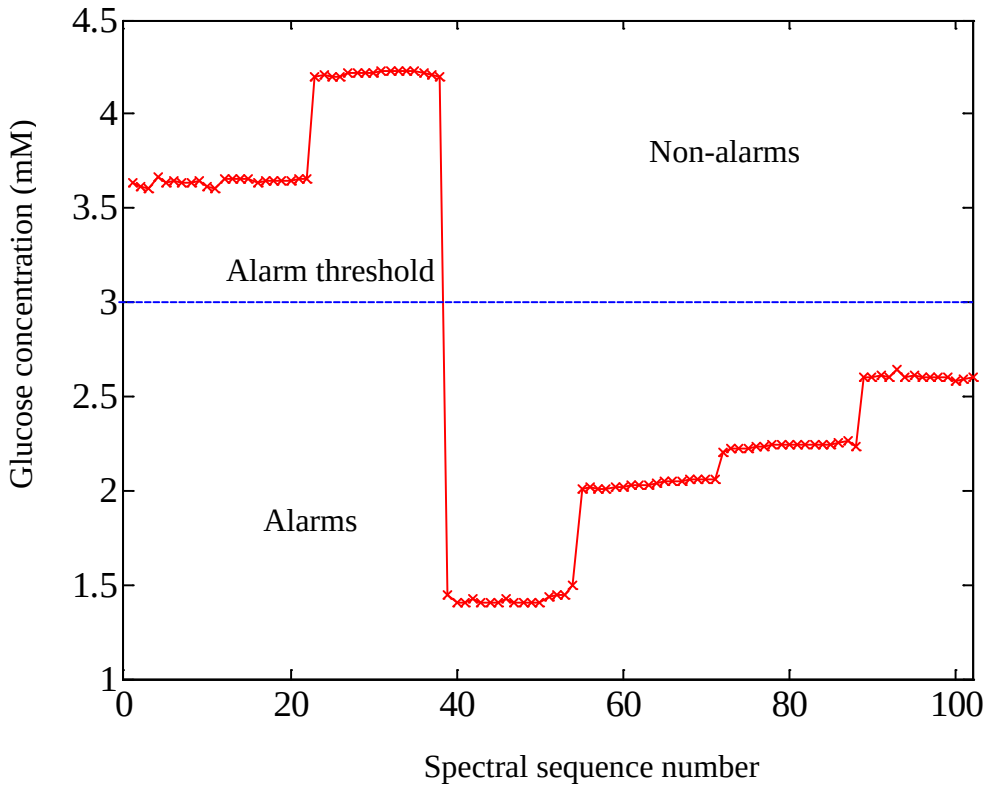


Figure 6.40 Glucose concentration profile for PS 6. The alarm threshold was 3.0 mM, and there were 64 and 38 alarm and non-alarm patterns, respectively.

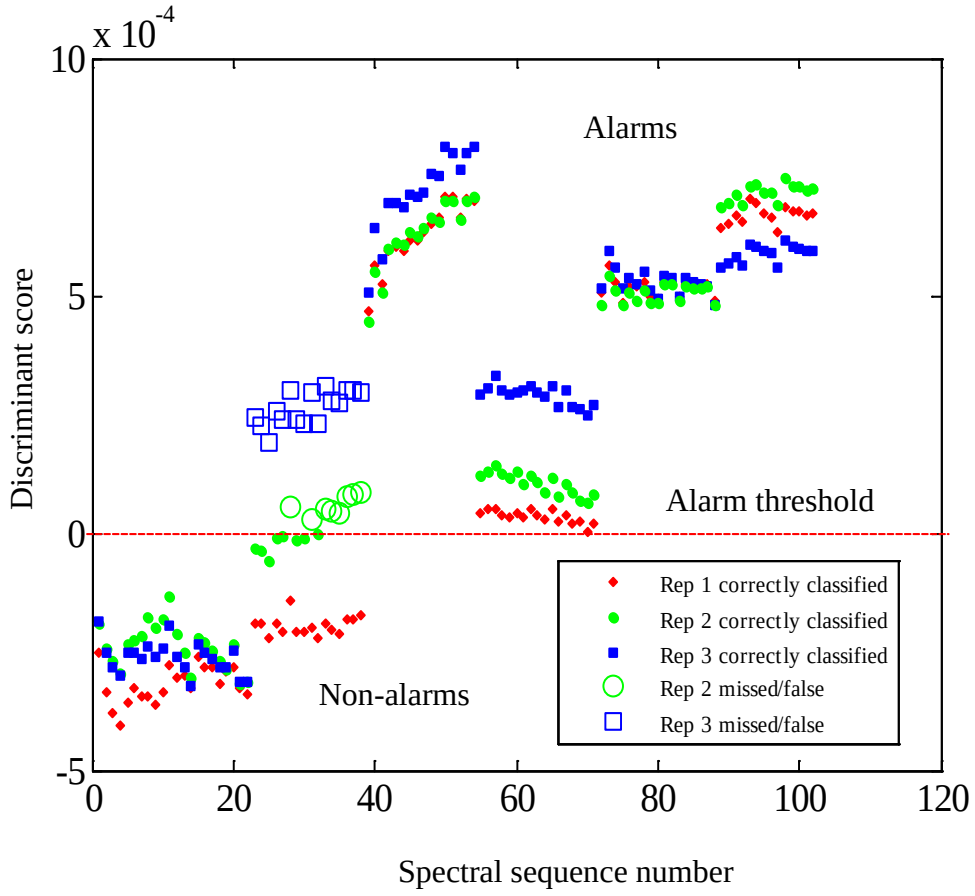


Figure 6.41 Discriminant score plot for PS 6. Discriminant scores for each of the three replicate classifiers are shown.

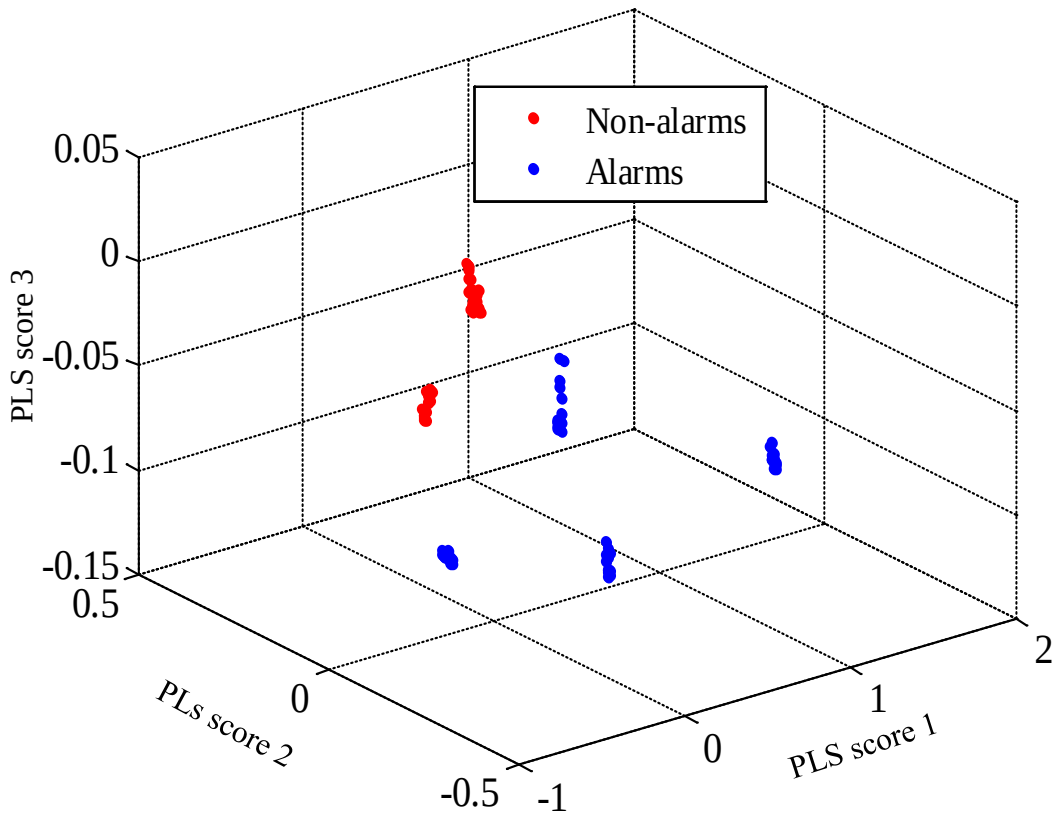


Figure 6.42 The 1st, 2nd and 3rd PLS scores plotted together for PS 6.

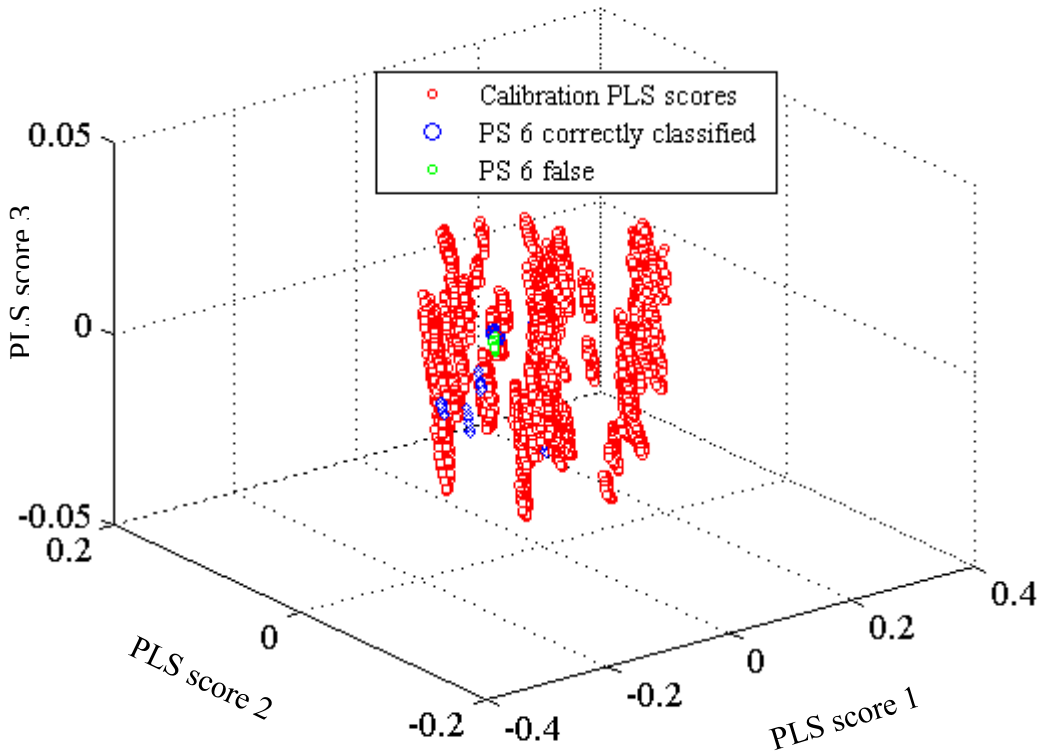


Figure 6.43 Second, third and fourth PLS scores plotted for the calibration data set and PS6. The PS 6 scores are not completely overlapped with the calibration scores as noted before.

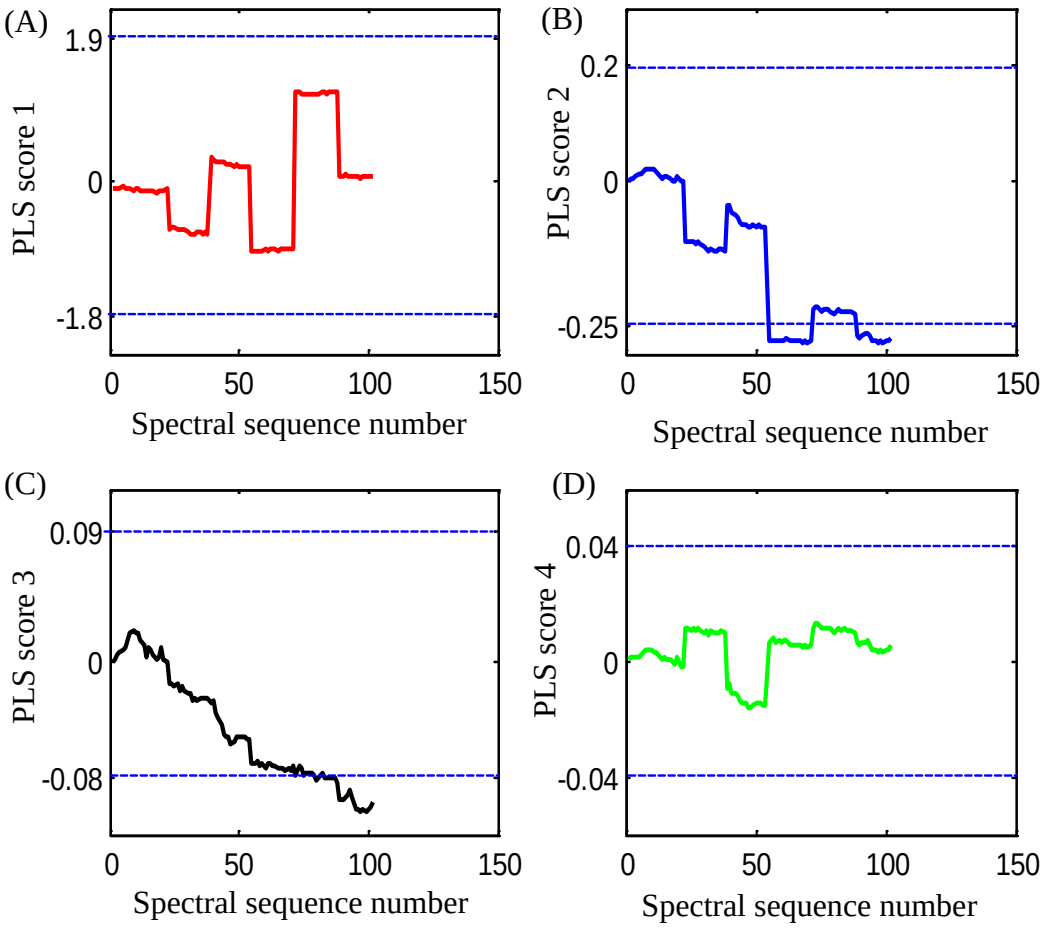


Figure 6.44 Plot of PLS scores with respect to spectral sequence number for PS 6. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as the dashed lines. Panels A, B, C, and D correspond to the scores along latent variables 1, 2, 3, and 4, respectively.

Conclusions

In this chapter, the nocturnal hypoglycemic alarm algorithm presented in Chapter 3 was tested with two *in vitro* simulation studies that served to mimic glucose excursions in the human body. These studies incorporated a tissue phantom composed of two main protein components present in human skin tissue, collagen (gelatin) and keratin. In addition to that, variable levels of urea were used as an overlapping spectral component.

The first dynamic study was comparatively simple in that the tissue phantom orientation was kept constant throughout the multiple days of data collection. The nocturnal hypoglycemic alarm algorithm performed well with all three prediction sets. There were no missed alarms and a total of four false alarms out of 121 non-alarm spectra across the three prediction sets. The false alarm rate was 3.3 %.

The second dynamic study was more complex as the tissue phantom orientation was changed throughout the data collection to nine different combinations of the gelatin and keratin films. This produced more complex data and simulated the case in which an *in vivo* measurement might be made across multiple locations on the body. The data were collected over multiple days and the hypoglycemic alarm algorithm performed well with the first of the three prediction sets, giving no missed or false alarms. However, the second and third prediction sets resulted in a total of seven missed alarms and eight false alarms. The overall recognition of the alarm events in the three prediction sets was 92.7 % while the false alarm rate was 7.9%.

Especially for the third prediction set, there was evidence that differential spectra in the prediction set were not well characterized by the calibration data. This underscores the importance of one of the key assumptions made when multivariate models are built:

the calibration data used to define the model coefficients must be accurate in representing the prediction data to which the model will be applied in the future. If this assumption is not correct, unpredictable results will be obtained in prediction.

The results presented in Chapters 4 and 6 clearly demonstrate that the proposed hypoglycemic alarm algorithm can function well in increasingly complex matrixes as long as the calibration data are representative of the data used for future predictions. These results also illustrate the importance of using replicate classification models and pooling their results to reach a decision regarding the data class of an unknown pattern. The utility of differential spectra to expand the data space of the available calibration spectra was also demonstrated. While the calculation of the differential spectra can also help to remove common background features, this simplification of the data does not occur when the background changes after the collection of the reference spectrum used to initialize the alarm algorithm. This was the case in the prediction data from DS 2 in which the orientations of the protein films were changed after the reference spectrum was defined. These concepts will be explored further in Chapter 7 in which a rat animal model is used to implement a true noninvasive measurement through tissue.

CHAPTER 7

IN VIVO SIMULATION STUDIES FOR THE DEVELOPMENT OF A NOCTURNAL HYPOGLYCEMIC ALARM ALGORITHM BASED ON NEAR-INFRARED SPECTROSCOPY

Introduction

Chapters 4, 5, and 6 discussed in detail *in vitro* simulation approaches for use in testing the proposed noninvasive nocturnal hypoglycemic alarm algorithm. The work described in this chapter extends the testing of the alarm methodology to the use of an animal model to implement true *in vivo* noninvasive blood glucose measurements.

Animal models provide a useful bridge between *in vitro* studies and ultimate testing with human subjects. An appropriate animal model allows the simulation of noninvasive human measurements and can provide key information to aid the development of a successful noninvasive glucose sensing technology. Animal models give the freedom to control several important parameters of the study, such as the range of *in vivo* glucose concentrations and the time, rate and the direction of glucose concentration changes.[88]

Optical Path Length Variation

One of the additional complexities that becomes important in moving from the *in vitro* studies described previously to the *in vivo* measurements discussed in this chapter is the issue of variation in optical path length. In the *in vitro* work, glucose was always

present in a sample cell with a fixed path length. For the *in vivo* work, however, a fixed aqueous path length for the glucose molecules cannot be guaranteed.

Glucose in living tissue can be found in intra- and extracellular fluids as well as in capillary blood. The living tissue also contains a significant amount of non-aqueous material. The tissue structure can thus be considered to contain an aqueous layer in which glucose is present and a solid or non-aqueous layer in which there is essentially no glucose that is accessible to the NIR measurement. Both layers are present within the optical path, but the effective path length of the aqueous layer determines the glucose sensitivity.

The Beer-Lambert law dictates that the product of concentration, path length, and molar absorptivity determines the measured absorbance at any given wavelength. Changes in the aqueous path length thus have a direct impact on a predicted glucose concentration or alarm decision derived from an *in vivo* measurement. In an *in vivo* absorbance measurement made relative to an open-beam air background, a path length estimate has to be made to correct for any variation in path length arising from changes in the measurement site or in the amount of tissue sampled.

One of the advantages of the differential spectrum calculation employed in this research is that path lengths are normalized to the reference spectrum and associated reference concentration obtained at the beginning of the sleep period. As long as path length variation is minimal after the reference spectrum is taken, the effect of path length changes on the alarm decision produced by the classification model is minimal.

Measurement Site

Another complexity associated with the *in vivo* application is the choice of tissue measurement site. Several issues impact this choice. The sensitivity of a transmission measurement is directly proportional to the length of the aqueous path in which the glucose molecules reside. Assuming constant noise, if there are more glucose molecules to absorb the incident light, the spectral features related to glucose will be easier to distinguish from those of background variance and noise. Multi-scattered photons will also effectively increase the optical path length.[9] Thus, selection of a measurement site that maximizes the aqueous path length is important.

The highly absorbing nature of water, as well as the absorbing and scattering properties of the tissue matrix, place limits on the thickness of tissue that can be sampled with currently available light sources. Spectral noise is typically dominated by the intrinsic noise of the detector, and thus maximizing the light throughput provides the easiest route to improving the signal-to-noise ratio (SNR) of the measurement. Thus, achieving both good sensitivity and a high SNR dictates a measurement site in which the aqueous component is as large as possible and the presence of other absorbing or scattering species is minimized.[77] As an example, fatty tissue possesses strong absorption bands near the glucose absorption bands at 4400 cm^{-1} and 4300 cm^{-1} in the combination region of the NIR spectrum. Thus, the presence of fat decreases the SNR at important wavelengths used to quantify the glucose absorbance. For this reason, any successful data collection procedure should be focused on minimizing the presence of fatty tissue.[77]

The thin skin on the upper shoulder area of Sprague Dawley rats was used in the work discussed in this chapter to simulate noninvasive human measurements. The rat

skin absorption spectrum given in Figure 7.1 clearly shows the small absorption features around 4300 cm^{-1} related to fatty tissue. An absorbance spectrum for human skin on the back of the hand is also displayed in Figure 7.1 and shows a similar shape to the spectrum collected from the rat skin tissue. The similarities in the shapes of the spectra indicate that the main chemical components are present in reasonably similar amounts. Arnold *et al.* reported that the differences in the absorbance values between human and rat tissue spectra are related to differences in light scattering by the skin tissue. It was further reported that the scattering in the rat skin tissue is typically higher than that of human tissue. This can be seen by the higher absorbance values across the spectral range in Figure 7.1.[10]

Overview of Rat Measurements

The results discussed in this chapter are based on two types of *in vivo* studies: (1) single-day experiments with anaesthetized rats and (2) multiple-day experiments with awake rats. The basic protocol for the nocturnal hypoglycemic data collection was designed to characterize the changes in skin spectra associated with hypoglycemia. This was accomplished by controlling blood glucose levels with a glucose clamp technique and collecting noninvasive spectra trough a fiber-optic interface. Rats fasted overnight to minimize glucose fluctuations before the data collection began and glucose transients were induced. Selected hypo- and hyperglycemic levels were maintained by delivering glucose or insulin as required. The lowest targeted hypoglycemic levels were around 2.5 mM (45 mg/dL). Spectra were collected continuously and were then used for implementing the nocturnal alarm algorithm as described in Chapter 3.

Animal models with multiple days of data collection were desired to build more robust sets of calibration data and also allow true external prediction outside the time span of the calibration. This creates a scenario much closer to a desired implementation protocol with a human subject. In between measurement days there may be variances associated with the skin tissue thickness, interface positioning etc. Multiple-day experiments allow the incorporation of these sources of variance into the study. The data discussed in this chapter allow a realistic assessment of the potential for implementing the proposed nocturnal hypoglycemic alarm with human subjects.

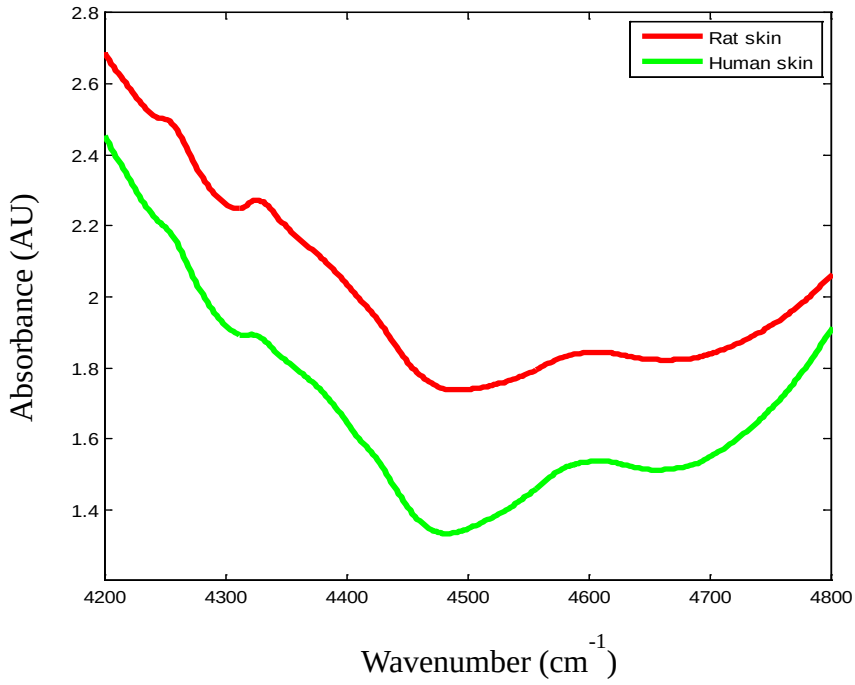


Figure 7.1. Noninvasive absorbance spectra relative to air collected from human (green) and rat (red) skin. The similarity in spectral shapes confirms the tissue composition is similar. The increased absorbance of the rat tissue across the spectral range suggests a higher degree of light scattering.

Experimental Methods

The *in vivo* measurements described below represent work performed jointly with our collaborators in the Arnold research group at the University of Iowa. The specific experiments described here represent a subset of an overall data collection effort related to noninvasive glucose measurements in the rat model. The data selected for use in testing the hypoglycemic alarm methodology were chosen on the basis of having glucose excursions above and below the targeted alarm threshold of 3.0 mM.

Reagents

Pentobarbital sodium injection solution (Ovation Pharmaceuticals, Deerfield, IL) was used to anesthetize rats. Dextrose (50 %, IVX Animal Health Inc., St. Joseph, MO) was used as the glucose injection solution. The solvent for insulin and heparinized saline was 0.9 % sodium chloride injection solution (saline) (Baxter Healthcare, Deerfield, IL). A 0.5 unit insulin injection solution was prepared by diluting 100 units/mL regular human insulin injection solution, Novoline® (Novo Nordisk Inc. Princeton, NJ). The insulin injection solution was prepared by mixing 0.075 mL Novolin® in 14.925 mL saline. The heparinized saline was prepared by mixing 1 mL of 1000 USP units heparin sodium in 49 mL of saline (American Pharmaceutical Partners, Inc., Schaumburg, IL). A 2 % Lidocaine HCL solution (Abbott Laboratories, Chicago, IL) was used topically during surgery for infiltration and nerve block. All reagents were obtained from the University of Iowa Hospitals and Clinics (UIHC) pharmacy. All procedures were approved by the University of Iowa Animal Care and Use Committee (ACURF # 0507182).

Surgical Procedures

Adult male Sprague-Dawley rats (retired breeders, weighing ~ 400 g) were used as the animal model in this work. A skin fold on the back neck of the rats was used as the spectral collection site based on previous work published by the Arnold research group.[10] As discussed previously, the similarities between the skin fold of the back of the human hand and the rat neck can be clearly seen in the spectra displayed in Figure 7.1. Rats were allowed to fast overnight before the experiment to allow a stable initial blood glucose concentration to be achieved.

The rat was anesthetized before the surgery by injecting a dosage of 50 mg/kg of pentobarbital sodium injection solution. For the single-day non-survival rat data collection, the anesthesia during the experiment was maintained by regulating a dosage of 25 mg/kg-hour pentobarbital. Hair on the upper back and femoral area was removed. During the surgery, two catheters were cannulated into a femoral vein and artery. The venous catheter was used for the infusion of glucose, insulin, saline and anesthetic. The arterial catheter was used for sampling reference blood samples.

After the surgery, the animal was transferred to the spectroscopy station, and the body temperature was maintained at 38.3 °C by a closed-loop temperature controller R/S 68900 (Barnant Co., Port Huron, MI). Supplemental oxygen was provided at a rate of 1.5 L/hour. For the single-day non-survival experiments, a mouseOX® pulse oximeter probe (STARR Life Sciences, Oakmount PA) was attached to the leg of the rat to monitor its pulse rate, pulse distention and oxygen saturation.

For the first rat used in the multiple day-data collection, a survival surgery was performed. Most of the surgical procedure was the same as described above. Two

catheters (Strategic Applications, Inc., Libertyville, IL) were inserted into the femoral artery and vein, respectively. The rest of the tubing was placed underneath the skin and the other end of the catheter was taken out from the skin on the back of the rat. This end of the catheter was then connected with one vascular access port (Strategic Applications, Inc.). The two ports were then sutured subcutaneously and wounds on the back and leg areas were also sutured. Once installed, the ports could be used for frequent arterial blood sampling and venous infusion. The mouseOX® pulse oximeter probe was attached to the tail of the rat to monitor its pulse rate, pulse distention, and oxygen saturation.

The second rat used in the multiple-day data collection was supplied with ports already inserted by Charles River Laboratories International Inc. (Chicago, IL). Other procedures used with this rat were as described above.

Instrumentation and Procedures

A Nicolet 670 FT spectrometer (Nicolet Analytical Instruments, Madison, WI) was used to collect the skin tissue spectra from the rats. Figure 7.2 displays a schematic of the instrumental setup. A 50-W tungsten-halogen light bulb (Gilway Technical Lamp, Peabody, MA) with an integrated, gold-coated reflector was used as the external light source. An E3633A 200-watt DC power supply (Agilent Technology, Van Nuys, CA) was used to keep the voltage of the light bulb at 12.00 V. The light from the spectrometer was brought to the skin tissue using a custom sapphire-rod interface. Another sapphire-rod interface was used to guide the transmitted light to an external detector. These interfaces contained two pieces of 3-mm-diameter sapphire rods, one for guiding the source interferogram to the skin tissue and the other for transporting the transmitted light

to the detector. Collected light was detected by a two-stage thermoelectrically cooled extended-wavelength InGaAs detector with a diameter of 1 mm and a 2.6-μm cutoff (Judson Technologies, Montgomeryville, PA). A low-noise transimpedance amplifier (Femto Messtechink GmbH, Berlin, Germany) was used to amplify the current signal. A K-band interference filter was used in the optical path to restrict the spectral range to 5000 to 4000 cm^{-1}. Figures 7.3 B and D display images of the experimental setup and the sampling interface, respectively.

Spectra were collected at a resolution of 8 cm^{-1} with 128 co-added scans for the data for the first rat discussed in this chapter and 32 scans for the data for all the other rats discussed. Each spectrum based on 128 scans required slightly more than 60 seconds for collection while for 32 scans it was close to 15 seconds. The interferograms were Fourier processed to single-beam spectra with triangular apodization, standard Mertz phase correction, and one level of zero filling. The resulting spectral point spacing was 1.928 cm^{-1}.

The rat positioned in the sampling interface for the single day non-survival studies is shown in Figures 7.3 C and D. Two metal blocks having central holes were glued onto the skin tissue to help fix the tissue sample in position and consequently minimize the impact of unconscious movement on spectral quality. Initially, spectra were collected without glucose or insulin perfusion and only saline was perfused to prevent dehydration. Since the animal had fasted overnight, the initial blood glucose reference values had minor fluctuations. The infusion rate was adjusted according to the body mass of the rat and the blood glucose value. Glucose and insulin were infused alternatively to achieve hyper- and hypoglycemia, respectively. At the end, blood glucose levels were allowed to return to baseline levels without any infusion of glucose or insulin.

Figure 7.4 shows a rat used for the multiple-day experiments after the surgery was performed and the rat had been placed in position to begin the data collection. The rat remained awake during the data collection for these experiments. Data collection procedures for each day were as described above for the single-day experiments.

During the spectral collection, single or multiple glucose transients were obtained by infusing glucose/insulin solutions alternatively through the venous catheter with two Dynamax peristaltic pumps (Rainin Instrument Co., Woburn, MA). A third pump was also used to infuse saline to compensate for body dehydration. Meanwhile, blood samples were taken from the arterial catheter every 5-10 minutes for reference blood glucose measurements. Blood glucose measurements were taken as the average readings of two Freestyle® glucose monitors (Abbott Laboratories, Chicago, IL).[147]

Computations

After collection and Fourier processing, spectra were transferred from the computer controlling the spectrometer to a Dell Precision 670 workstation (Dell Computer Corp.) running under Red Hat Linux (Version 5.3, Red Hat, Inc., Raleigh, NC). The majority of subsequent calculations were performed on this computer using the Matlab development environment (Version 7.4.0 (R2007a), The MathWorks, Inc., Natick, MA). Some calculations performed in Matlab made use of functions from the PLS Toolbox (Version 5.2, Eigenvector Research, Wenatchee, WA). Software for the calculation of piecewise linear discriminants used in-house software written in Fortran and compiled with the Intel Fortran Compiler for Linux (Version 10.0, Intel Corp., Santa Clara, CA).

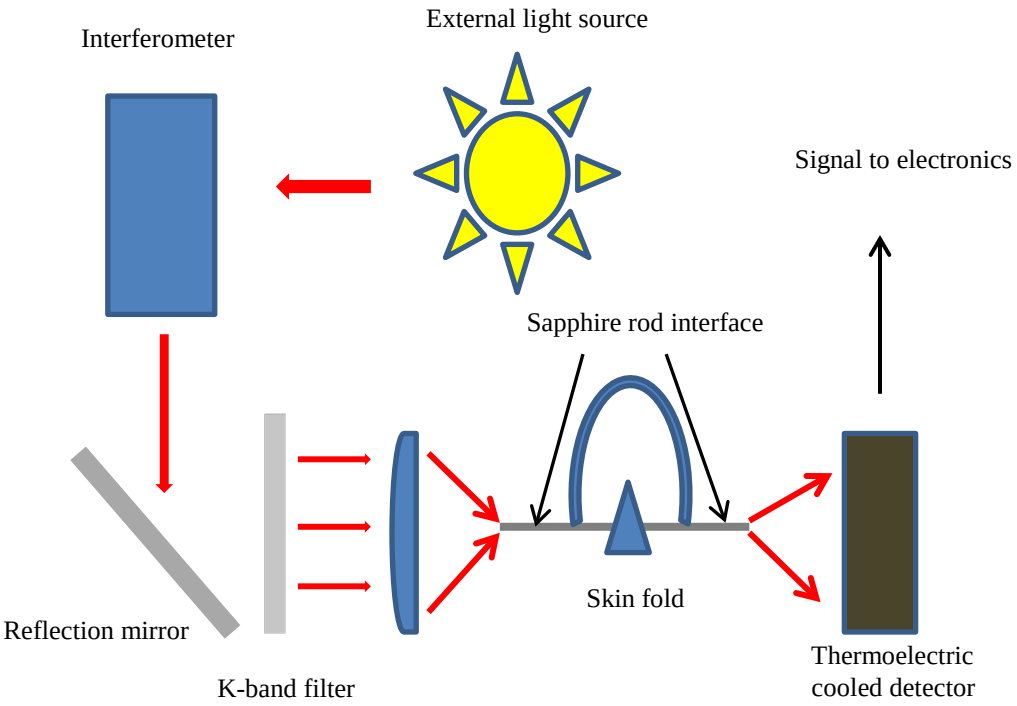


Figure 7.2. Schematic representation of the experimental setup.

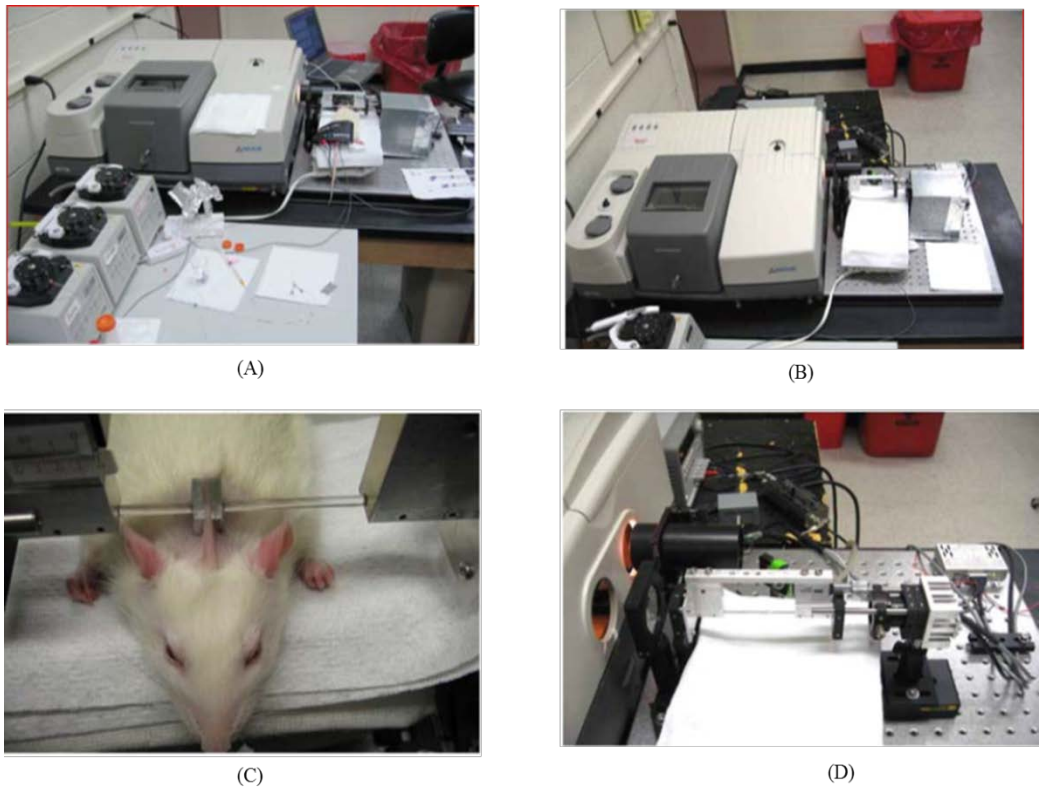


Figure 7.3. Experimental setup used for the single day non-survival rat data collection.[142] A. Overview of the instrumentation and hyper- and hypoglycemic pump system. B. Customized FT spectrometer with external light source, detector and sapphire rod interface. C. An anesthetized rat being clamped during a typical study. D. Close-up view of external light source, sapphire rod interface, and detector.

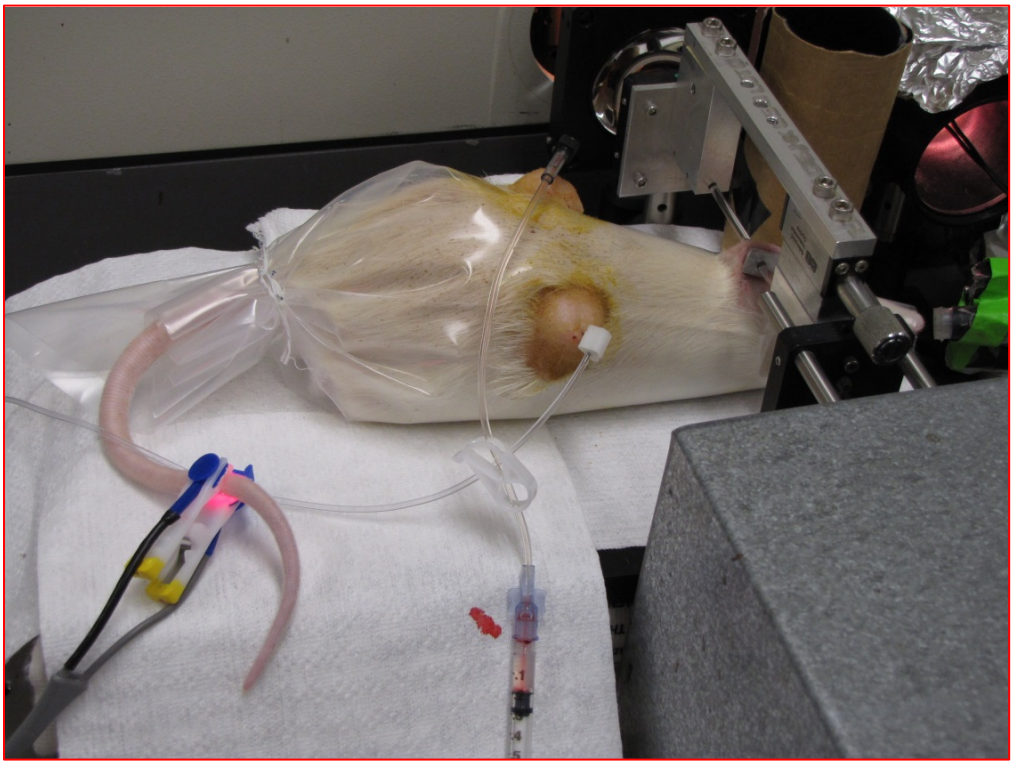


Figure 7.4. Typical rat used for the multiple-day data collection being positioned in the sample station. Two vascular access ports can be seen in the back of the rat. A mouseOX® pulse oximeter probe is attached to the tail of the rat.

Results and Discussion

This chapter discusses the data acquired from four rats that will be referenced by their serial numbers (rats 106, 114, 134 and 169). For all rats, absorbance spectra were calculated by taking the negative logarithm of the ratio of each rat skin single-beam spectrum to a mean air spectrum computed from the air spectra collected at the start and the end of the data collection session. As an illustration of the overall spectral shape and quality, Figure 7.5 displays a series of absorbance spectra collected from rat 106. Spectra were collected continuously at 128 scans while the reference blood measurements were

taken every 5-10 minutes. The absorbance spectra shown in Figure 7.5 clearly indicate water as the main spectral contributor, with the edges of the spectrum increasing in absorbance in the direction of the intense water bands located at 3300 and 5200 cm^{-1}. The peaks at 4250 cm^{-1} and at 4350 cm^{-1} arise from fat absorbance in the skin tissue and the peak at 4600 cm^{-1} corresponds to proteins, mainly keratin and collagen, in the skin tissue.

Single-Day Rat Studies

Analysis of Rat 106

Rat 106 was a male rat weighing around 438 g. The surgery was performed and the glucose transients for the hypoglycemic studies were performed on the same day. The short-term noise for rat 106 was evaluated by computing 100 % lines for pairs of consecutive single-beam spectra. These 100 % lines were converted to AU, and the region of 4300-4500 cm^{-1} was fitted to a third-order polynomial model as described in Chapter 3. The RMS noise was then computed about the polynomial fit to obtain the intrinsic measurement noise. The average RMS noise calculated for consecutive pairs of spectra collected at 128 scans was approximately 22 μAU. The RMS noise values computed for each consecutive pair of spectra are shown in Figure 7.6.

The quality of the data was initially evaluated by use of PCA. The skin tissue absorbance spectra were mean-centered, and the PCA calculation was applied over the 4900-4200 cm^{-1} range. The first two principal component scores were plotted together as shown in Figure 7.7. It can be clearly seen that the spectra collected at the beginning of the data collection, during which the skin tissue is adjusting to the interface, show different clustering from the rest of the data.

As described in Chapter 3, unusual observations can be identified relative to the principal component model by the calculation of Q and T^2. Figure 7.8 is a plot of Q vs. T^2 for the first two principal components for rat 106. This plot illustrates the larger T^2 values for the first 16 spectra of the data collection. Initial attempts to build PLS models using spectra from the beginning of the data set led to models that did not perform well in external predictions. Hence, the first 16 spectra collected at the beginning of the experiment were removed from further analysis.

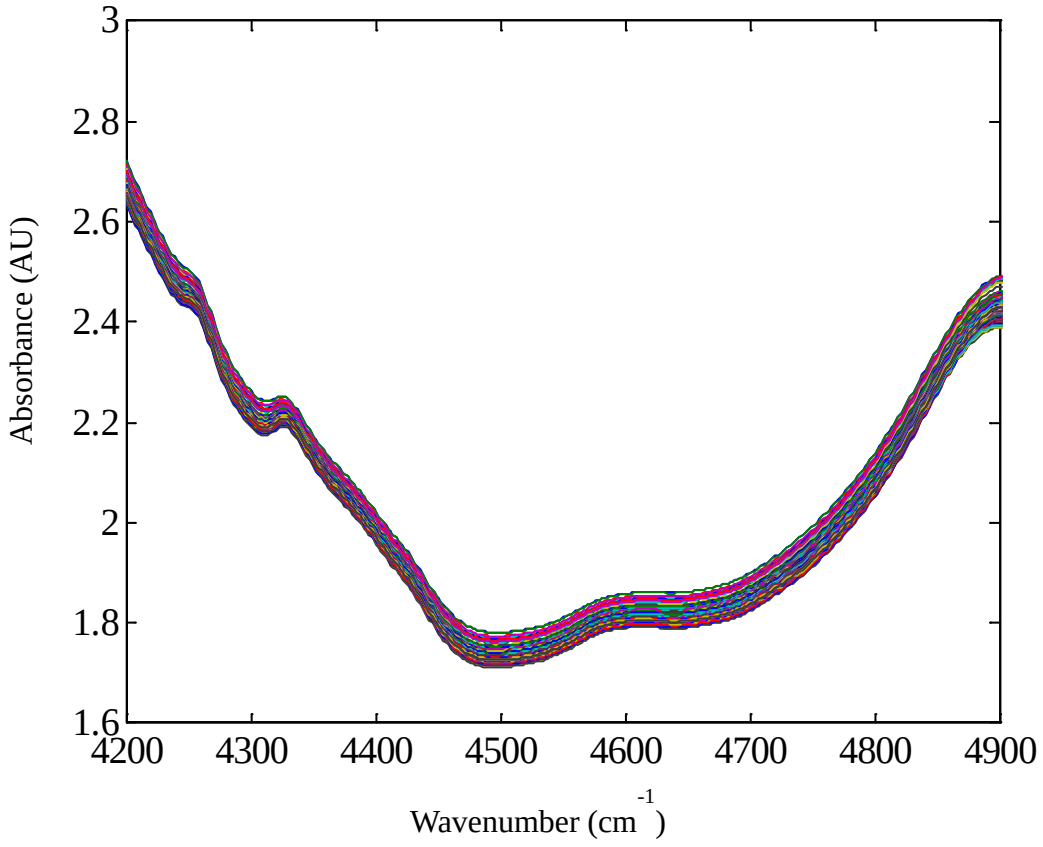


Figure 7.5. Plot of 462 absorbance spectra for rat skin tissue relative to air. Fat absorbance features can be seen at 4250 cm^{-1} and 4350 cm^{-1}, while the peak at 4600 cm^{-1} arises from skin tissue proteins. These spectra were collected from rat 106.

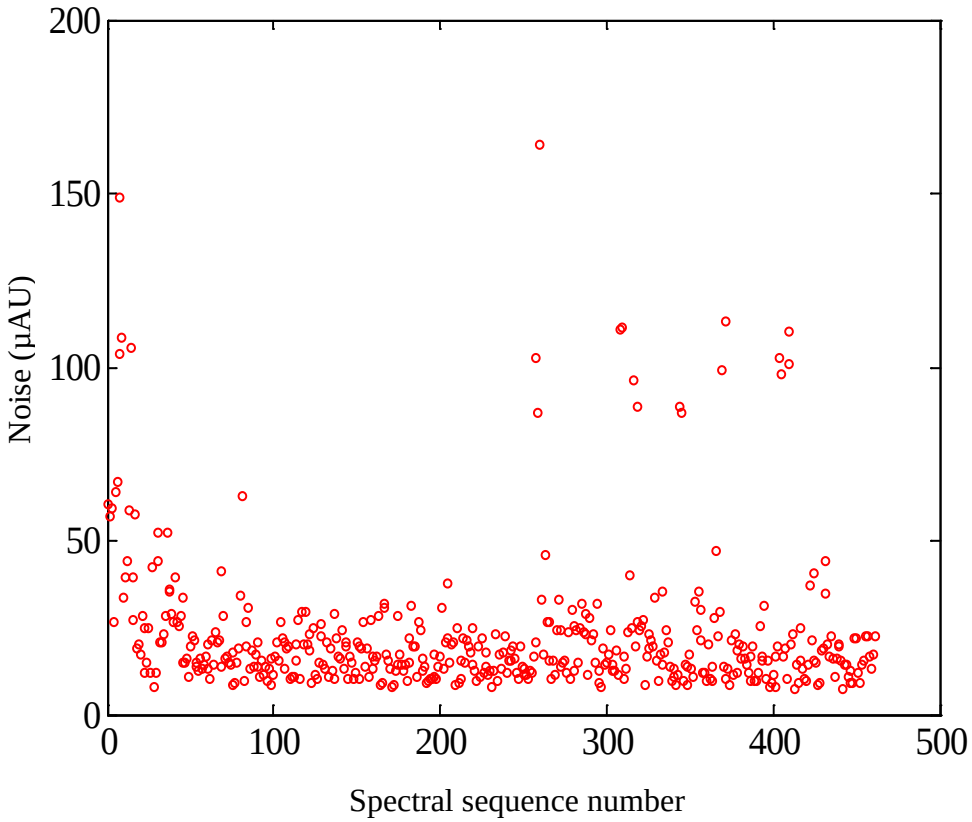


Figure 7.6. Values of RMS noise in µAU computed for consecutive pairs of spectra of rat 106. The 4500-4300 cm^{-1} region was used. Noise values were computed by fitting the 100% line in the given wavenumber region to a third-order polynomial and computing the residuals about the fit. The average noise value was approximately 22 µAU.

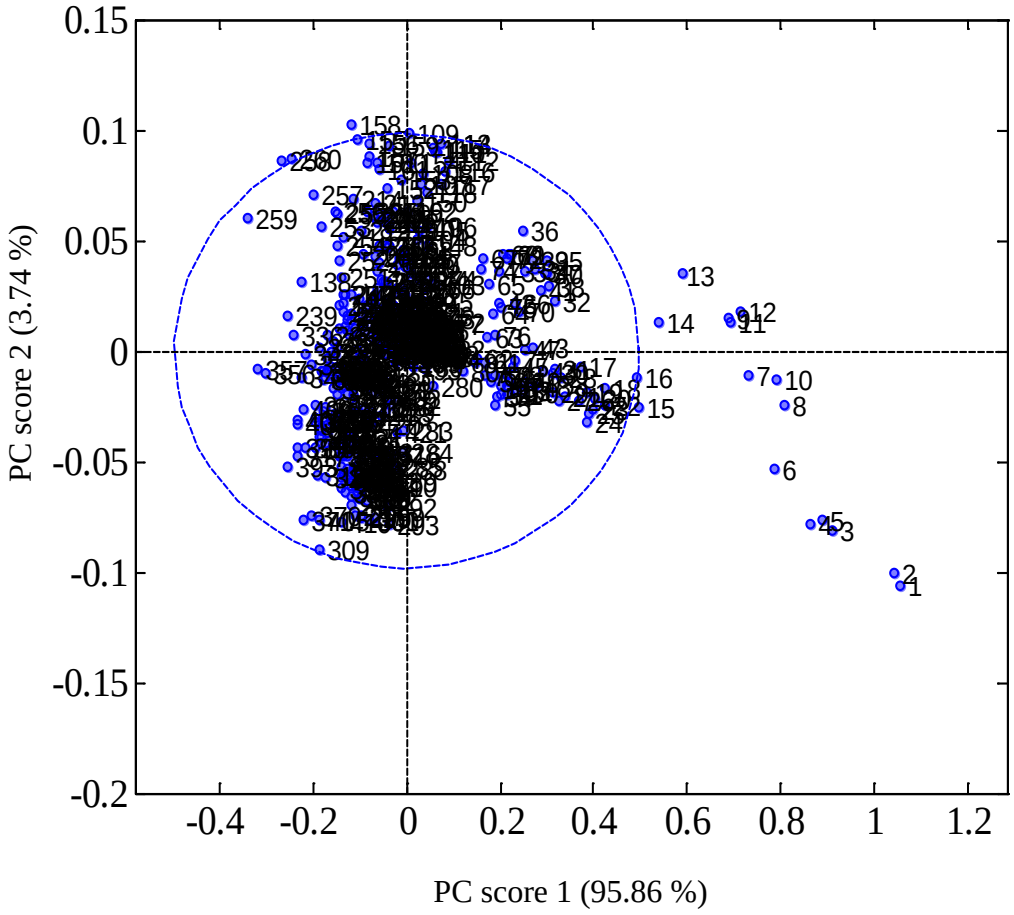


Figure 7.7. Score plot for rat 106 based on the first two principal components. Data labels indicate the spectral sequence number. Mean-centered absorbance data computed relative to an air background served as inputs to the PCA calculation. The wavenumber range of 4900-4200 cm^{-1} was used. The first two principal components accounted for a total of 99.6 % of the data variance. The ellipse shown here corresponds to the 95 % confidence interval for the scores.

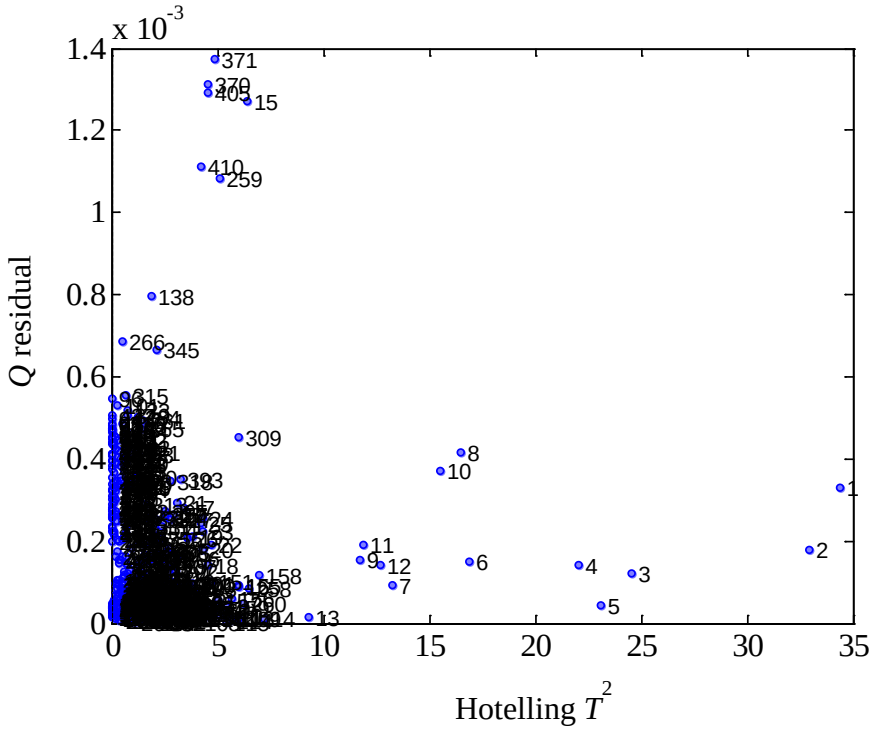


Figure 7.8. Plot of values of Q with respect to Hotelling's T^2 for the absorbance data from rat 106. The first two principal components were used for the analysis. Data labels indicate the spectral sequence number.

The glucose concentration profile for rat 106 is shown in Figure 7.9. A reference glucose concentration needed to be assigned to each of the spectra collected. The spectral collection time was linearly interpolated with the reference glucose concentrations and corresponding measurement times to assign a glucose concentration to each of the noninvasive spectra.

Accurate concentration assignment, however, requires the consideration of the time delay of glucose in the arterial blood and the interstitial fluid.[148] Reports of many research groups indicate that the glucose in arterial blood requires some time before the

corresponding concentration of glucose is observed either in the subcutaneous tissue,[35,88,149] or in the dermis layer of the skin.[150]

Assuming a delay or lag time of 0-20 min in 1-minute increments, a glucose concentration was assigned to each of the noninvasive spectra collected. A cross-validation PLS calculation was performed with the absorbance spectra, leaving out 10 % of the data each cycle as the internal prediction set. The wavelength range was 4900-4200 cm^{-1} and 1-16 PLS factors were used. The glucose concentration profile that gave the best SECV value was assumed to correspond to the optimal delay time assignment. For each delay time, the SECV values were plotted with respect to the number of latent variables in the PLS model to determine the minimum SECV value. This plot is shown in Figure 7.10. The delay time corresponding to the minimum SECV was taken as the optimal time that is required for the arterial blood to be equilibrated in the skin tissue matrix. For rat 106, the optimal lag time was found to be 11 minutes and the glucose concentrations were assigned to each spectrum on the basis of this time.

The glucose concentration profile was then partitioned into a training set, a monitoring set and a prediction set as shown in Figure 7.11. The training set, monitoring set and the prediction set contained 164, 83 and 199 single-beam-spectra, respectively. All the single-beam spectra were collected continuously in the same day. Assuming that the background information is the same throughout and that path length variation is minimal, all the calibration single-beam spectra were put into a single block for calculating differential spectra.

Ratios of all the combinations of single-beam spectra in the training set were computed and the resulting differential spectra and corresponding differential

concentrations were used to optimize the wavenumber and number of PLS factors that could be applied to the alarm classification of the monitoring set. Here, the monitoring set was a subset of the calibration set to test the parameters for PLS that would subsequently be used in the implementation of the alarm algorithm with the prediction set. The calculation of differential spectra for the training set produced 13,466 total differential spectra. A grid search analysis using the same protocol described previously in Chapters 4 and 6 was performed to identify the best wavenumber and latent variable combinations.

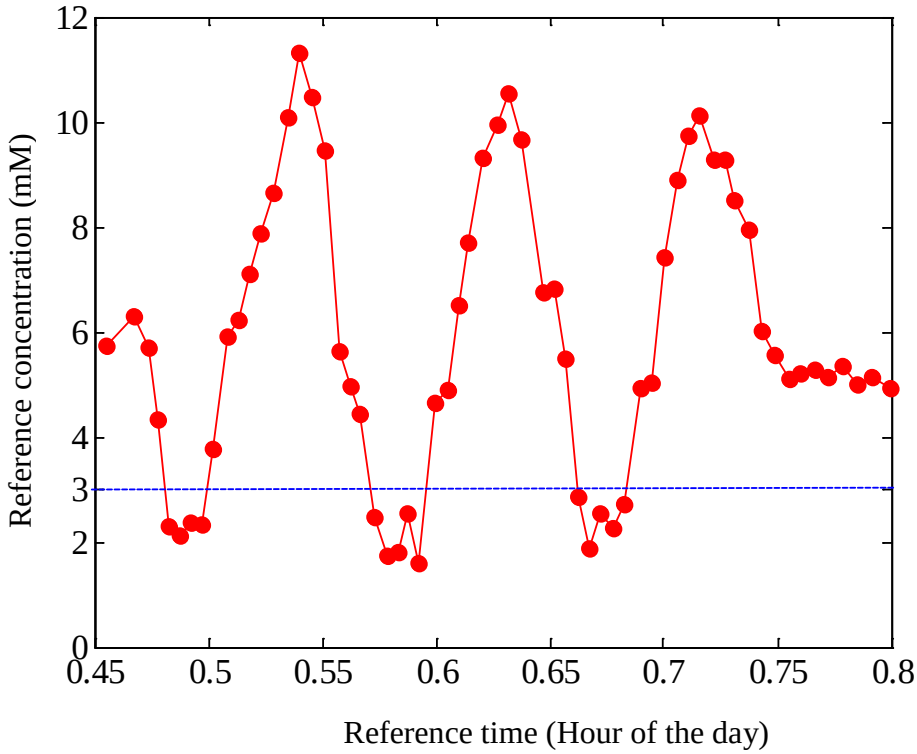


Figure 7.9. Glucose concentration profile for rat 106. An interpolation was required to assign a glucose concentration to each of the spectra collected. the horizontal line denotes the alarm threshold of 3.0 mM. Reference times are given in fractions of a day.

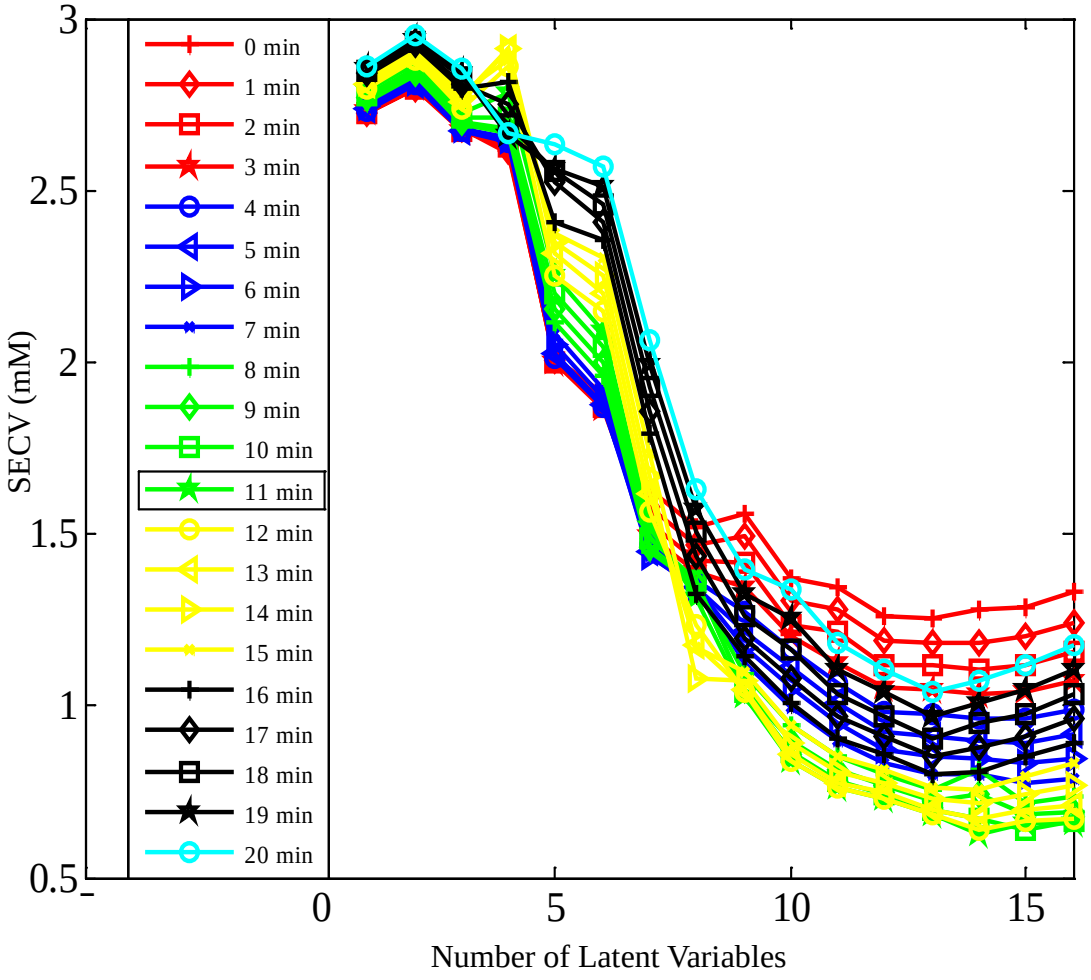


Figure 7.10. Plot of SECV vs. the number of PLS latent variables for rat 106. The best lag time was 11 minutes based on the minimum SECV value of 0.63 mM.

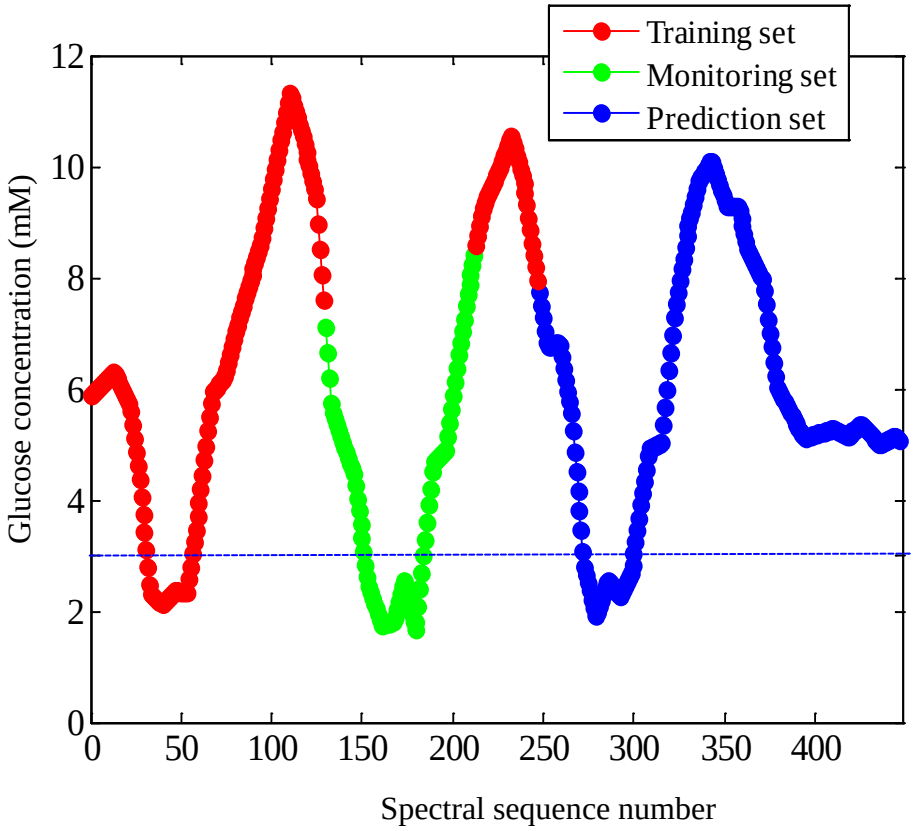


Figure 7.11. Lag-corrected glucose concentration profile for rat 106 employing the selected delay time of 11 minutes. The concentration profile was partitioned into a training set, a monitoring set and a prediction set. The horizontal line identifies the alarm threshold of 3.0 mM.

The grid search was based on sliding a window of fixed spectral width in 25 cm^{-1} increments across the 4900-4100 cm^{-1} range. At each step, PLS models were constructed using 3-16 LVs. The starting width of 100 cm^{-1} was incremented in 25 cm^{-1} increments up to a maximum width of 700 cm^{-1}. At each step, the cross-validation procedure was applied, withholding 10 % of the data for prediction. Table 7.1 summarizes the four optimal wavenumber ranges and the relevant LVs (*F*-test corrected) that resulted from the grid search analysis.

Figure 7.12 displays a plot of SECV vs. the number of latent variables for the top wavenumber range (4875-4175 cm^{-1}) that resulted from the grid search analysis. As observed in previous chapters, the SECV is changing very little by the time the model size reaches 16 factors and thus the consideration of larger numbers of latent variables was deemed unnecessary.

The top four wavenumber ranges listed in Table 7.1 and latent variables 9-14 were then used to build classification models with PLDA. The PLS calculation was applied to all the wavenumber and PLS factor combinations to produce the corresponding sets of scores. The critical concentration of the monitoring set was then used to partition the differential spectra in the training set into alarm and non-alarm classes for use with PLDA.

Table 7.1. Top four wavenumber-latent variable combinations based on minimum SECV values from the grid search analysis of training differential spectra

Wavenumber (cm^{-1})	Latent variables	SECV (mM)
4875-4175	16	0.38_8
4850-4175	16	0.40_2
4900-4200	16	0.40_8
4900-4300	16	0.41_2

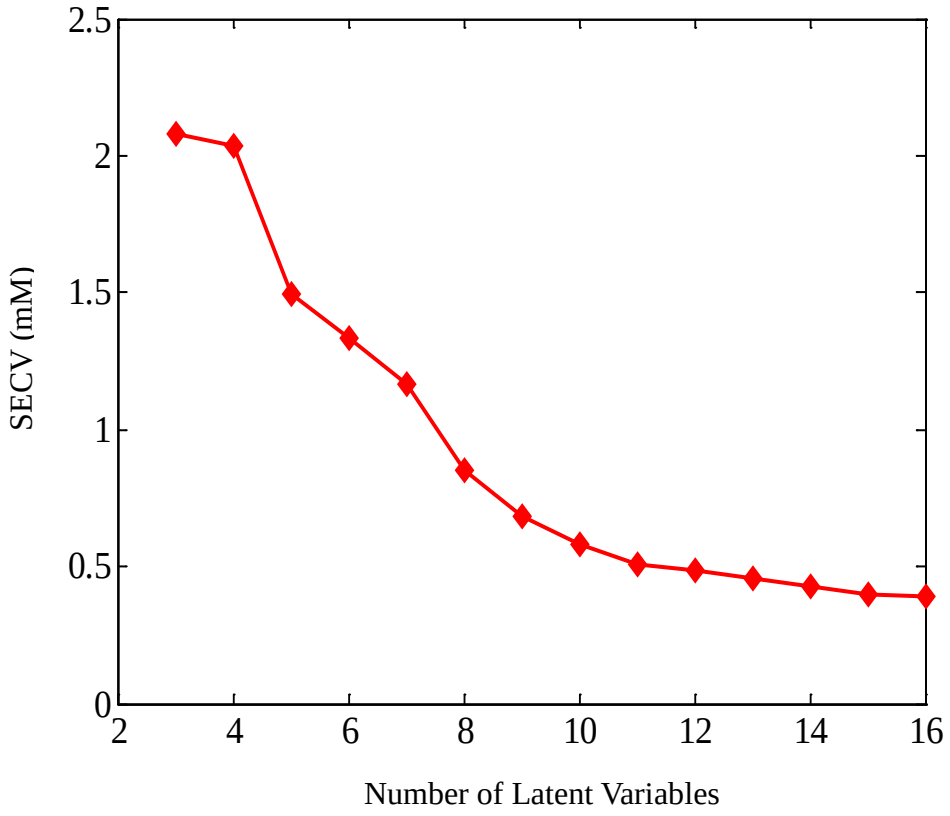


Figure 7.12. Plot of SECV vs. the number of PLS latent variables for the 4875-4175 cm^{-1} range. This is the top wavenumber range that resulted from the grid search analysis. The minimum SECV was 0.39 mM.

The reference concentration, the first concentration in the monitoring set, was 7.12 mM and the critical concentration was –4.12 mM for an alarm threshold of 3.0 mM. There were 4234 alarm patterns and 9132 non- alarm patterns in the training set. Table 7.2 summarizes the percentage (average ± standard deviation) of missed and false alarms when PLDA was employed. Three replicate classifiers were computed and each classifier was based on three discriminant functions. The classifiers were applied to the 30 alarm and 50 non-alarm patterns in the monitoring set. The optimal wavenumber-latent variable

combination was 4900-4200 cm^{-1} and 10 latent variables based on the minimum percentage of missed and false alarms.

Table 7.2. Average percentage of missed and false alarms for the monitoring set

Spectral range, cm^{-1}		Latent variables					
		9	10	11	12	13	14
4875-4175	AM (%)[a] ± SD	53.5± 7.0	59.6 ± 12.6	68.7 ± 9.7	62.6 ±12.6	45.4 ± 10.9	47.5 ± 9.74
	AF (%)[b] ± SD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[c] ± SD	69.0 ± 2.7	71.8 ± 3.8	74.1 ± 0.7	76.3 ± 1.0	77 1± 0.9	79.4 ± 4.8
4850-4175	AM (%)[a] ± SD	54.5 ± 0.0	48.5 ± 0.0	66.7 ± 6.07	84.8 ± 16.9	47.5± 7.00	54.5 ±18.2
	AF (%)[b] ± SD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[c] ± SD	71.4 ± 1.0	69.6 ± 0.9	72.7 ± 1.1	76.7 ± 0.5	75.5 ± 0.9	76.3 ± 0.6
4900-4200	AM (%)[a] ± SD	51.5 ± 3.03	19.2 ± 3.50	32.3 ± 9.26	47.5 ± 9.26	89.9 ± 1.75	52.5 ± 24.3
	AF (%)[b] ± SD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[c] ± SD	68.7 ± 2.2	72.9 ± 2.8	75.6 ± 1.4	78.2 ± 0.7	78.1 ± 1.5	75.9 ± 5.0
4900-4300	AM (%)[a] ± SD	30.3 ± 0.0	25.2 ± 3.50	19.2± 14.0	36.4 ± 9.1	42.4 ± 3.03	57.6 ± 10.5
	AF (%)[b] ± SD	1.33 ± 1.15	1.33 ± 1.15	4.00 ± 2.83	0.67± 0.94	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[c] ± SD	80.0 ± 0.6	69.9 ± 0.9	74.3 ± 2.3	75.7 ± 1.9	80.0 ± 0.7	80.2 ± 1.4

[a]AM (%) ± SD: Average percentage of missed alarms ± standard deviation.

[b]AF (%) ± SD: Average percentage of false alarms ± standard deviation.

[c]DS (%) ± SD: Average percentage of separated alarm patterns with three discriminants (DS) ± standard deviation.

For application to the prediction set, the training set and monitoring set were then combined to form an overall calibration set which consisted of 247 single-beam spectra. The reference concentration for the prediction set, alarm threshold value and the critical concentration were 7.72, 3.00 and -4.72 mM, respectively. A total of 30,381 differential spectra were generated from the calibration data and the PLS scores computed with the

optimal wavenumber-latent variable combination were partitioned into 8346 alarm and 22,035 non-alarm patterns. Three replicate classifiers were computed with the calibration PLS patterns.

The prediction set contained 27 alarm and 172 non-alarm patterns. Figure 7.13 shows the prediction glucose concentration profile for rat 106. Three discriminants from each of the replicate classifiers were applied to classify the alarm and non-alarm patterns in the prediction set. The prediction result for the three replicate discriminants yielded 4, 2 and 0 missed alarms and 3, 8 and 14 false alarms. The committee decision rule used previously in Chapters 4 and 6 was applied to these results (i.e., two out of three discriminant scores must be positive for an alarm detection). When the committee rule was applied, there were 2 missed alarms out of 27 alarm patterns and 8 false alarms out of 172 non-alarms patterns. This corresponded to the successful detection of 92.6 % of the alarm events with a false alarm rate of 4.7%.

The discriminant score plot for the prediction set is shown in Figure 7.14. The concentration trend in the prediction profile is clearly explained by the discriminant score plot. This dependence of discriminant scores on the glucose concentration further confirms the data classification by PLDA.

A comparison between the discriminant score plot in Figure 7.14 and the prediction concentration profile in Figure 7.13 confirms that the false and missed patterns correspond to concentrations close to the alarm threshold. Small differences in concentrations at the alarm threshold are challenging to detect because the classification of these patterns is extremely sensitive to the positioning of the separating surface

between the data classes. This factor explains why benefit is gained from training replicate classifiers and using them collectively to make the classification.

Another factor that must be considered is imprecision in the assignment of the reference concentrations. As described previously, concentrations were assigned to each spectrum on the basis of interpolating the reference concentration measurements. Imprecision in these assignments can also lead to what appear to be missed or false alarms at the alarm threshold. Imprecision in the determination of the concentration lag can have a similar effect.

The scores along the first two PLS latent variables are plotted together for the prediction set in Figure 7.15. Good separation between alarm and non-alarm patterns is noted. A further test of the selection of the spectral range and number of latent variables was obtained by performing a cross-validation calculation on all the absorbance spectra for rat 106 over the range of 4900-4200 cm^{-1}. Leaving out 10 % of the data for prediction at each cycle in cross-validation produced a SECV value of 0.84 mM for the 10-factor model. The time-based profiles of cross-validated predicted glucose concentrations and assigned reference concentrations shown in Figure 7.16 A and the correlation plot in Figure 7.16 B show good correlation and further validate the selected model parameters.

The absorbance measurements are highly sensitive to the optical path length which is significantly affected by different skin thicknesses, pressure from the interface, and even slight movement in the sensing location. In addition, changes in skin hydration and dehydration during the measurement can vary the optical path length which can lead to erroneous predictions. To better understand the skin tissue variance during the data collection, a simple linear regression using Eq. 5.1 was performed using four known pure

components, water, collagen, keratin, and fat, to estimate the thickness of the skin components and aqueous optical path length. In addition to these four chemical components, a constant term was included to compensate for scattering effects.

When each pure-component spectrum is normalized to a thickness of 1 mm, the regression coefficients represent the physical thickness of each of these pure components in the tissue matrix. The regression coefficients for each of the spectra collected for rat 106 are plotted in Figure 7.17. It can be clearly observed that the regression coefficients of each component have different magnitudes of variations during the data collection period. These fluctuations could be due to different reasons, including the interface pressure applied to the skin tissue, slight variations in the tissue components, environmental variations, such as temperature and humidity, unconscious body movements, etc. For example, a gradual decline in the thickness of the water layer is observed over the time course of the experiment. This suggests the pressure applied by the interface is slowly pushing water out of the optical path.

The PLS method must be applied with precautions against overmodeling. One diagnostic to employ in this regard is to examine the computed PLS loading weights and spectral loadings to verify that the captured information is relatively devoid of noise-like structure. Given that NIR spectral features have known widths, any extremely high-frequency signatures are likely based on noise features. The incorporation of such features into the loading weights or spectral loadings is a sign of overfitting (i.e., noise features that correlate by chance with the dependent variable of concentrations have been extracted into the PLS factors).

The eighth, ninth and tenth loading weights and spectral loadings computed from the calibration differential spectra are shown in Figures 7.18 and 7.19, respectively. Only very minor contamination of these factors by noise-like structures is apparent. This result helps to confirm the use of 10 PLS factors in the classification model.

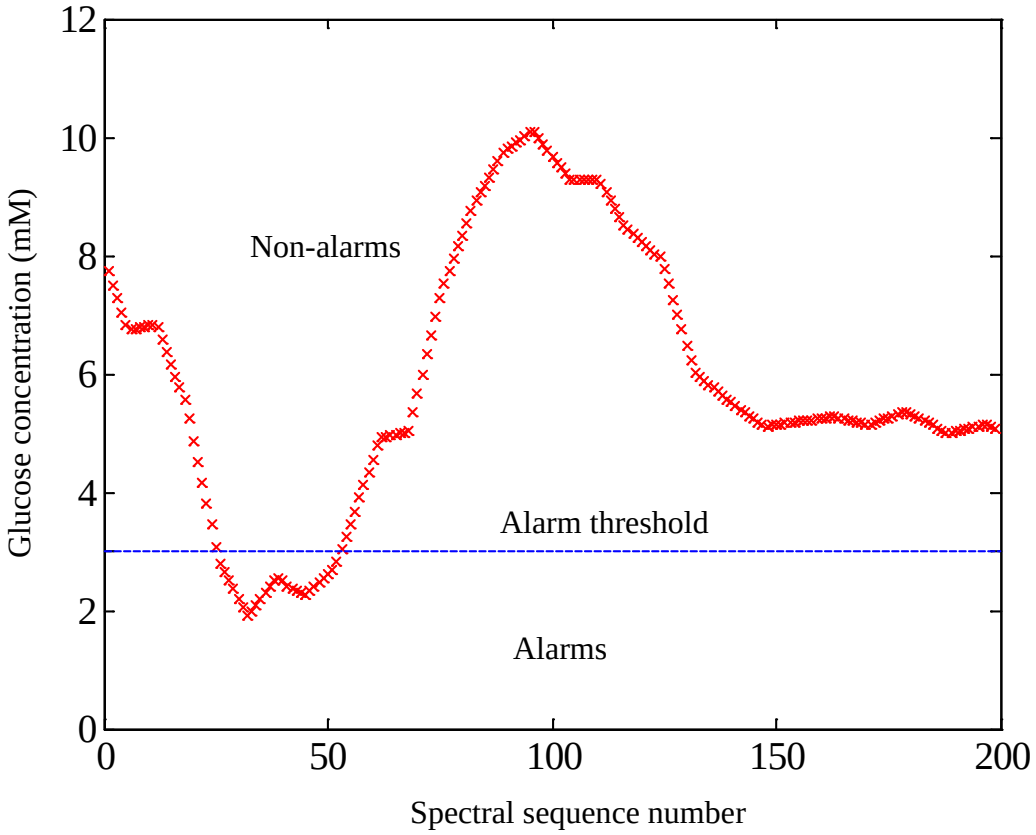


Figure 7.13. Prediction glucose concentration profile for rat 106. There are 27 alarm patterns and 172 non-alarm patterns. The alarm threshold of 3.0 mM is shown by the horizontal line.

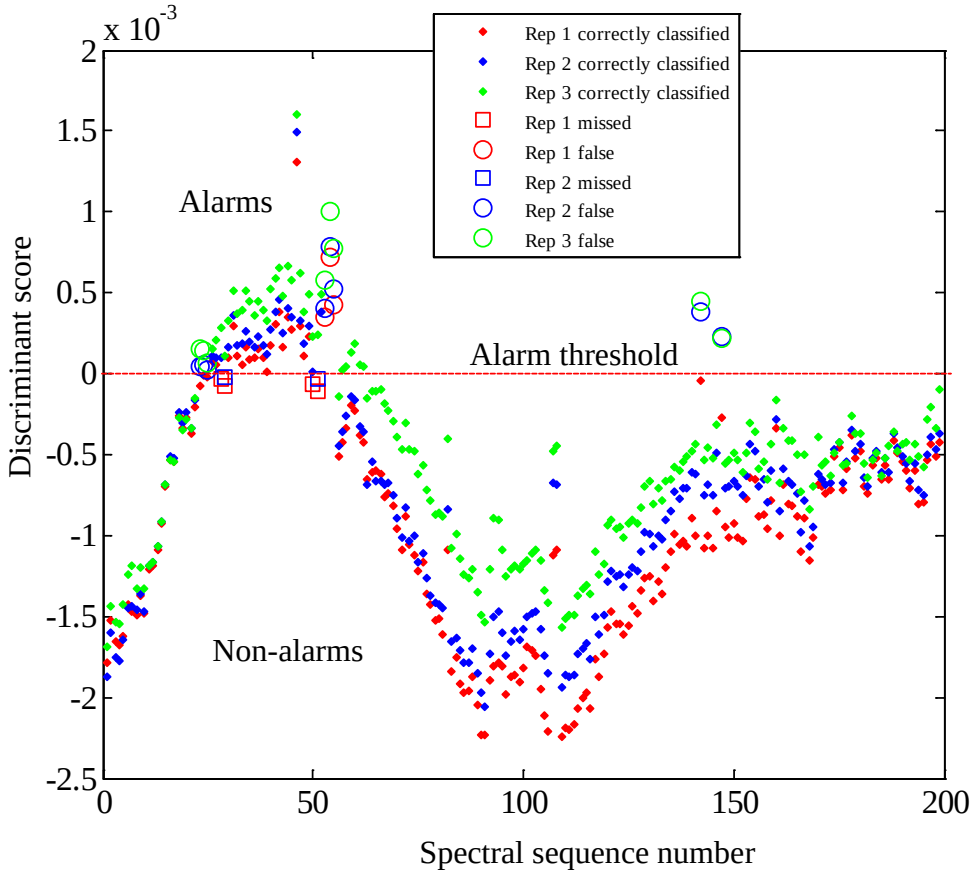


Figure 7.14. Discriminant score plot for the prediction set with three replicate classifiers. Missed and false alarms for each classifier are indicated.

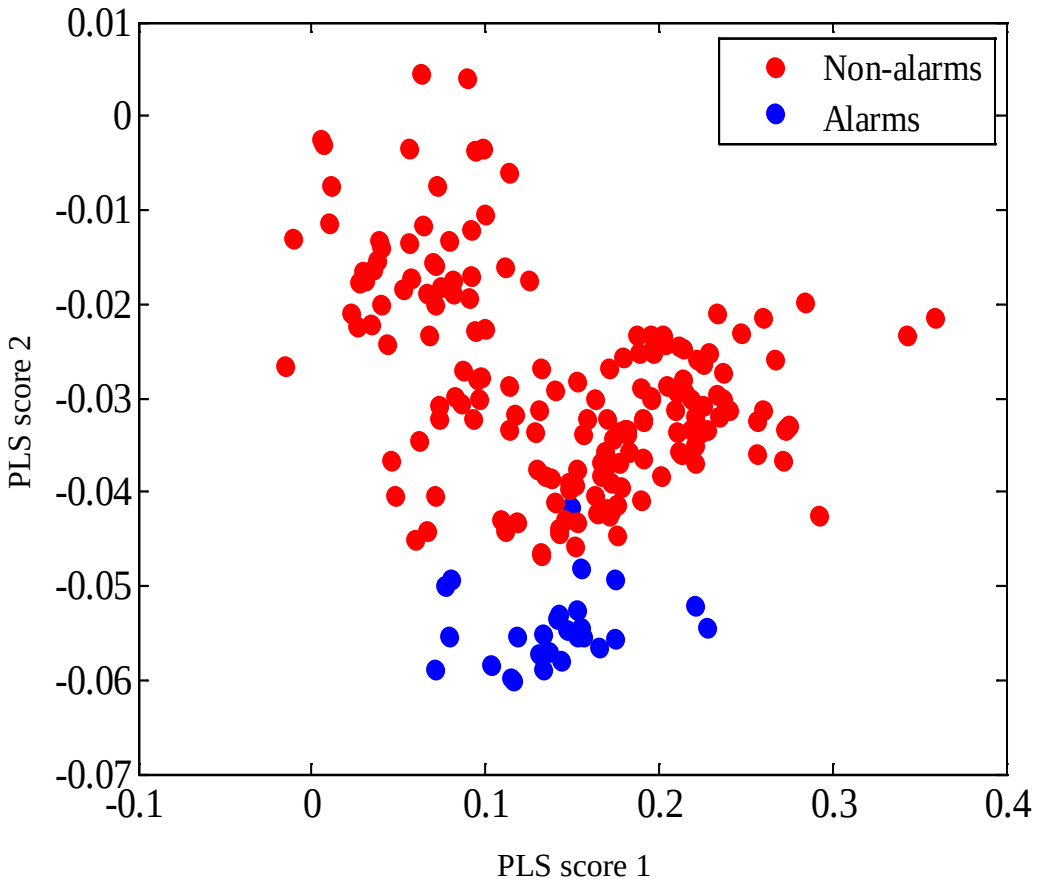


Figure 7.15. Plot of first two PLS scores for the prediction set for rat 106. The wavenumber range was 4900-4200 cm^{-1}. Good separation between alarm and non-alarm patterns is noted.

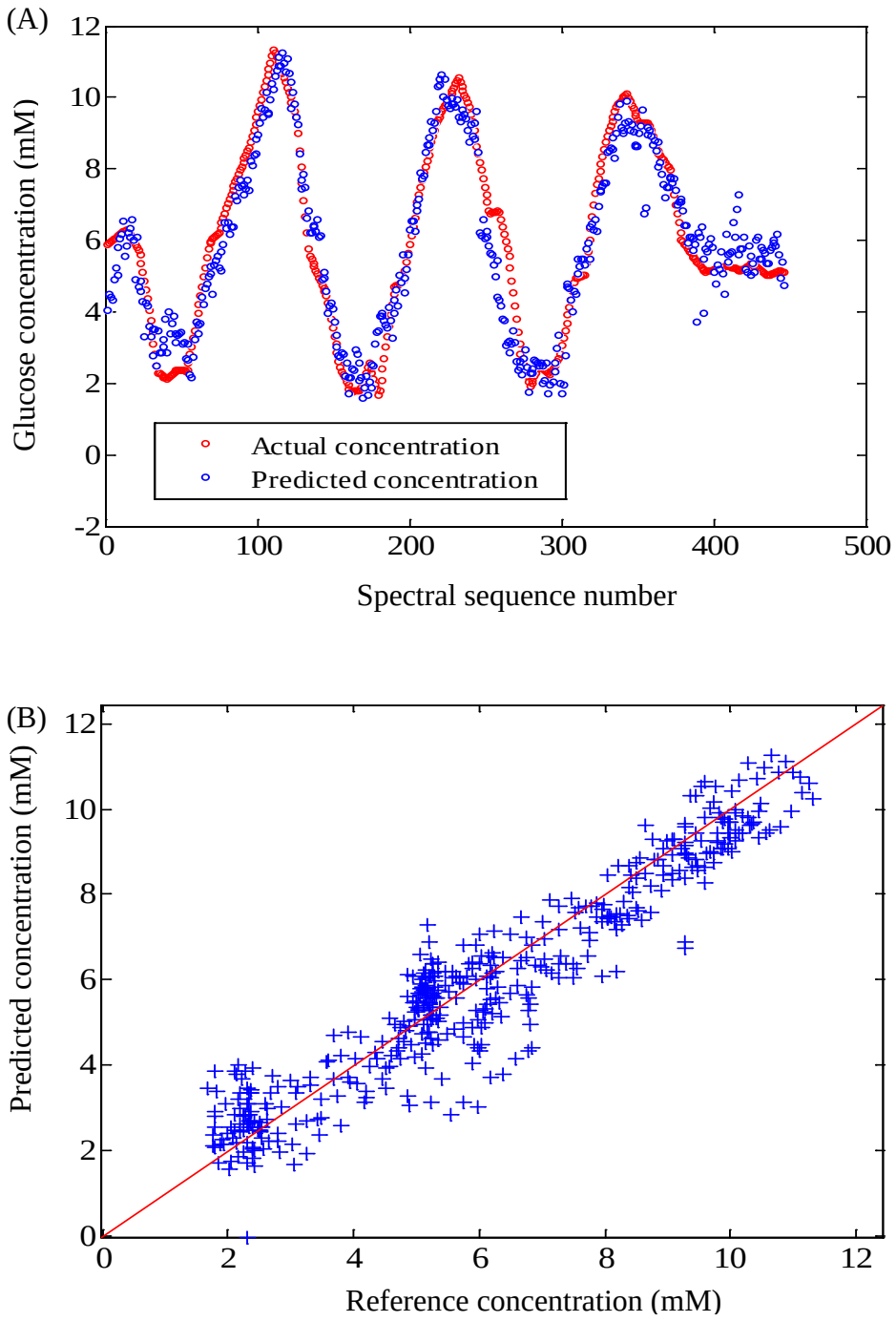


Figure 7.16. A. Cross-validated predicted and assigned reference glucose concentrations are plotted with respect to time (spectral sequence number) for rat 106. Ten PLS factors and a wavenumber range of 4900-4200 cm^{-1} were used for the cross-validation. B. Correlation plot of cross-validated predicted vs. reference concentrations.

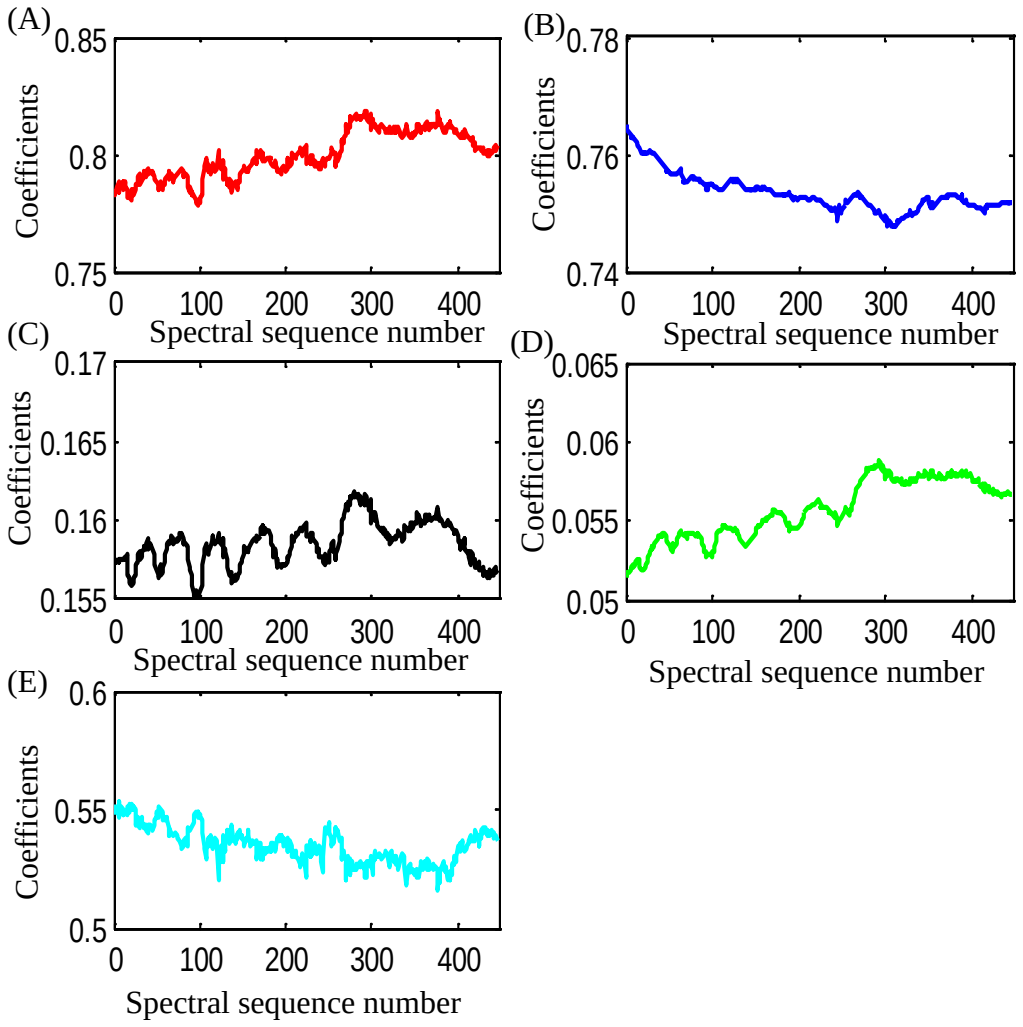


Figure 7.17. Regression coefficients computed for each of the spectra collected for rat 106. Panels A-E correspond to regression coefficients for collagen, water, keratin, fat, and an intercept term. The regression coefficients represent the thickness of each component in the skin tissue matrix in units of mm.

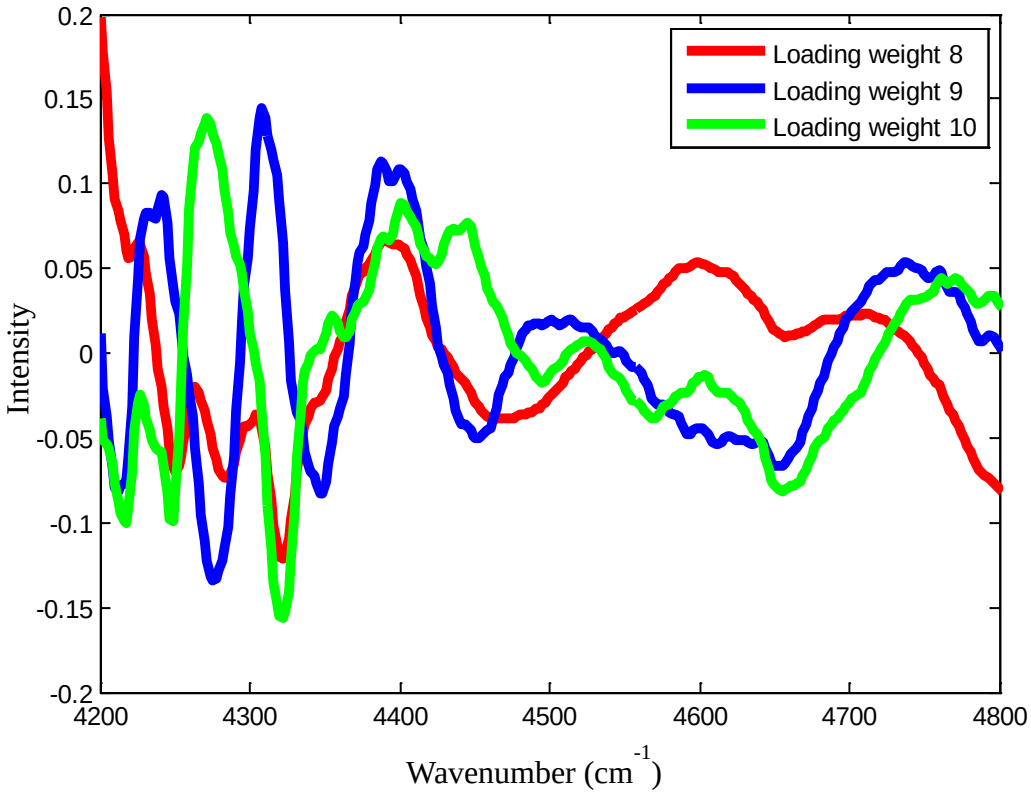


Figure 7.18. Loading weights computed from the calibration differential spectra. The eighth, ninth and tenth loading weights show spectral features, not noise. The 4900-4200 cm^{-1} range was used.

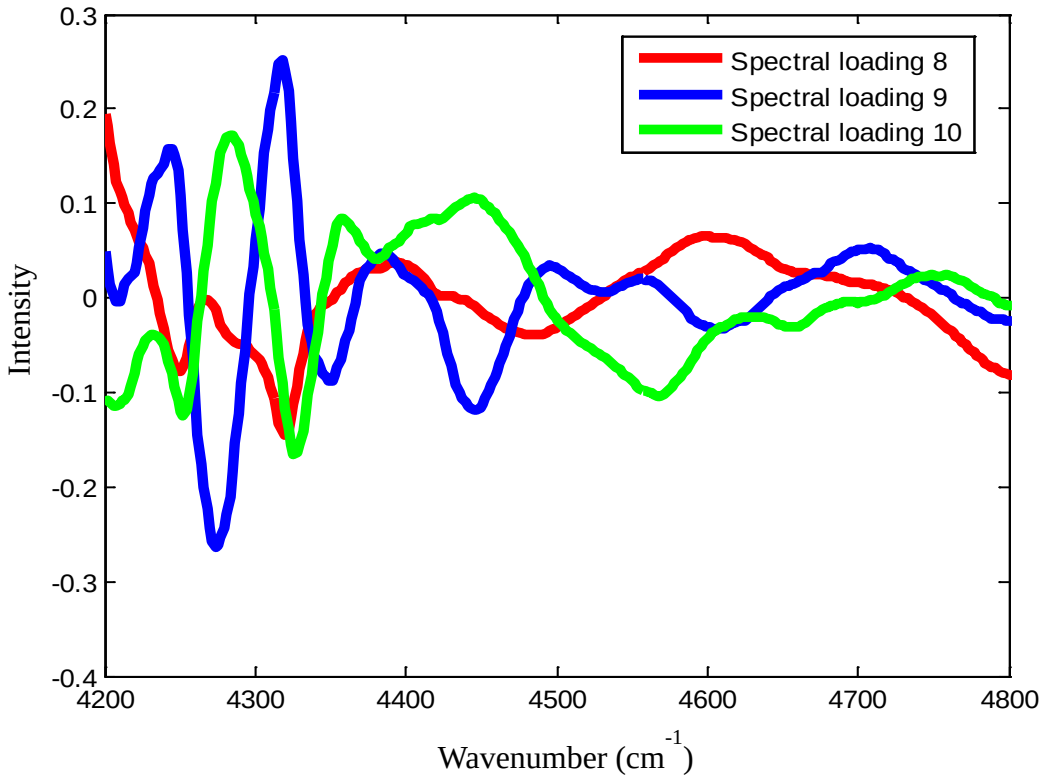


Figure 7.19. Spectral loadings computed from the calibration differential spectra. The eighth, ninth and tenth spectral loadings show spectral features, not noise. The 4900-4200 cm^{-1} range was used.

Analysis of Rat 114

Rat 114 was a male rat weighting around 340 g. A non-survival surgery for the rat was performed and the glucose transient study for hypoglycemia was performed on the same day. Spectral data for rat 114 were collected on the basis of 32 averaged scans. Assuming random noise, the SNR can be increased by a factor of 2 when four

consecutive spectra are averaged. For the subsequent data analysis, four consecutive single-beam spectra were averaged to increase the SNR.

The short-term noise for rat 114 was evaluated for the averaged single-beam spectra by taking the ratio of pairs of consecutive scans and converting the resulting transmittance data to absorbance. As described previously, the wavenumber region of 4300-4500 cm^{-1} was used, and the RMS noise was calculated about a third-order polynomial fit. The mean noise value was approximately 65 μAU. A plot of the computed RMS noise values is provided in Figure 7.20. These noise values are approximately three times higher than those reported above for rat 114. Reasons for this increase in noise will be explored at the end of this section.

As performed previously in the analysis of rat 106, the quality of the data was initially studied by performing PCA on the absorbance spectra computed relative to an average air reference collected at the beginning and end of the of the experiment. The skin tissue absorbance spectra were mean-centered before PCA was performed, and the spectral range was restricted to 4900-4200 cm^{-1}. The first three principal component scores are plotted together in Figure 7.21. It can be seen that spectra 1-7 are separated from the rest. This can be more clearly observed in the Q vs. T^2 plot for the first three principal components presented in Figure 7.22. On the basis of these results, the first seven signal-averaged spectra were removed from the subsequent analysis.

The glucose concentration profile for rat 114 is shown in Figure 7.23. A reference concentration was assigned to each of the spectra by interpolation of the reference measurements as discussed previously. Assuming a delay time of 0-20 min, a glucose concentration was assigned to each of the noninvasive single-beam spectra

collected at 32 scans. A cross-validation was performed with the absorbance spectra as described previously. Figure 7.24 plots the computed SECV values with respect to the number of PLS latent variables. The delay time corresponding to the minimal SECV value was taken as the optimal time required for the arterial blood to be equilibrated in the skin tissue matrix. For rat 114, the optimal lag time was found to be 7 minutes and the glucose concentrations were assigned to each single-beam spectrum on the basis of this time.

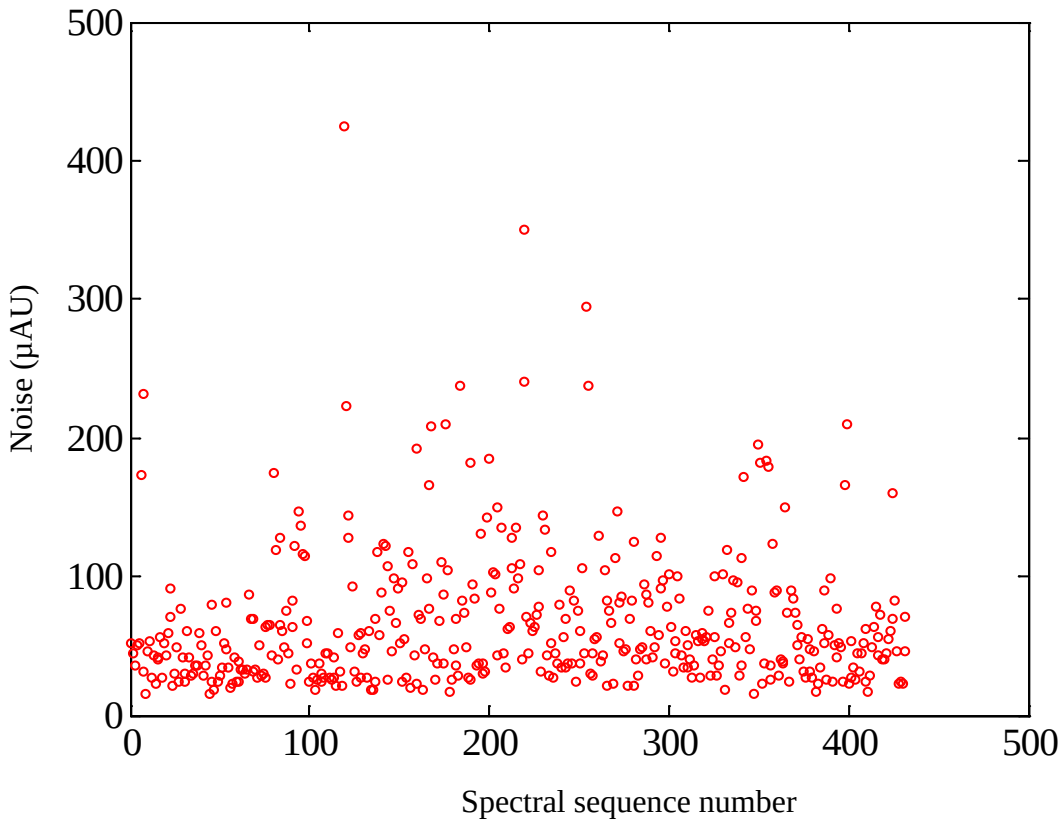


Figure 7.20. Values of RMS noise in µAU computed for each spectrum of rat 114 over the 4500-4300 cm^{-1} region. Noise values were computed by fitting the 100% line in the given wavenumber range to a third-order polynomial and calculating the noise about the fitted value.

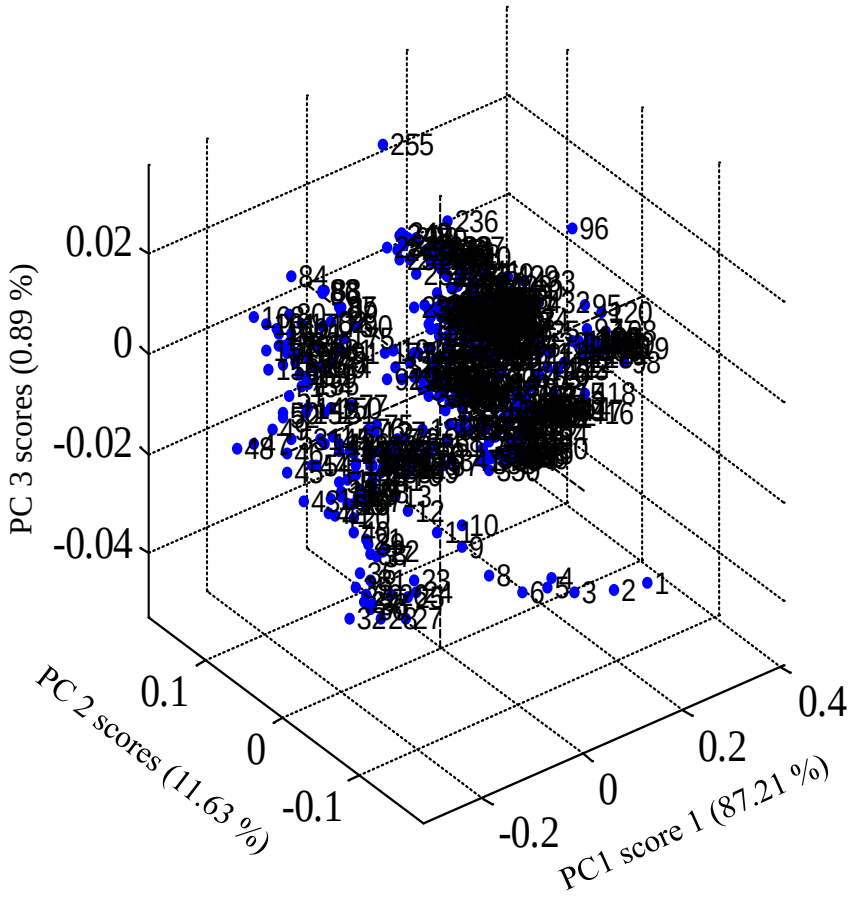


Figure 7.21. Principal component score plot for rat 114 based on the first three latent variables. The absorbance rat skin tissue spectra in the wavenumber range of 4900-4200 cm^{-1} were used for the analysis. Points are labeled according to their time sequence.

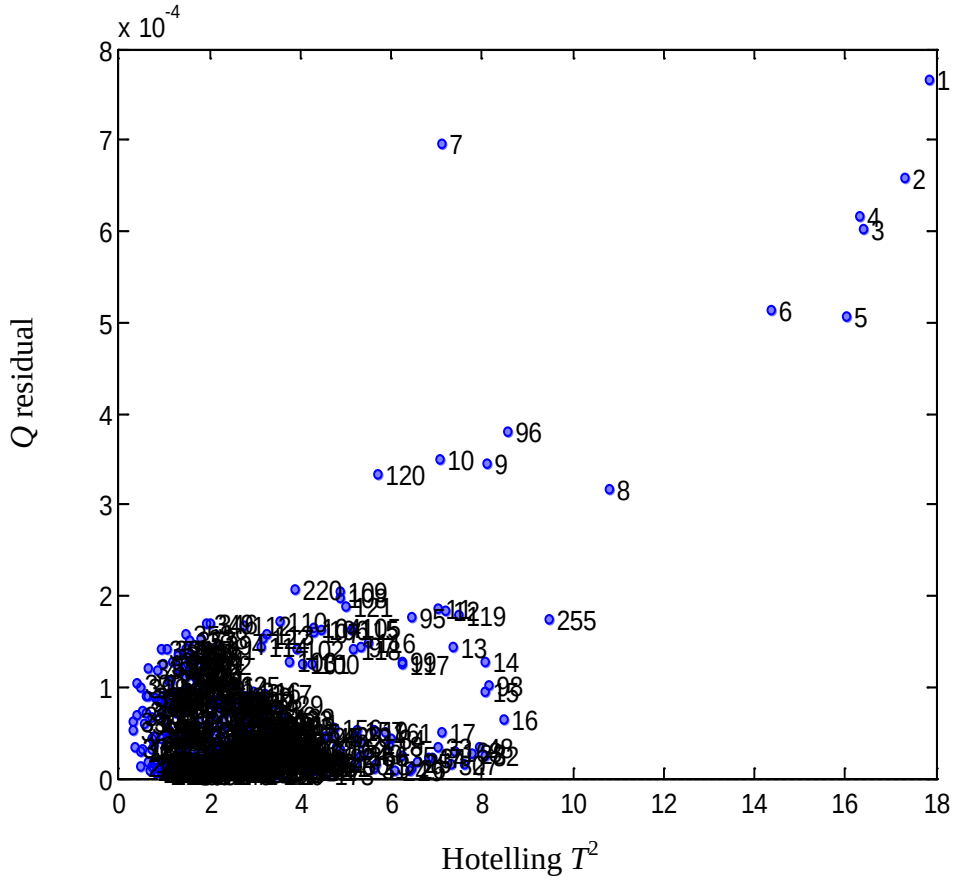


Figure 7.22. Plot of Q vs. Hotelling's T^2 for the absorbance spectra from rat 114. The first three principal components were used for the analysis. Spectra 1-7 are judged to be extreme outliers.

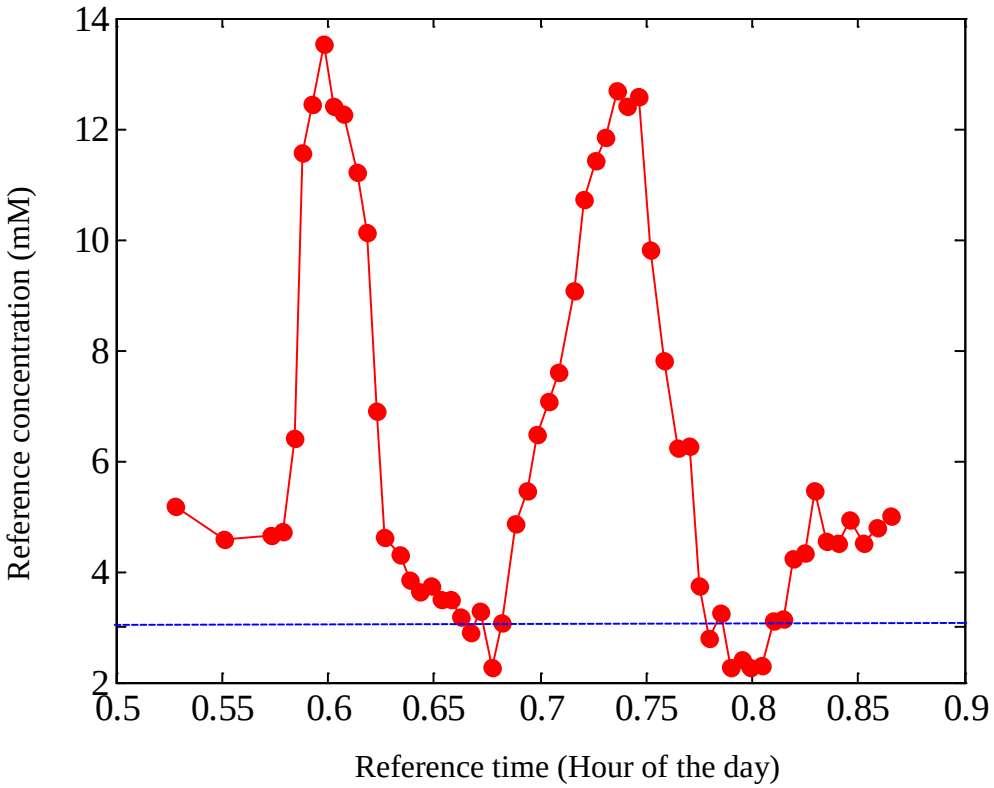


Figure 7.23. Glucose concentration profile for rat 114. An interpolation is required to assign a glucose concentration to each of the spectra collected. The horizontal line denotes the alarm threshold of 3.0 mM. Reference times are given in fractions of a day.

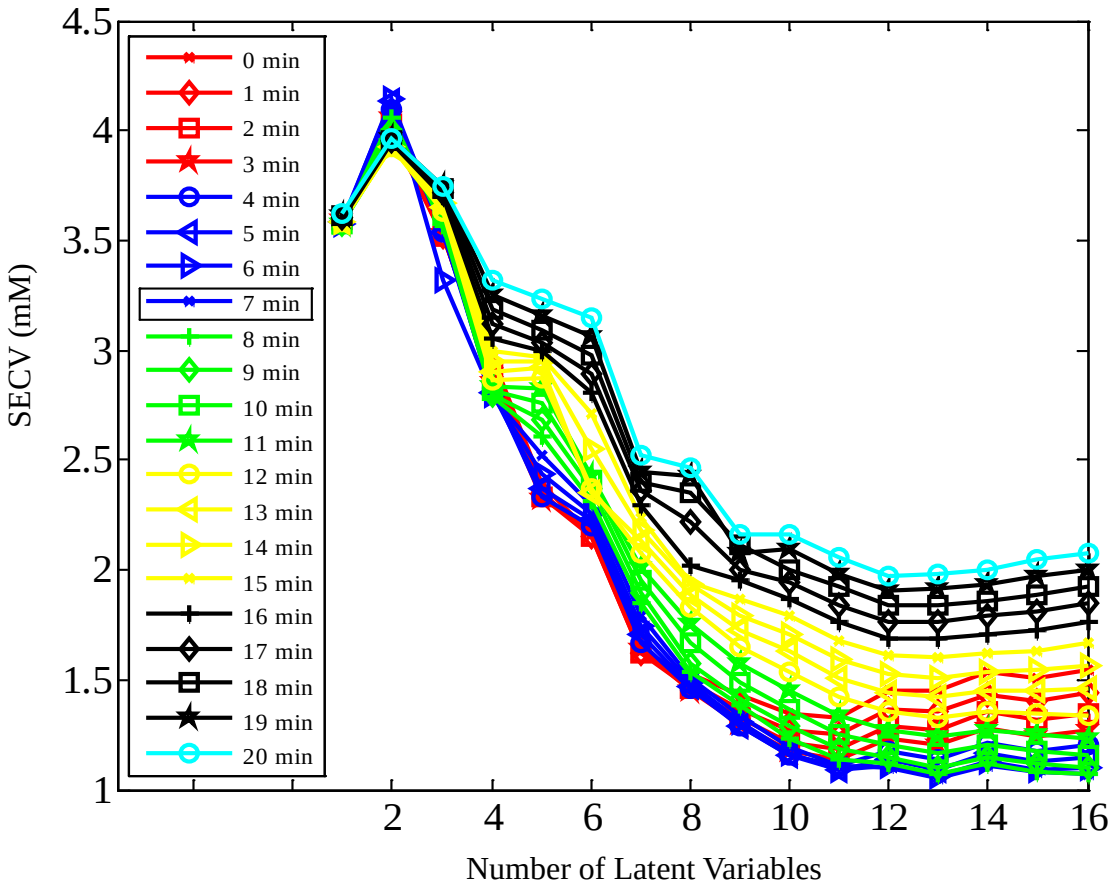


Figure 7.24. Plot of SECV vs. latent variables for rat 114. The best lag time was 7 minutes based on the minimum SECV value of 1.05 mM.

The glucose concentration profile was then partitioned into a training set, a monitoring set and a prediction set as shown in Figure 7.25. The training set, monitoring set and the prediction set contained 178, 49 and 199 signal-averaged single-beam spectra, respectively. The training set yielded 15,753 differential spectra. As described previously, a grid search analysis based on cross-validation was performed on the differential spectra

to identify the best possible wavenumber and latent variable combinations for further study. The grid search was based on sliding a window of fixed spectral width in 50 cm^{-1} increments across the 4900-4100 cm^{-1} range. Widths employed were 100 to 700 cm^{-1} in 50 cm^{-1} increments. The top four wavenumber ranges and associated numbers of latent variables identified by the grid search are shown in Table 7.3. Figure 7.26 plots values of SECV vs. the number of latent variables for the top wavenumber range of 4900-4200 cm^{-1}.

The spectral ranges listed in Table 7.3 and latent variables 13-16 were then used to build classification models with PLDA for implementation of the alarm algorithm with the monitoring set. Partial least-squares scores were calculated with all the wavenumber-PLS factor combinations and were partitioned into alarm and non-alarm patterns based on the critical concentration. The reference, critical and alarm threshold concentrations for the monitoring set were 5.48, – 1.08 and 4.40 mM, respectively. The alarm threshold was set higher in this case than the value of 3.0 mM used previously because of the small number of spectra below 3.0 mM in the monitoring set.

There were 9,601 alarm and 6,152 non-alarm patterns in the training set. Table 7.4 summarizes the percentage (average ± standard deviation) of missed and false alarms when three replicate classifiers were applied to the 26 alarm and 23 non-alarm patterns in the monitoring set. Each classifier was based on two discriminant functions. A third discriminant was not used here because it did not separate a significant number of patterns in the training set. The optimum wavenumber-latent variable combination was 4850-4250 cm^{-1} and 14 latent variables based on the minimum percentage of missed and false alarms.

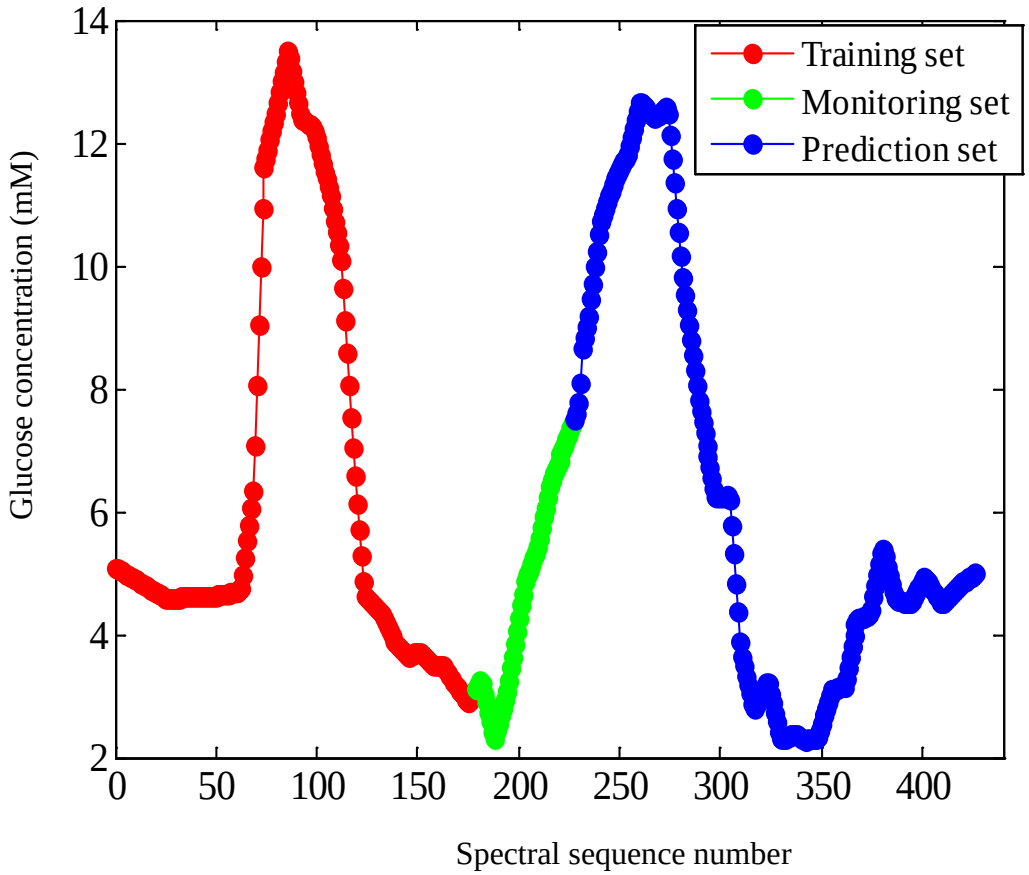


Figure 7.25. Lag-corrected glucose concentration profile for rat 114. The concentration profile was partitioned into training, monitoring, and prediction sets.

Table 7.3. Top four wavenumber-latent variable combinations based on minimum SECV values from the grid search analysis of training differential spectra

Wavenumber (cm^{-1})	Latent variables	SECV (mM)
4900-4200	16	0.51_1
4900-4300	16	0.51_2
4850-4300	16	0.52_3
4850-4250	16	0.53_8

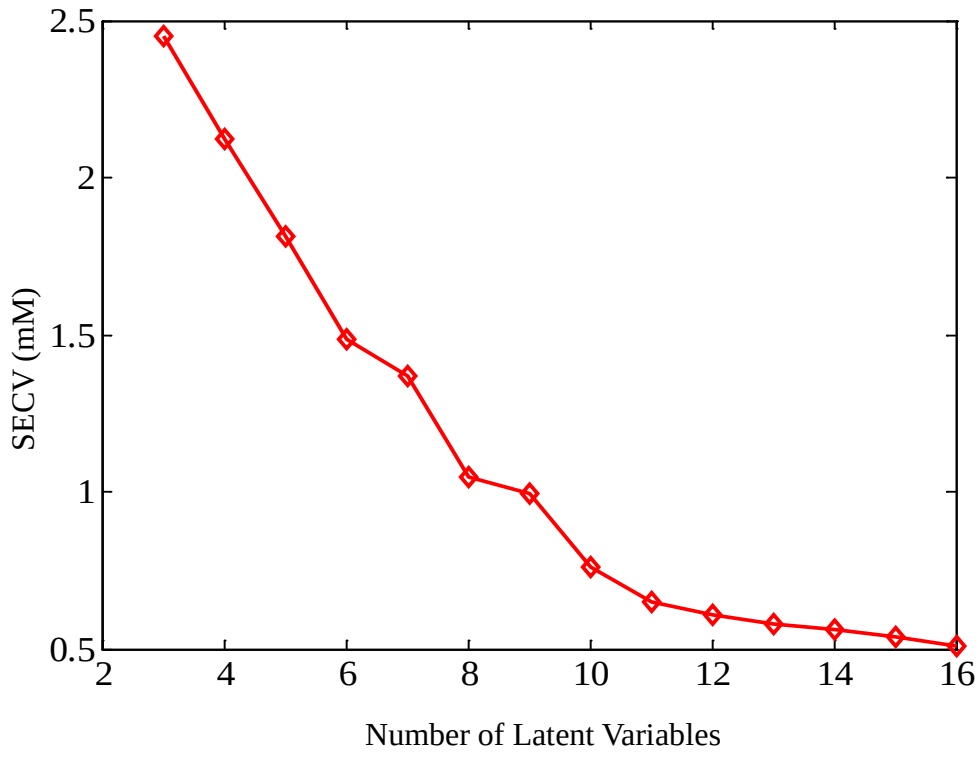


Figure 7.26. Plot of SECV vs. latent variables for the top wavenumber range (4900-4200 cm^{-1}) that resulted from the grid search. The minimum SECV was 0.51 mM.

Table 7.4. Average percentage of missed and false alarms for the monitoring set

Spectral range ,cm^{-1}	Latent variables				
		13	14	15	16
4900-4200	AM (%)[a] ± SD	100 ± 0.0	23.1 ± 20.4	24.4 ± 42.2	41.0 ± 37.0
	AF (%)[b] ± SD	0.0 ± 0.0	11.6 ± 16.5	5.8 ± 10.0	11.6 ± 20.1
	DS (%)[c] ± SD	80.0 ± 1.9	79.9 ± 1.4	80.9 ± 2.1	80.2 ± 0.4
4900-4300	AM (%)[a] ± SD	65.4 ± 32.9	71.8± 17.3	48.7 ± 8.90	57.0 ±37.0
	AF (%)[b] ± SD	2.9 ± 2.5	2.9 ± 2.5	4.3 ± 0.0	4.3 ± 7.5
	DS (%)[c] ± SD	78.7 ± 0.8	83.2 ± 4.3	78.2 ± 3.6	81.5 ± 2.5
4850-4300	AM (%)[a] ± SD	100 ± 0.0	100 ± 0.0	100 ± 0.0	47.4 ± 42.2
	AF (%)[b] ± SD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[c] ± SD	83.1 ± 0.3	75.9 ± 0.8	81.8 ± 0.3	82.1 ± 1.1
4850-4250	AM (%)[a] ± SD	61.5± 30.0	29.5± 21.9	100 ± 0.0	100 ± 0.0
	AF (%)[b] ± SD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[c] ± SD	78.8 ± 2.2	82.7 ± 0.9	80.8 ± 0.8	83.3 ± 0.8

[a]AM (%) ± SD: Average percentage of missed alarms ± standard deviation.

[b]AF (%) ± SD: Average percentage of false alarms ± standard deviation.

[c]DS (%) ± SD: Average percentage of separated alarm patterns with two discriminants (DS) ± standard deviation.

The training set and monitoring set were then combined to build a calibration set which consisted of 227 averaged single-beam spectra. A total of 25,651 differential spectra were produced and used to compute the PLS scores that formed the patterns.

The alarm algorithm was then implemented with the prediction set. The PLS scores computed with the optimal wavenumber-latent variable combination were partitioned into 8008 alarm and 17,643 non-alarm patterns. The reference concentration, alarm threshold value and the critical concentration for the prediction set were 7.21, 3.00

and -4.21 mM, respectively. Three replicate classifiers were computed with the calibration PLS patterns.

The prediction set contained 34 alarm concentrations and 165 non-alarm concentrations. Figure 7.27 shows the prediction glucose concentration profile for rat 114. The two discriminant functions from each of the replicate classifiers were applied to classify the alarm and non-alarm patterns in the prediction set. The prediction results for the three replicate discriminants were 4, 6 and 6 missed alarms and 29, 16 and 14 false alarms. Based on the alarm decision rule described previously, there were 6 missed alarms out of 34 alarm patterns and 16 false alarms out of 165 non-alarms patterns. This corresponded to the successful detection of 82.4 % of the alarm events and a false alarm rate of 9.7 %. This performance is worse than that observed previously with rat 106.

The discriminant score plot for the prediction set is shown in Figure 7.28. The dependence of the discriminant scores on the concentrations can also be observed for rat 114. As observed previously with rat 106, the majority of the missed and false alarms occur in the region near the alarm threshold.

The first three PLS scores for the prediction set are plotted in Figure 7.29. Good separation between the alarm and non-alarm patterns is noted, although the classification results suggest the discriminants computed from the calibration data are not positioned optimally with respect to the prediction data.

A cross-validation was performed with the averaged skin tissue absorbance spectra for the wavenumber-latent variable combination used for the PLDA prediction.

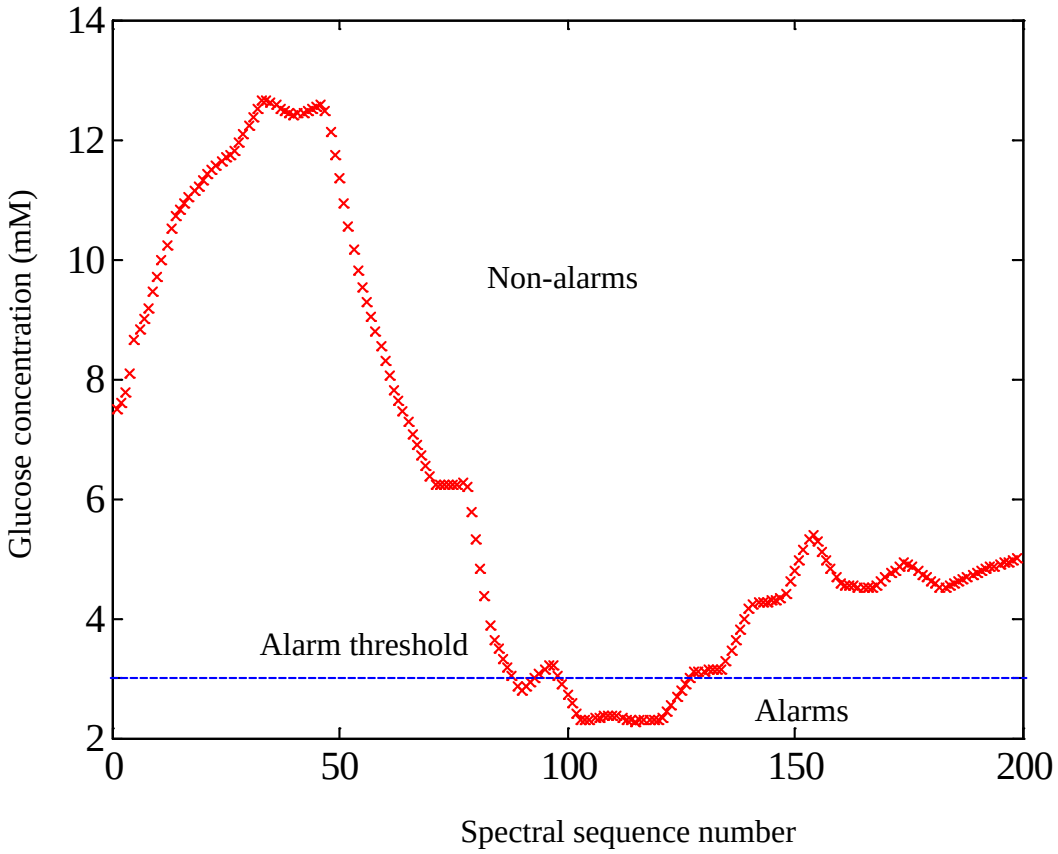


Figure 7.27. Prediction glucose concentration profile for rat 114. There are 34 alarm patterns and 165 non-alarm patterns. The alarm threshold of 3.0 mM is shown by the horizontal line.

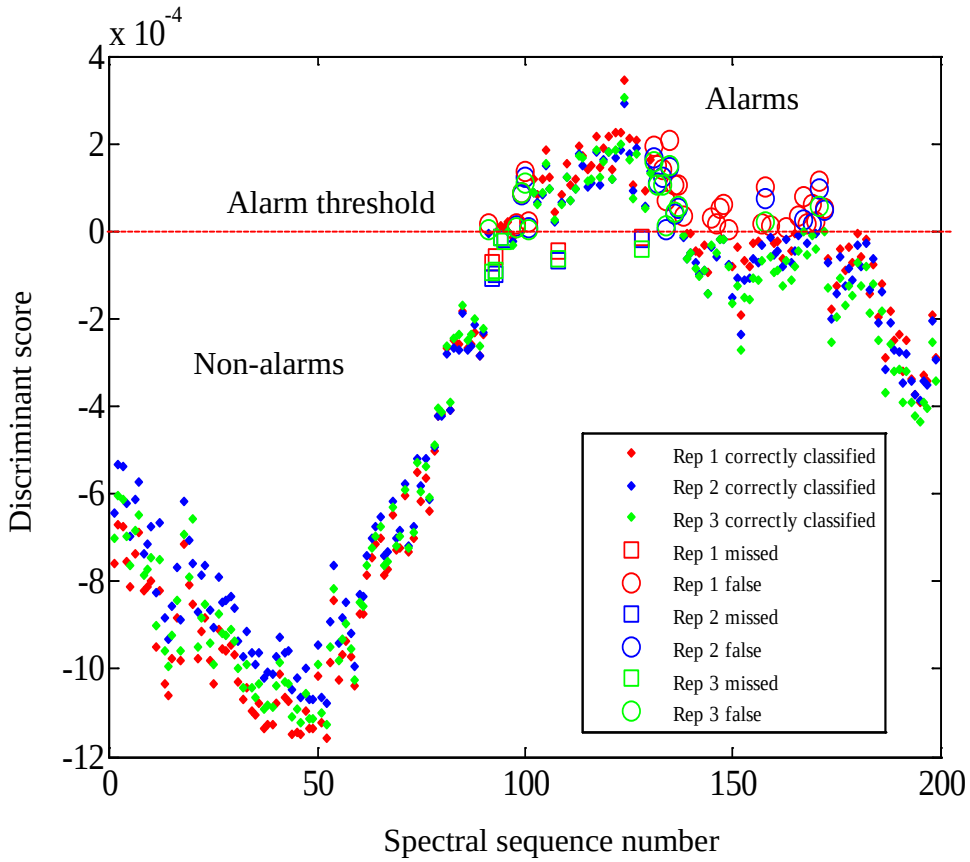


Figure 7.28. Discriminant score plot for the prediction set with three replicate classifiers. Missed and false alarms for each classifier are indicated.

Leaving out 10 % of the data for prediction in each cross-validation cycle resulted in an SECV value of 1.06 mM for 14 PLS factors and the 4850-4250 cm^{-1} wavenumber range. The concentration profiles derived from the assigned reference concentrations and cross-validated predicted concentrations are shown in Figure 7.30 A and B. These results show good correlation, but the prediction errors are higher than those obtained previously for rat 106.

The 12th, 13th, and 14th loading weights and spectral loadings computed from the differential spectra in the calibration set are shown in Figures 7.31 and 7.32, respectively. There are no significant noise features observed in the spectral loadings, but the loading weights clearly become increasingly contaminated with noise-like features as the factor number increases. The three-times higher noise level for rat 114 increases the likelihood of noise contamination in the PLS factors. Noise contamination in the loading weights is more likely than in the spectral loadings because the loading weights arise from the correlation between the reference glucose concentrations and the residual spectra (i.e., the spectral information remaining after the components modeled by the previous PLS factors have been subtracted). Chance correlations between noise features and concentrations will thus be evident in the loading weights.

Regression coefficients computed from a fit of the absorbance spectra to pure-component component spectra (Eq. 5.1) are plotted in Figure 7.33. The plotted traces show the variation in component thicknesses observed over the time course of the experiment. As previously observed with rat 106, a gradual decline in the thickness of the water layer suggests the pressure applied by the interface is causing water to move out of the optical path.

A comparison to the corresponding plot for rat 106 displayed in Figure 7.17 reveals that an overall thicker section of tissue has been sampled for rat 114. This can be seen easily by inspection of the thickness of the water layer (increase from an average of ~ 0.75 mm to an average of ~ 0.95 mm). Given that the incident source intensity was approximately the same for the two rats, the increased thickness of the tissue causes a decrease in the radiant power detected. This is most likely the reason for the increased

RMS noise observed for rat 114. The increased noise level is also most likely the cause of the decreased performance of the alarm algorithm for this rat.

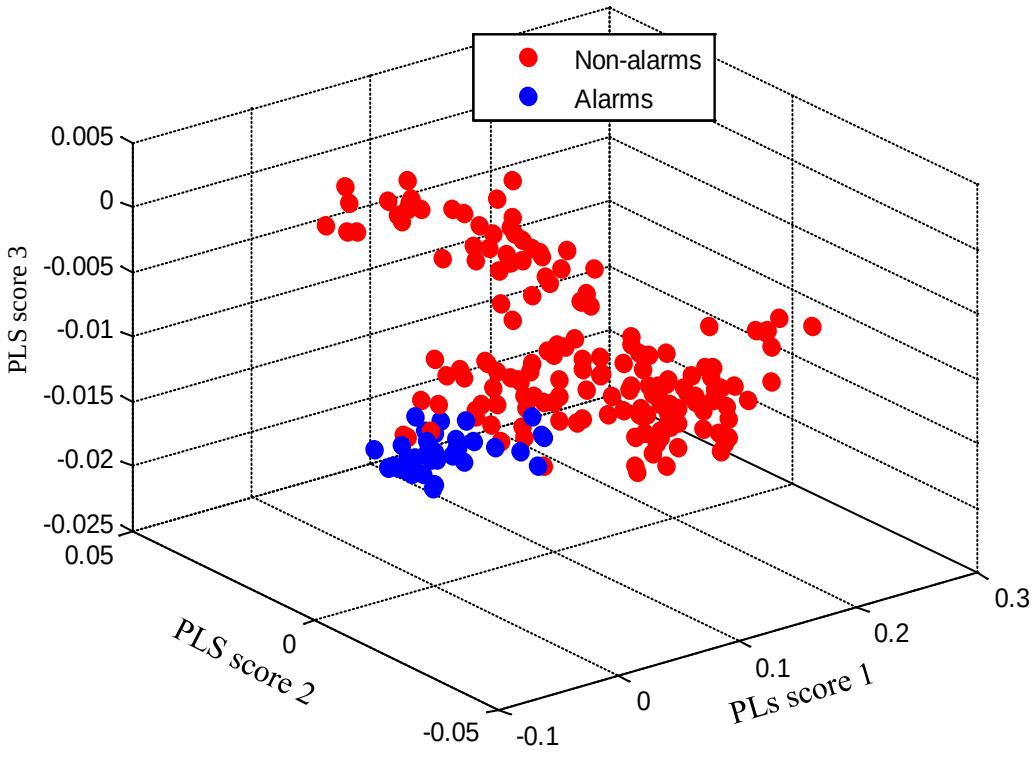


Figure 7.29. First three PLS scores for the patterns in the prediction set for rat 114. The wavenumber range was 4850-4250 cm^{-1}. Clear separation between the data classes is noted, although the class overlap is greater than that observed previously in Figure 7.15.

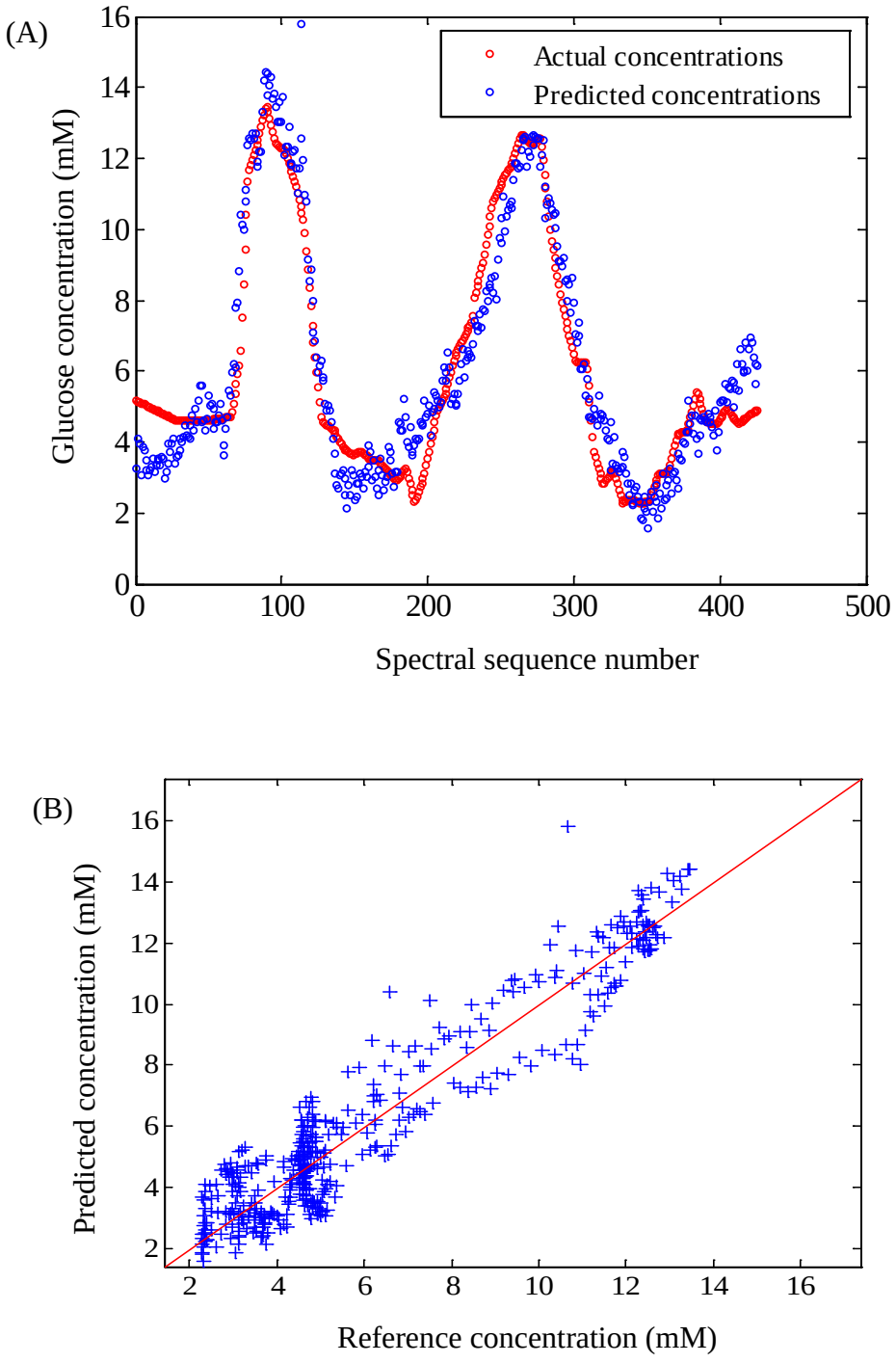


Figure 7.30. A. Time profile of assigned reference and cross-validated predicted concentrations for rat 114. Fourteen PLS factors and a wavenumber range of 4850-4250 cm^{-1} were used for the cross-validation calculations. B. Correlation plot of cross-validated predicted vs. reference concentrations.

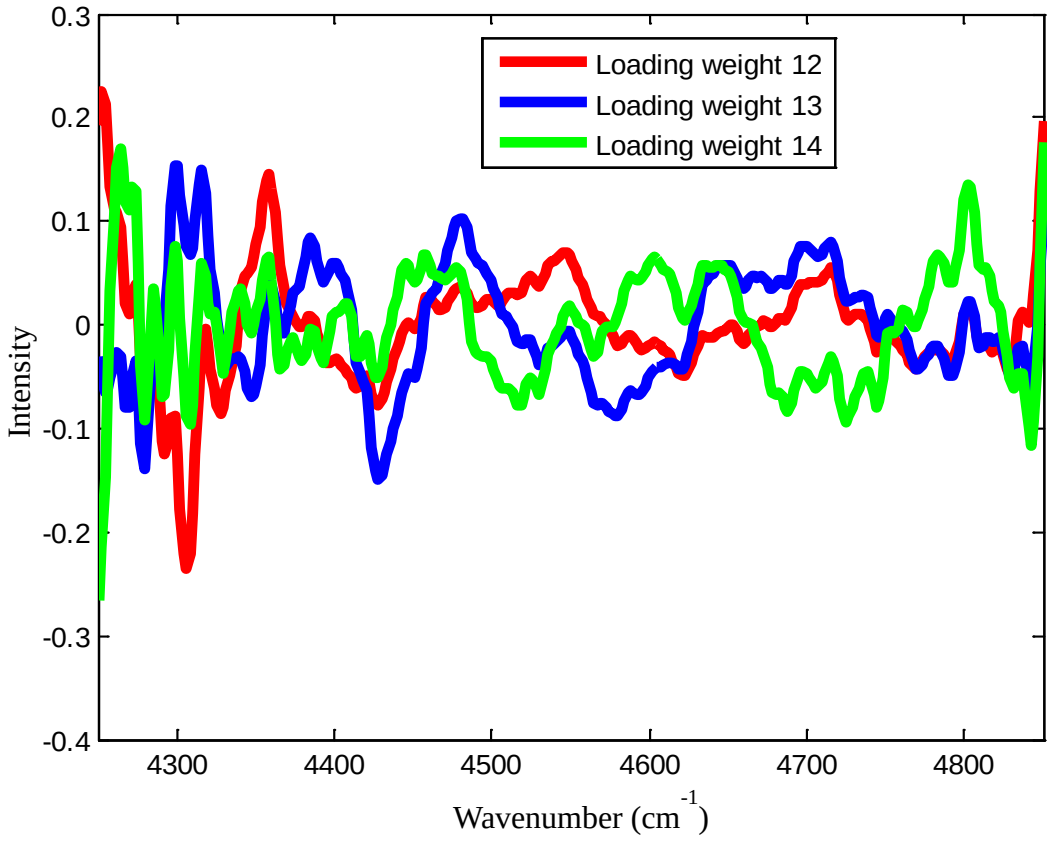


Figure 7.31. Loading weights plot for the calibration differential spectra. The 12th, 13th, and 14th loading weights show some broad spectral shapes but they are clearly becoming contaminated with noise.

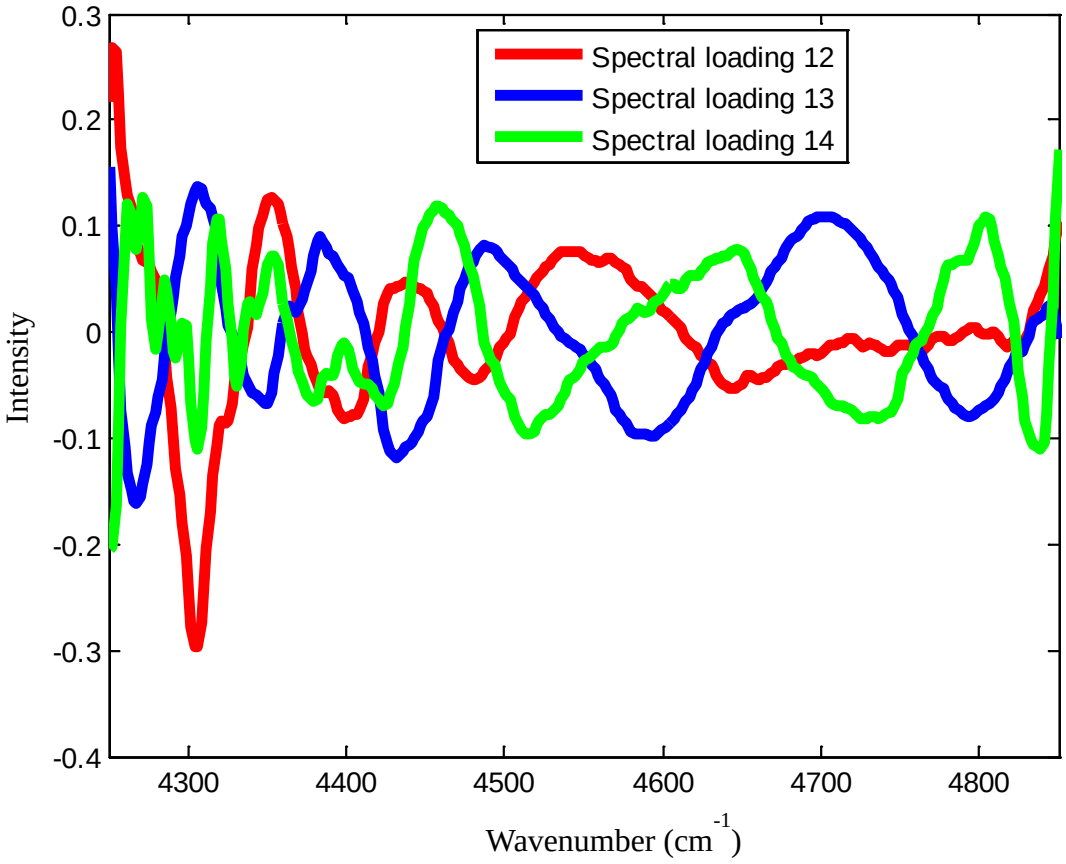


Figure 7.32. Spectral loadings plot for the calibration differential spectra. Little evidence of noise is observed in the 12th, 13th, and 14th spectral loadings.

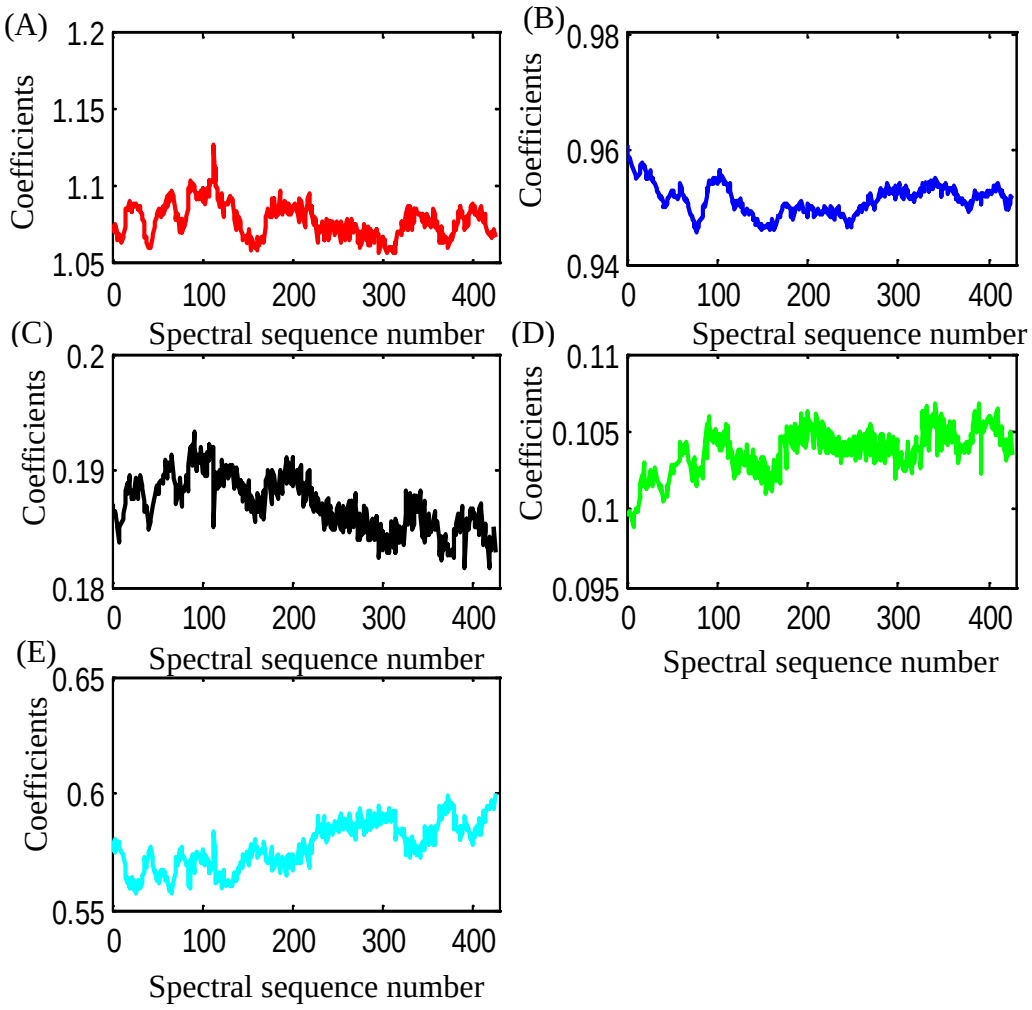


Figure 7.33. Regression coefficients computed for each of the spectra collected for rat 114. Panels A-E correspond to regression coefficients for collagen, water, keratin, fat, and an intercept term. The regression coefficients represent the thickness of each component in the skin tissue matrix.

Multiple-Day Rat Studies

Analysis of Rat 134

The glucose transient studies for rat 134 were performed over multiple days. The rat was a male rat weighing around 336 g when the survival surgery was performed on day 1. Glucose transient data were collected on five different days and the first day of the data collection was used as the calibration. The rat was awake during the data collection. Data for multiple-day rat experiments were collected as 32 co-added scans. As described previously, four consecutive single-beam spectra were averaged to increase the SNR. The RMS noise calculated with a third-order polynomial fit for the calibration data was ~ 40 μAU in the region of 4300-4500 cm^{-1} and the corresponding plot is shown in Figure 7.34. Some evidence of increasing noise is seen over the time course of the experiment.

Principal component analysis performed with the averaged absorbance spectra of the calibration day clearly showed a different spectral variance for the first 24 spectra collected. Those spectra were removed from building the calibration database. The first two principal component scores plotted together for the signal-averaged absorbance spectra in the wavenumber range of 4900-4200 cm^{-1} are shown in Figure 7.35. The *Q* vs. T^2 plot for rat 134 is shown in Figure 7.36.

A cross-validation was performed with the rat skin tissue absorbance spectra to find the best lag time. This procedure produced an optimal value of 29 minutes. Extending the calculation past 30 minutes produced increasing SECV values. A plot of SECV vs. the number of latent variables is shown in Figure 7.37. A lag time of 29 minutes seems too long on the basis of previous results and its uncertainty must be considered in evaluating further results obtained with this rat. A similar computation for

the prediction days resulted in lag time estimates of 0, 0, 34 and 0 minutes, respectively, for prediction days 1 to 4.

The glucose concentration profile for the calibration day is shown in Figure 7.38. The concentration profile was partitioned into training and monitoring sets as shown in the figure. The training and monitoring sets contained 159 and 73 signal-averaged single-beam spectra, respectively. The calculation of differential spectra for the training set produced 12,403 total spectra. Unlike rats 106 and 114, the monitoring set used rat with 134 was non-contiguous with respect to time. This was required in order for it to have both alarm and non-alarm spectra.

A grid search analysis was performed and the top four wavenumber ranges, identified on the basis of the *F*-test corrected minimum SECV values are shown in Table 7.5. Figure 7.39 shows a plot of SECV vs. the number of latent variables for the top wavenumber range that resulted from the grid search. The top wavenumber ranges and latent variables 12-16 were then used to build classification models for the monitoring set. The reference concentration, alarm threshold concentration and the critical concentration for the monitoring set were 8.95 mM, 5.60 mM and -3.35 mM, respectively. A high alarm threshold was used because of the small amount of data below 3.0 mM. When applied to the monitoring set, there were 485 alarm and 11,918 non-alarm patterns in the training set. The monitoring set contained 32 and 41 alarm and non-alarm spectra, respectively.

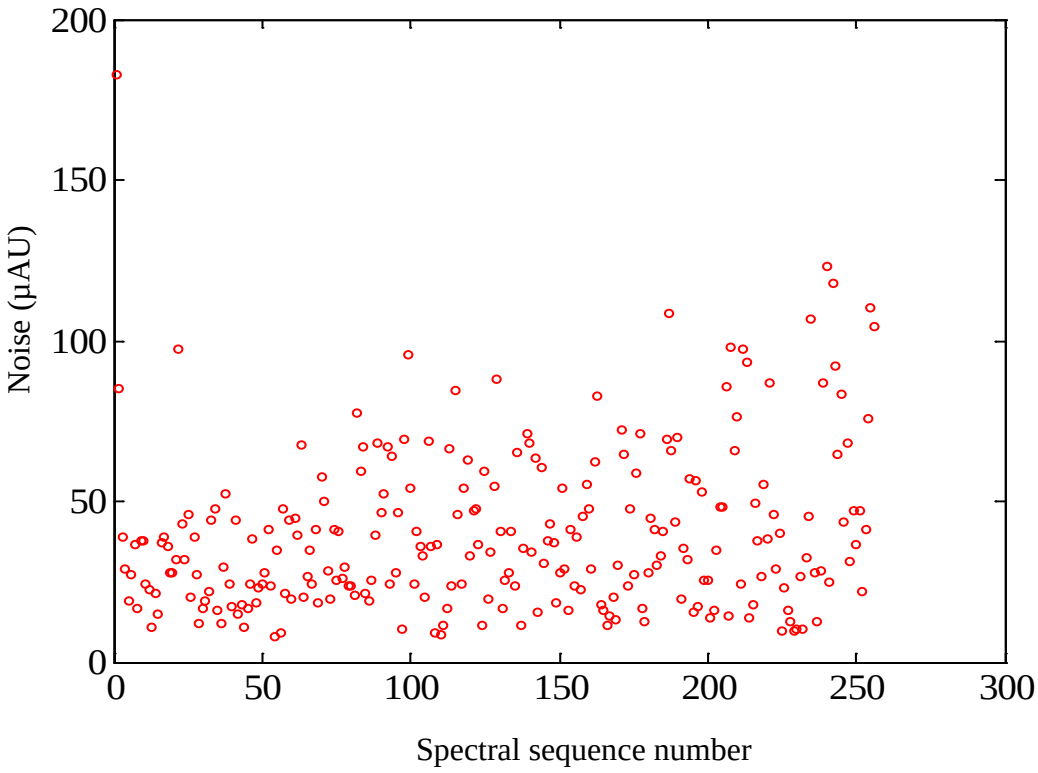


Figure 7.34. Values of RMS noise (in µAU) computed for each pair of consecutive signal-averaged spectra collected on the calibration day for rat 134. The region of 4500-4300 cm^{-1} was used. Noise values were computed by fitting the 100% line in the given wavenumber region to a third-order polynomial and then computing the noise about the fit. The average noise was ~ 40 µAU. An increase in noise can be observed over the time course of the experiment.

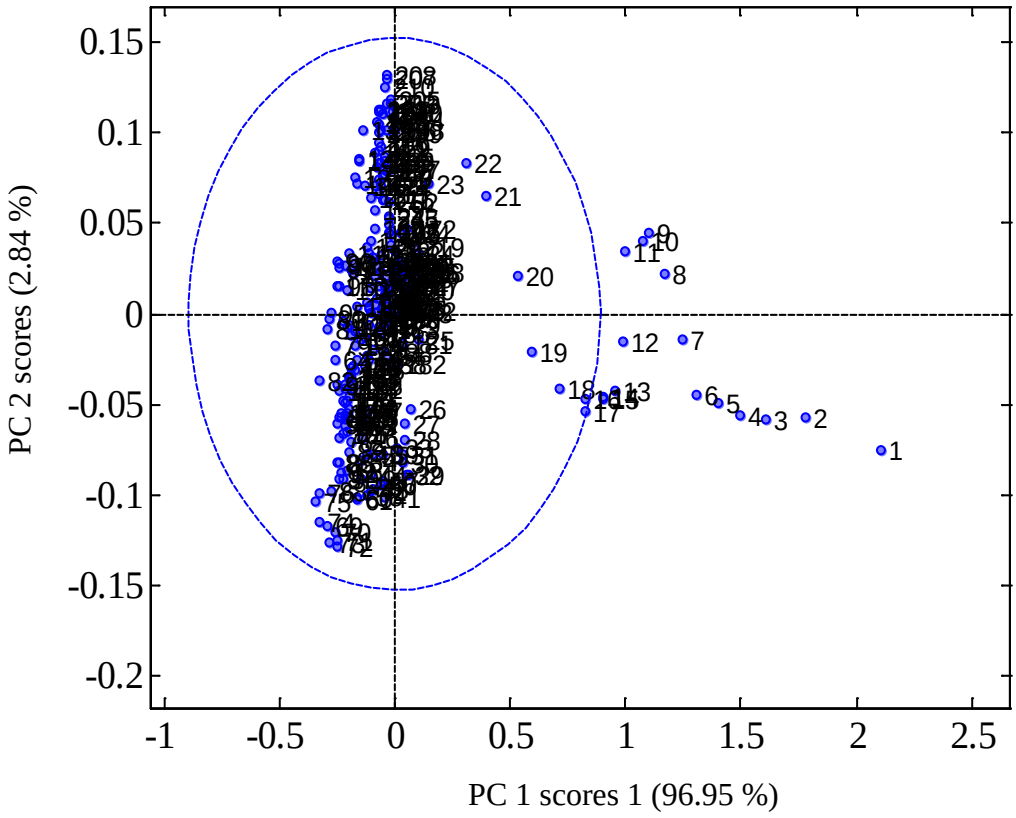


Figure 7.35. Principal component score plot for rat 134 for the calibration day. Data labels indicate the spectral sequence number. A different data clustering was observed for the first 24 averaged spectra and these spectra were removed from building the calibration database. The ellipse corresponds to the 95 % confidence interval for the scores.

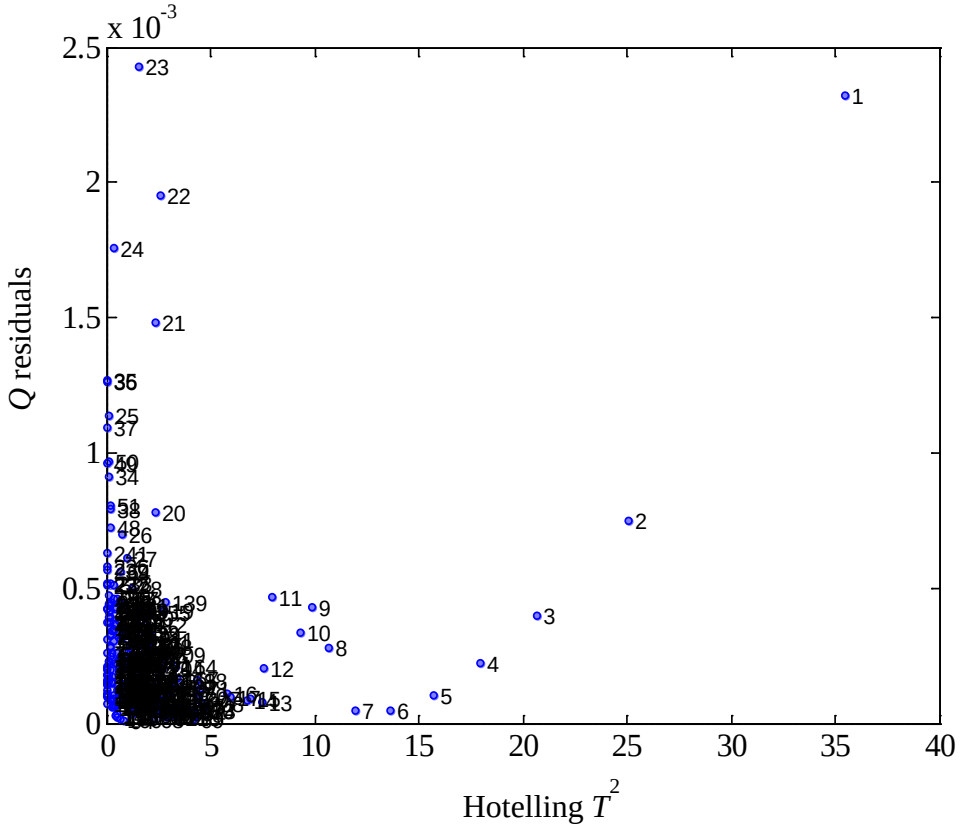


Figure 7.36. Plot of Q vs. Hotelling's T^2 for the rat absorbance spectra collected on the calibration day. The first two principal components were used for the analysis. Data labels indicate the spectral sequence number.

Table 7.6 summarizes the percentage (average ± standard deviation) of missed and false alarms when PLDA was employed with three replicate classifiers. Each classifier was based on two discriminant functions. The lowest number of PLS factors and the corresponding wavenumber range that gave the minimal missed and false alarm percentages was taken as the optimal combination for use with the prediction data. For rat 134, these optimal parameter values were a spectral range of 4850-4250 cm^{-1} and 15 PLS factors.

Comparison of Table 7.6 with Tables 7.2 and 7.4 reveals a much lower percentage of training patterns separated by the piecewise linear discriminant. The number of alarm patterns was very low and among those most of the patterns had differential concentrations very close to the critical concentration. The critical concentration was -3.4 mM and 330 patterns out of 485 total patterns had concentrations in the range of -3.4 to -4.0 mM. This likely led to the poor data separation observed with this rat.

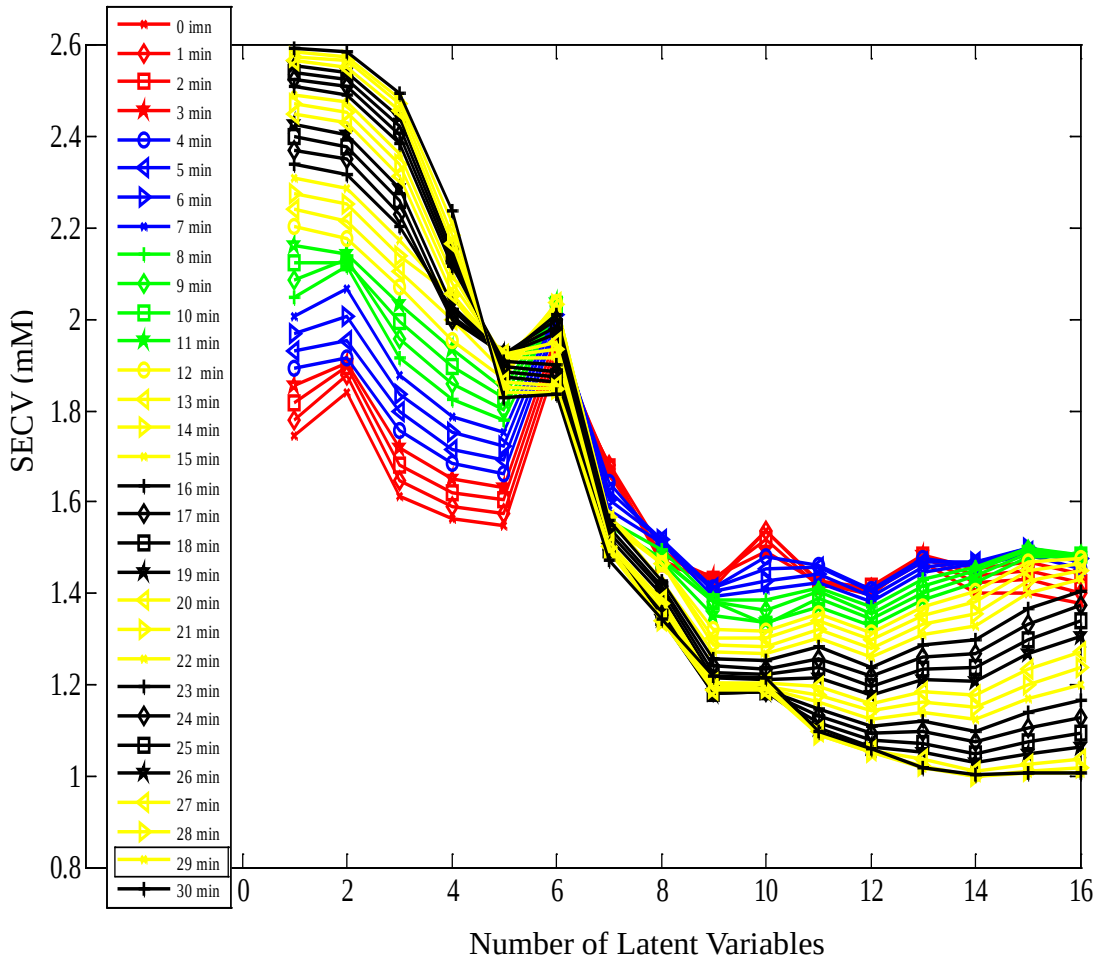


Figure 7.37. Plot of SECV vs. the number of latent variables for the calibration day of rat 134. The best lag time was 29 minutes based on the minimum SECV value of 1.00 mM.

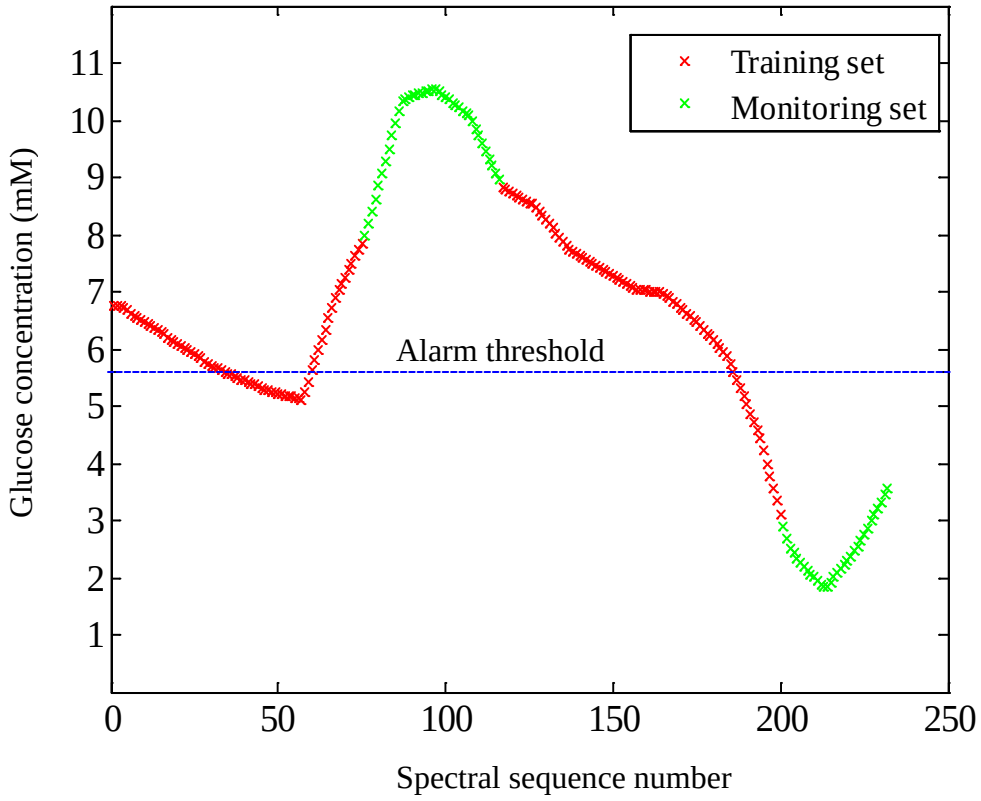


Figure 7.38. Glucose concentration profile for the calibration day. The concentration profile was partitioned into a training set which contained 159 spectra and a monitoring set which had 73 spectra. The horizontal line denotes the alarm threshold used with the monitoring set.

Table 7.5. Top four wavenumber-latent variable combinations based on minimum SECV values from the grid search analysis of training differential spectra

Wavenumber (cm^{-1})	Latent variables	SECV (mM)
4850-4200	16	0.47_2
4900-4200	16	0.47_5
4800-4200	16	0.47_7
4850-4250	16	0.47_9

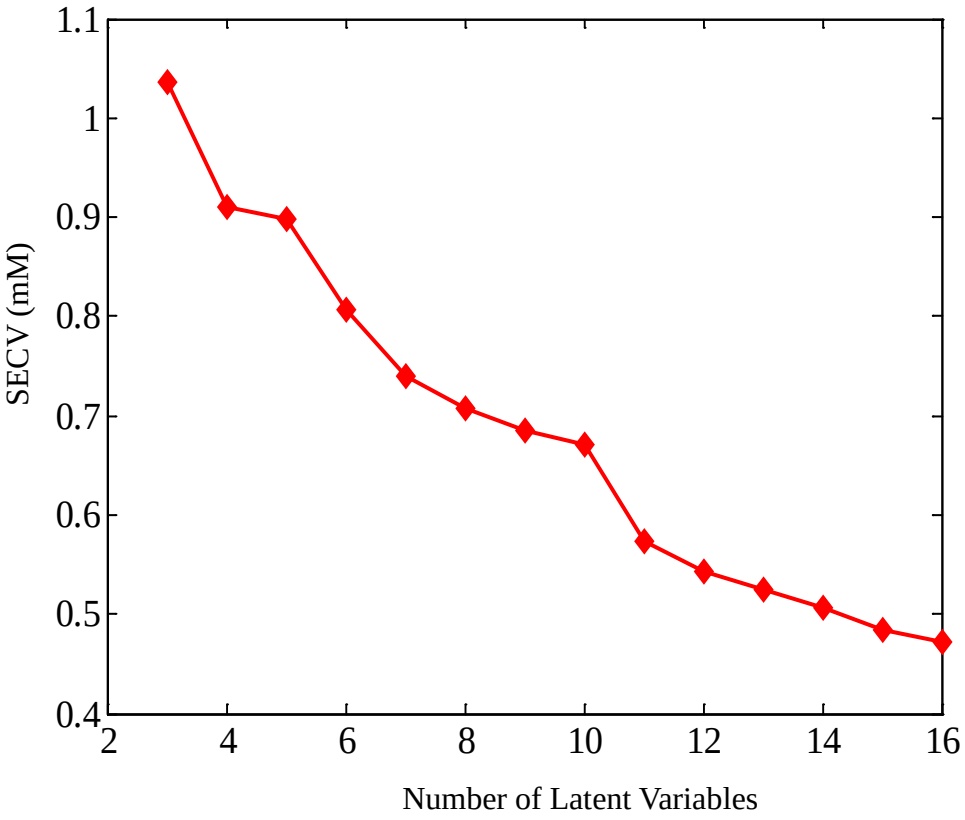


Figure 7.39. Plot of SECV vs. the number of latent variables for the top wavenumber range that resulted from the grid search analysis (4850-4200 cm^{-1}). The minimum SECV was 0.47 mM.

Table 7.6. Average percentage of missed and false alarms for the monitoring set

Spectral range,cm^{-1}		Latent variables 12	13	14	15	16
4850-4200	AM (%)[a] ± SD	3.12 ± 0.0	3.12 ± 0.0	3.12 ± 0.0	1.04 ± 1.80	0.0 ± 0.0
	AF (%)[b] ± SD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[b] ± SD	37.4 ± 3.7	42.7 ± 1.9	40.6 ± 0.2	45.9 ± 2.8	52.4 ± 3.7
4900-4200	AM (%)[a] ± SD	3.12 ± 0.0	3.12 ± 3.12	1.04 ± 1.80	4.17 ± 3.61	2.08 ±1.80
	AF (%)[b] ± SD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[b] ± SD	43.9 ± 5.6	44.4 ± 2.6	40.2 ± 0.7	44.2 ± 4.1	43.0 ± 1.0
4800-4200	AM (%)[a] ± SD	22.9 ±10.0	4.17 ± 1.80	4.17 ± 1.80	1.04 ± 1.80	1.04 ± 1.80
	AF (%)[b] ± SD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[b] ± SD	42.1 ± 1.9	43.1 ± 4.1	43.7 ± 3.6	43.6 ± 0.9	46.6 ± 0.2
4850-4250	AM (%)[a] ± SD	27.1± 1.8	6.25± 10.8	1.04 ± 1.80	0.0 ± 0.0	0.0 ± 0.0
	AF (%)[b] ± SD	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[b] ± SD	36.9 ± 0.9	44.7 ± 4.6	46.0 ± 4.2	41.9 ± 1.6	44.0 ± 0.5

[a]AM (%) ± SD: Average percentage of missed alarms ± standard deviation.

[b]AF (%) ± SD: Average percentage of false alarms ± standard deviation.

[c]DS (%) ± SD: Average percentage of separated alarm patterns with two discriminants (DS) ± standard deviation.

Each of the single-beam spectra collected on the first prediction day was assigned to a glucose concentration using the calibration lag time of 29 minutes. The alarm algorithm discussed in previous chapters used the first spectrum of the prediction set as the reference. However, as observed previously, there is a period of time at the beginning of the data collection in which the rat is adjusting to the spectrometer interface. As shown in Figures 7.40 and 7.41 for the third prediction day, these initial spectra consistently

exhibit different clustering in score plots derived from PCA and in Q vs T^2 plots. Use of the initial spectrum as a reference leads to poor classification performance.

To address this problem, a protocol is needed for determining the correct adjustment period before the reference spectrum is taken. Plots such as those in Figures 7.40 and 7.41 are retrospective (i.e., require the entire data set) and thus are not viable for use in real-time decision making. Examining the data for all the rats, it was judged that an acceptable protocol for implementing the alarm on the prediction day would be to assign 15 minutes as a static adjustment period before the assignment of the reference spectrum.

The glucose concentration profile for the first prediction day is shown in Figure 7.42. Data for the first prediction set were collected three days after the calibration day. The average RMS noise value was 37 μAU for the first prediction day. A plot of RMS noise values is shown in Figure 7.43. Noise calculations were performed as described previously. The computed values are comparable in magnitude to the calibration day. However, rather than slowly increasing with time, the values appear to stabilize after approximately 50 minutes.

The first spectrum after removing the initial 15 signal-averaged spectra (~15 minutes) had a concentration of 6.08 mM and was used as the reference spectrum. The alarm threshold was 3.00 mM and the critical concentration was -3.08 mM. All the data collected on the calibration day were put into a single block to compute the differential spectra. This yielded 26,796 total spectra and differential concentrations that were used to compute the PLS scores. The PLS scores computed with the optimal wavenumber-latent variable combination determined previously were partitioned into 9914 alarm and 16,882 non-alarm patterns.

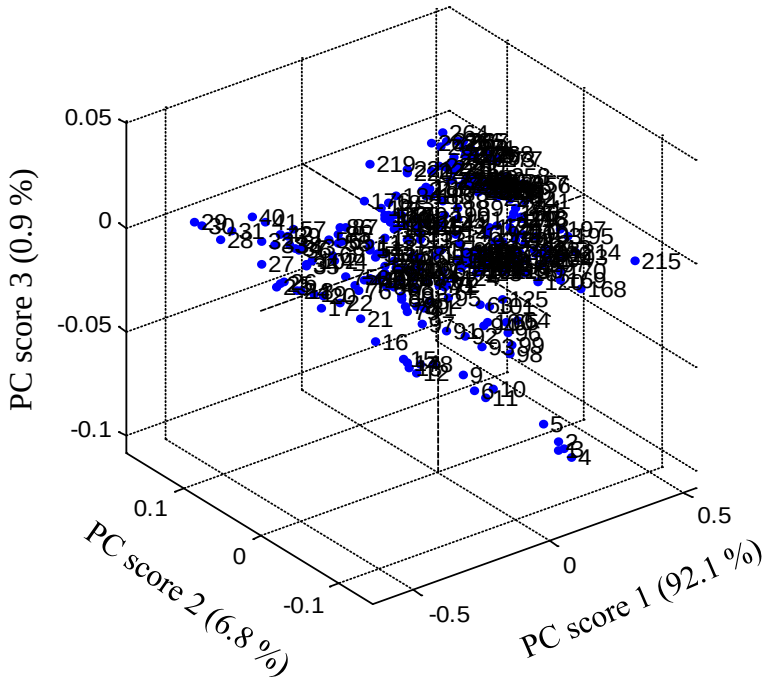


Figure 7.40. Principal component score plot for rat 134 for the third prediction day. Data labels indicate the spectral sequence number. A different clustering was observed for the spectra at the beginning of the data collection.

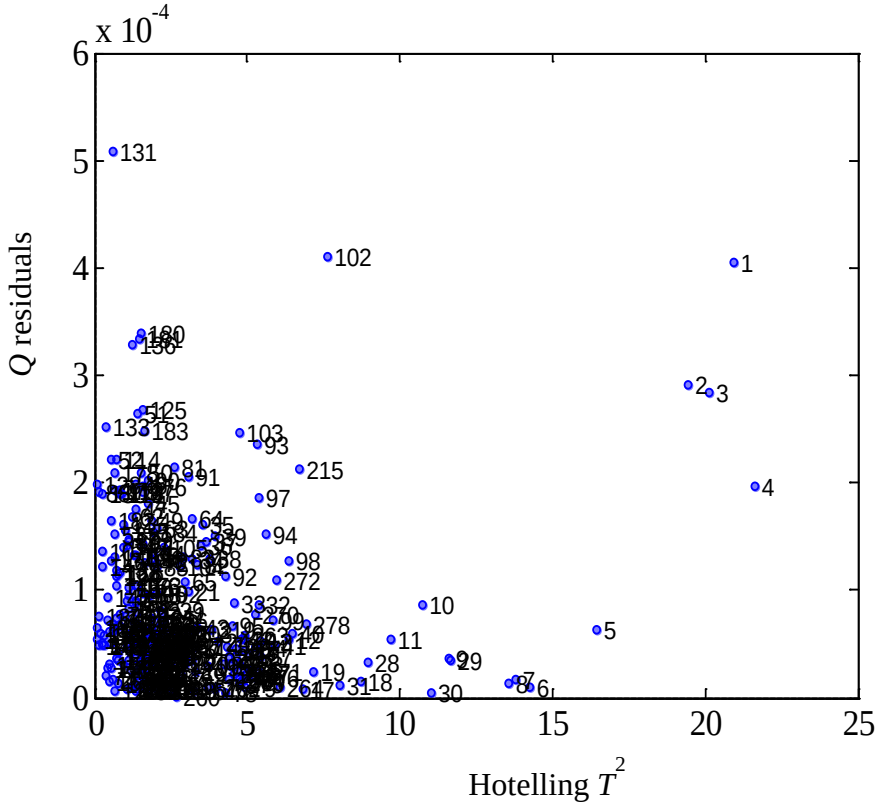


Figure 7.41. Plot of Q vs. Hotelling's T^2 for the averaged rat absorbance spectra collected on the third prediction day. The first two principal components were used for the analysis. Data labels indicate the spectral sequence number.

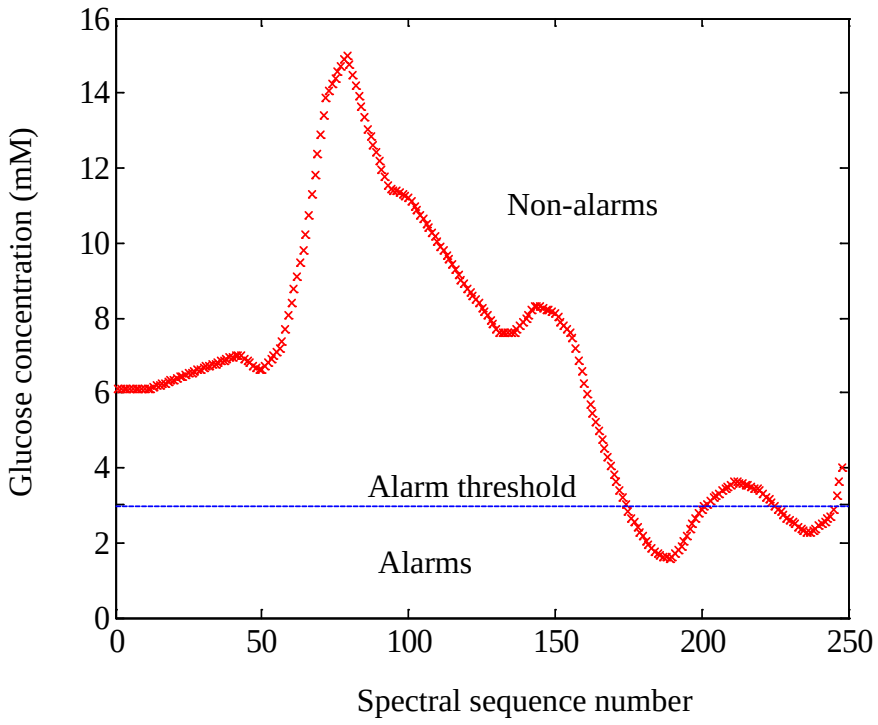


Figure 7.42. Glucose concentration profile for the first prediction day for rat 134. The concentration profile contained 248 total spectra, 48 alarm spectra and 200 non-alarm spectra. The alarm threshold of 3.0 mM is shown by the horizontal line.

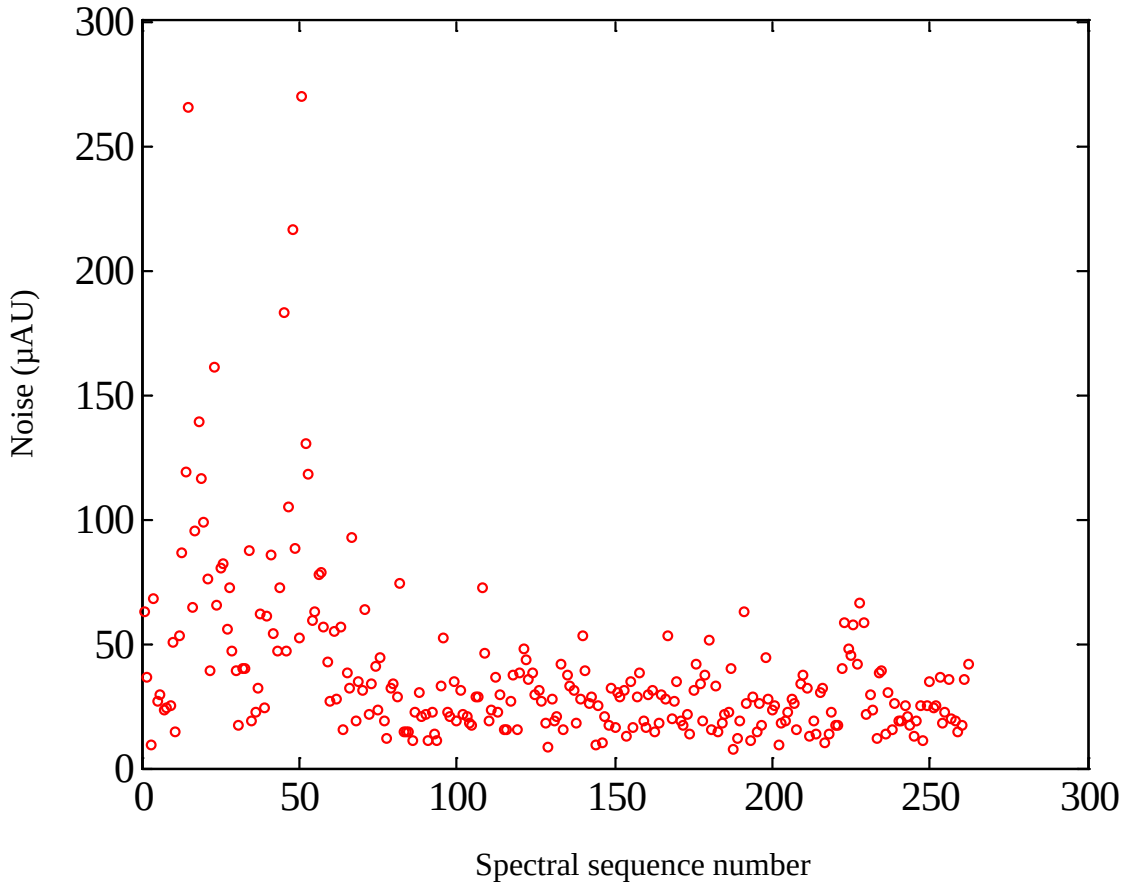


Figure 7.43. Plot of RMS noise values (in µAU) computed for each consecutive pair of signal-averaged spectra collected on the first prediction day for rat 134. The 4500-4300 cm^{-1} region was used for the noise calculation. Noise values were computed by fitting the 100% line in the given wavenumber region to a third-order polynomial and computing the noise about the fit. Noise levels are comparable to the calibration day. Some stabilization of the noise is apparent after the first hour of data collection.

Three replicate classifiers were computed with the calibration PLS patterns. Each classifier was based on two discriminant functions. Across the three replicate discriminants, an average of 34 % of the alarm patterns in the calibration set were separated. As with the monitoring data, this value is much lower than ideal and indicates difficulty in separating the alarm and non-alarm patterns. When the alarm decision rule was applied, there were 11 missed alarms out of 48 alarm patterns and there were 8 false alarms out of 200 non-alarm patterns. This corresponds to a successful detection of 77.1 % of alarm events with a false alarm rate of 4.6 %.

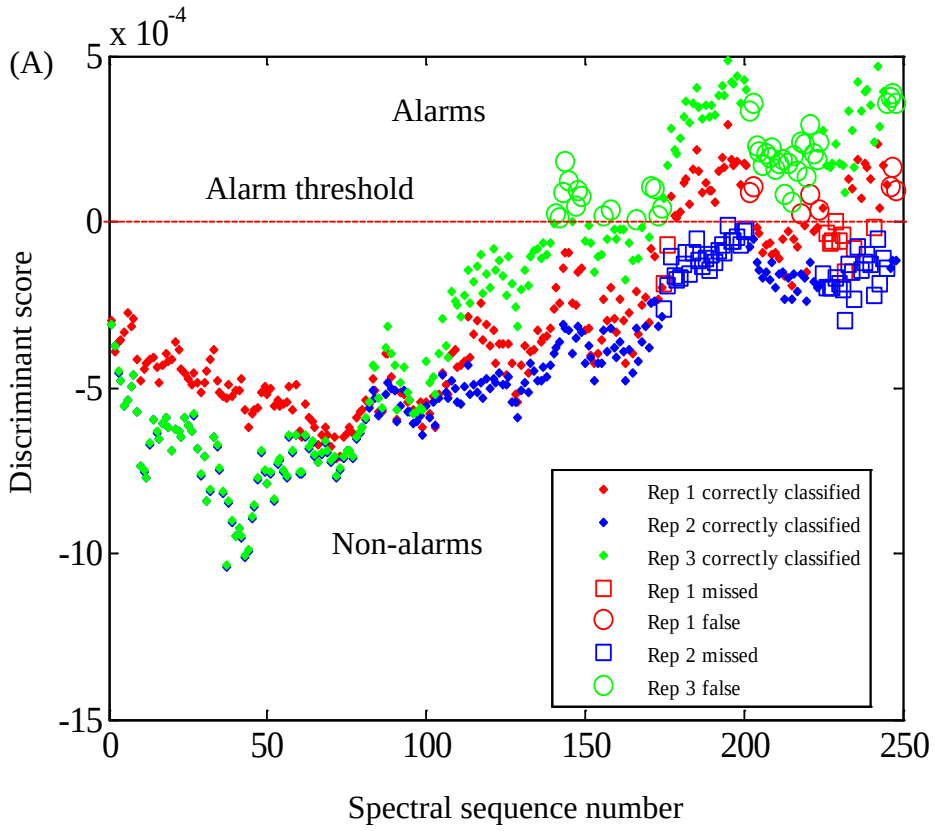
The discriminant scores plot for the three replicate classifiers is shown in Figure 7.44 A and the combined committee result is shown in Figure 7.44 B. For the committee result, the discriminant score plotted for patterns placed in the alarm class is the average score for the two or three replicate classifiers that produced positive discriminant scores. For those patterns classified as non-alarms, the average of the negative discriminant scores is plotted. Also overlaid on Figure 7.44 B with the solid red line is the glucose reference concentration assigned to each spectrum (lag time of 29 minutes). In addition, the red dashed lines display glucose concentration profiles for lag times of 0 and 15 minutes. The alternate lag times allow an assessment of the accuracy of the value of 29 minutes that was determined from the data collected on the calibration day.

Inspection of Figure 7.44 B reveals that the classifier performs well until near the end of the day. The concentration profile defines a very challenging region at the end in which the assigned reference value goes in and out of alarm as judged by the 3.0 mM threshold. This is also a region where imprecision in the lag can be affecting the assignment of which patterns should be designated as alarms and non-alarms.

The regression coefficients for each of the spectra collected for the first prediction day of rat 134 are shown in Figure 7.45. This figure shows the thickness variations of each pure component during the data collection. As described previously, these variations indicate changes in the background matrix during the data collection.

The first four PLS scores for the first prediction day are plotted with respect to the spectral sequence number in Figure 7.46. For a successful data prediction, the prediction PLS scores should be within the ranges defined by the calibration PLS scores. For the prediction data, PLS score 2 is within the range defined by the calibration data. In the other plots, however, it can be observed that some of the prediction PLS scores occupy a region outside the range of the calibration scores.

As noted previously, most of the missed alarm patterns were observed for spectral sequence numbers 225-245 acquired near the end of the day. The PLS scores for these patterns were outside the calibration PLS scores as can be seen in Figures 7.46 A, C, and D. In Figure 7.47, the differential absorbance value at 4400 cm^{-1} for the prediction day is plotted vs. the spectral sequence number to check for any spectral variance that occurred during the data collection at the peak wavenumber of the most important glucose absorbance band. It can be clearly seen that the absorbance value shows a sequential drift with time. A new background component may have arisen during the data collection. This suggests that the reference spectrum taken at the beginning of the day may no longer be valid by the end of the data collection.


(A)
x 10^-4
5
0
-5
-10
-15
Discriminant score
Alarms
Alarm threshold
Non-alarms
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Rep 1 missed
Rep 1 false
Rep 2 missed
Rep 3 false
0
50
100
150
200
250
Spectral sequence number

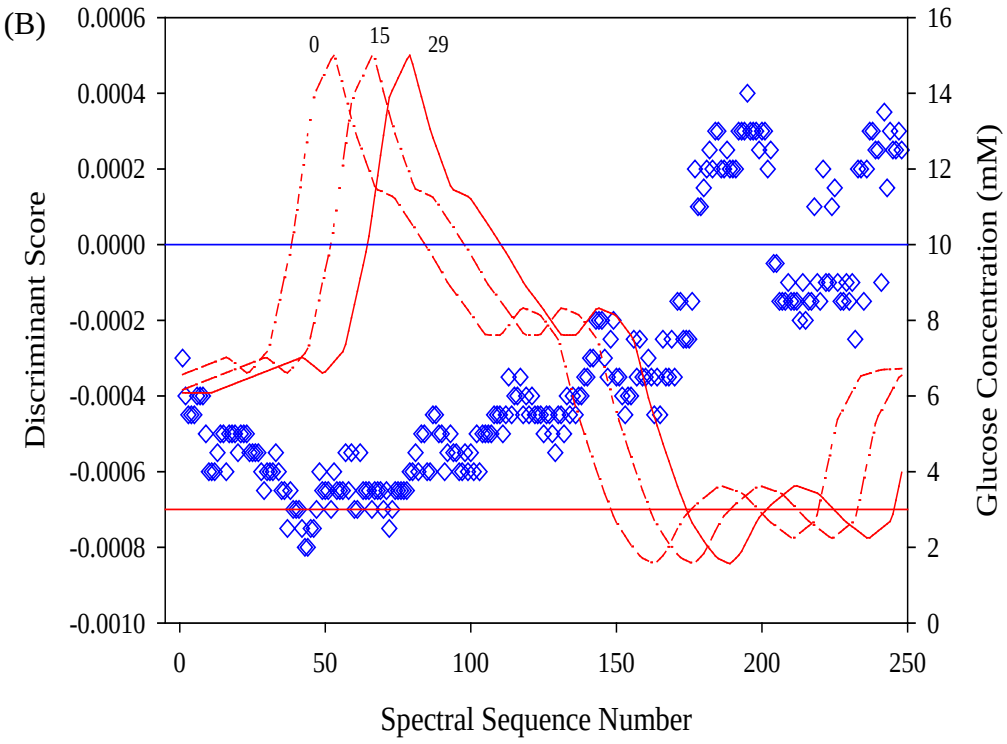


Figure 7.44. A. Discriminant score plot for the first prediction day of rat 134 with three replicate classifiers. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). The glucose transients related to three lag times, 0, 15 and 29, minutes are shown in the figure.

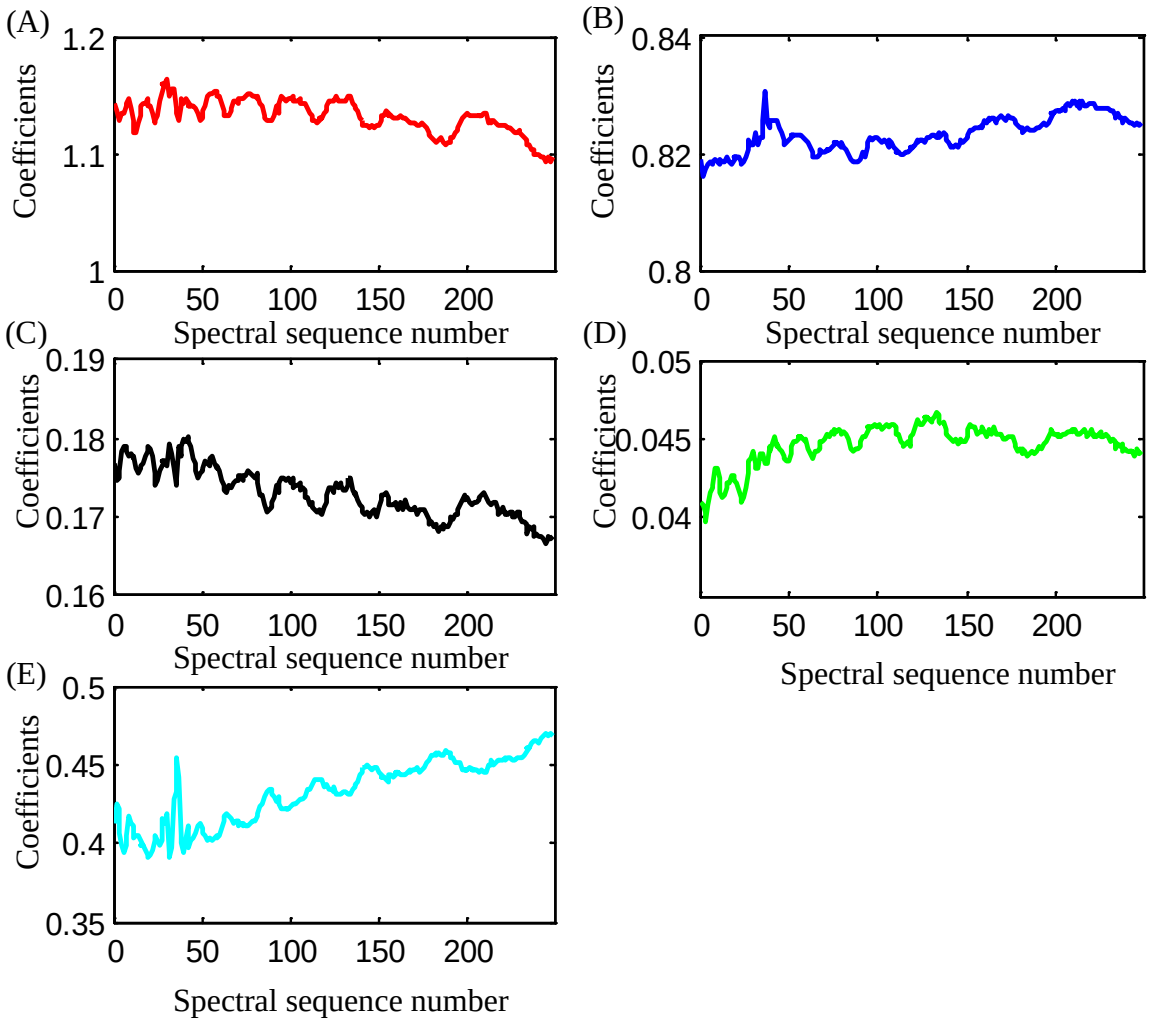


Figure 7.45. Regression coefficients computed for each of the spectra collected for the first prediction day of rat 134. Panels A-E correspond to regression coefficients for collagen, water, keratin, fat, and an intercept term. The regression coefficients represent the thickness of each component in the skin tissue matrix.

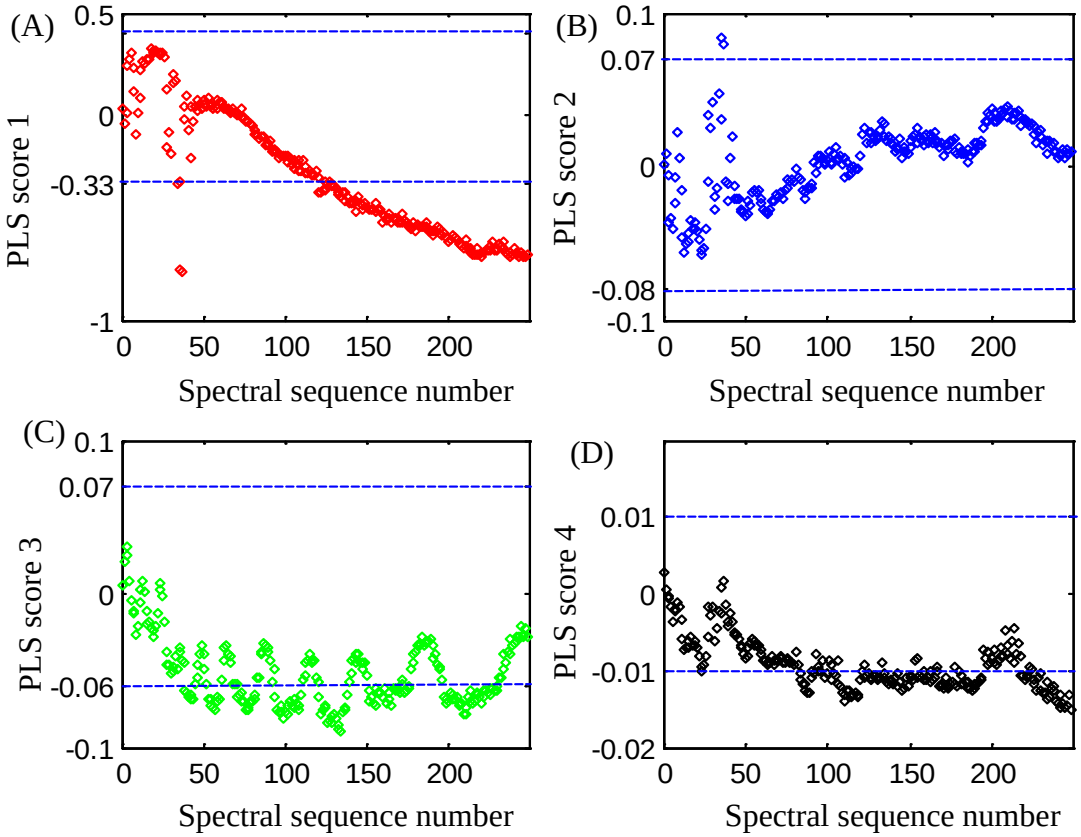


Figure 7.46. Plot of PLS scores vs. spectral sequence number for the differential spectra collected on the first prediction day. The 4850-4250 cm^{-1} wavenumber range was used. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as dashed lines. Panels A, B, C, and D correspond to the scores for latent variables 1-4, respectively.

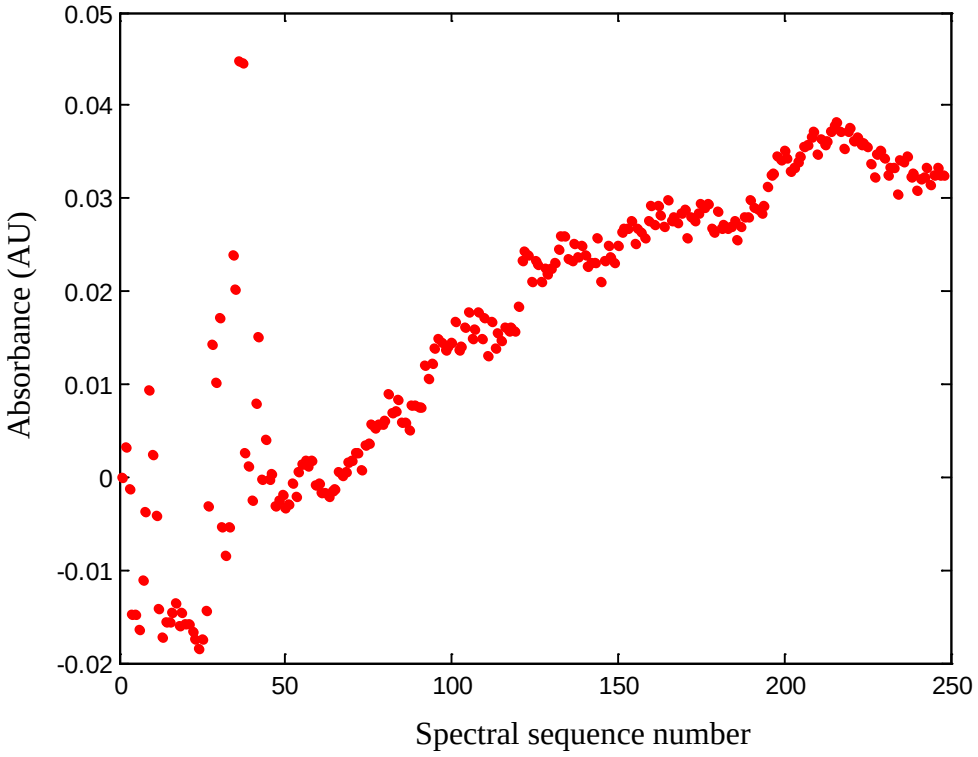


Figure 7.47. Absorbance at 4400 cm^{-1} plotted vs. the spectral sequence number for the differential spectra for the first prediction day.

Each of the single-beam spectra collected on the second prediction day was assigned to a glucose concentration using the calibration lag time of 29 minutes. The data for the second prediction set were collected eight days after the calibration day. As discussed previously, the data collected in the first 15 minutes of the prediction day (15 signal-averaged single-beam spectra) were removed from the data prediction assuming it takes that long to stabilize the rat within the interface. The glucose concentration profile for the second prediction day is shown in Figure 7.48. All of these spectra are above the alarm threshold of 3.0 mM (i.e., all are non-alarms). The signal-averaged single-beam spectra were used for the RMS noise value calculations and with 128 scans and a third-order polynomial fit the average RMS noise value was 30 μAU. A plot of RMS noise values is shown in Figure 7.49.

The first spectrum after removing the first 15 signal-averaged spectra had a concentration of 6.64 mM and was used as the reference spectrum. The alarm threshold was 3.00 mM and the critical concentration was -3.64 mM. The PLS scores computed with the optimal wavenumber-latent variable combination were partitioned into 7467 alarm and 19,329 non-alarm patterns. Three replicate classifiers were computed with the calibration PLS patterns and the first two discriminants from each of the replicates were used for the prediction. Across the three replicate discriminants, an average of 38 % of the alarm patterns in the calibration set were separated. This again represents a very low degree of separation of the alarm and non-alarm patterns.

When the alarm decision rule was applied, there were 13 false alarms out of the 226 non-alarm patterns. This corresponded to an occurrence of 5.8 %. The discriminant scores plot for the replicate classifiers is shown in Figure 7.50 A, and the committee

result is plotted as before in Figure 7.50 B. While staying below the alarm threshold, the committee result begins to approach the decision boundary as the experiment proceeds. This is an indication that the background is changing and that the discriminant boundary is no longer positioned correctly.

The regression coefficients for each of the spectra collected for the second prediction day of rat 134 are shown in Figure 7.51. This figure shows the thickness variations of each of the individual components during the data collection. Compared to the first prediction day, the water thickness is approximately 20 % lower. This may be responsible for the lower RMS noise values observed for the second prediction day, as less water translates to more transmitted light. In principle, the calculation of the differential spectra should remove the effects of variation in pathlength from day to day.

The first four prediction PLS scores plotted for the second prediction day are shown in Figure 7.52. The prediction PLS score 4 is within the space defined by the calibration PLS score 4. In the other PLS scores plots, it can be observed that some prediction scores occupy a space outside that defined by the calibration scores. Most of the false alarm patterns were observed for the spectral sequence numbers of 134-150. The PLS scores for these patterns were outside the calibration scores as can be seen in Figures 7.52 A, B, and C. The differential absorbance value at 4400 cm^{-1} was plotted as shown in Figure 7.53. The same sequential drift observed earlier can also be seen for the second prediction day. This suggests the presence of a new background component that is not being accounted for in the calculation of the differential spectra. This is consistent with the profile of the discriminant scores discussed previously.

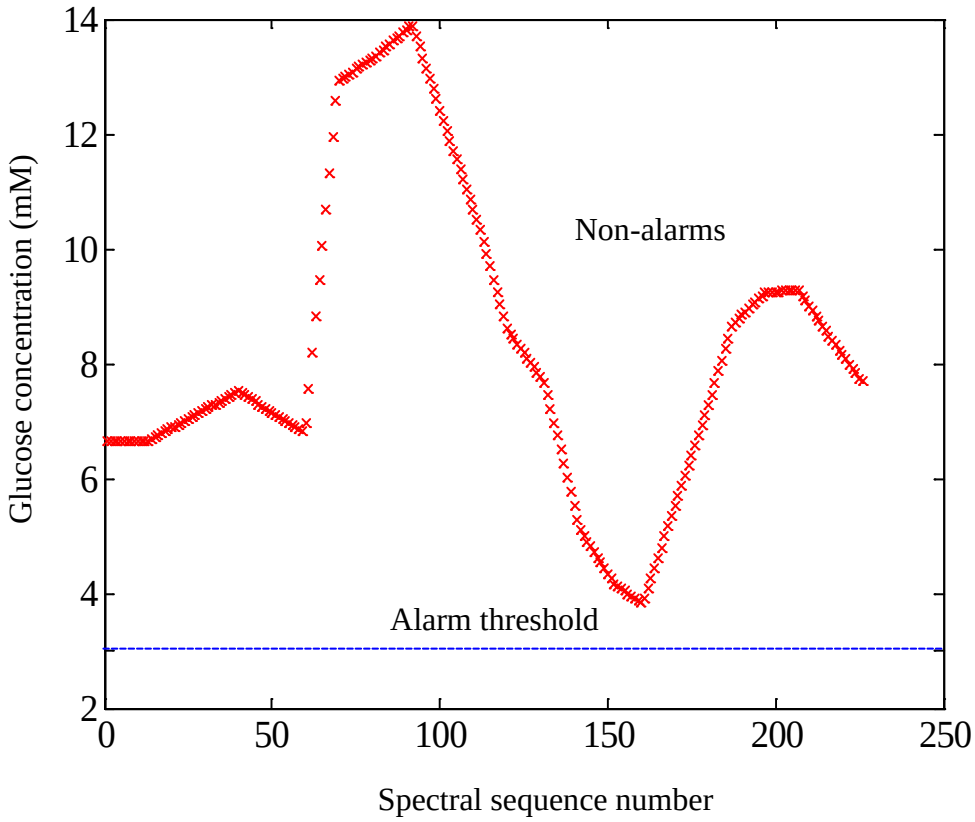


Figure 7.48. Glucose concentration profile for the second prediction day. The concentration profile contained 226 total spectra and all had concentrations above the 3.00 mM alarm threshold shown by the horizontal line.

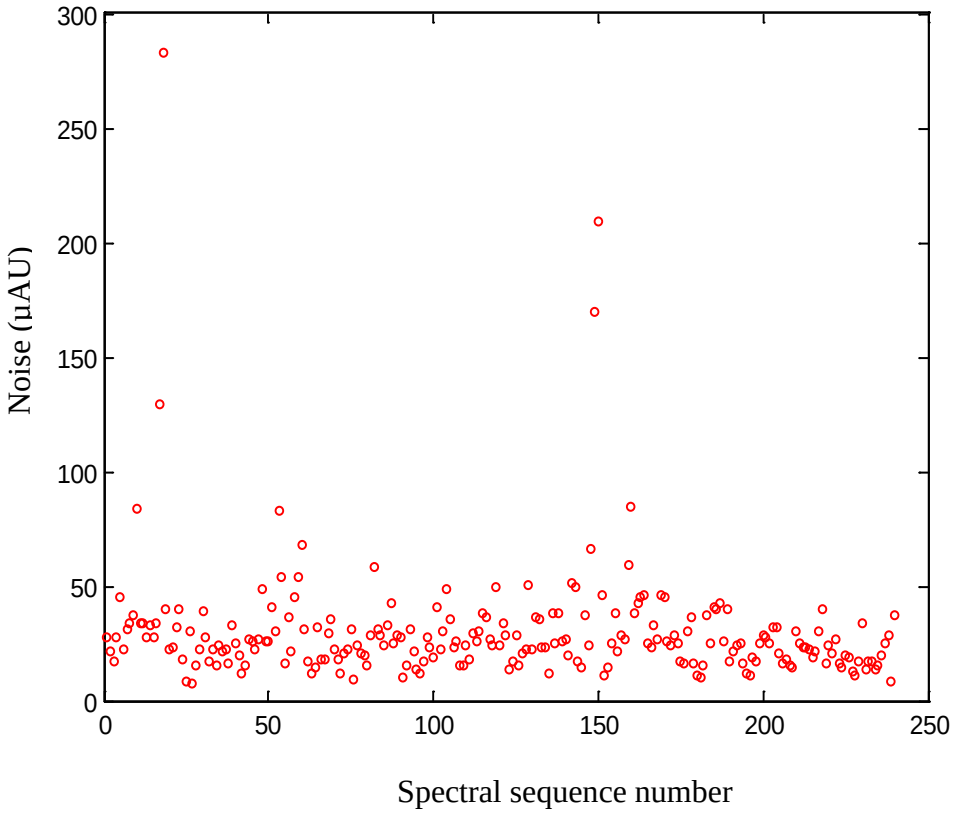
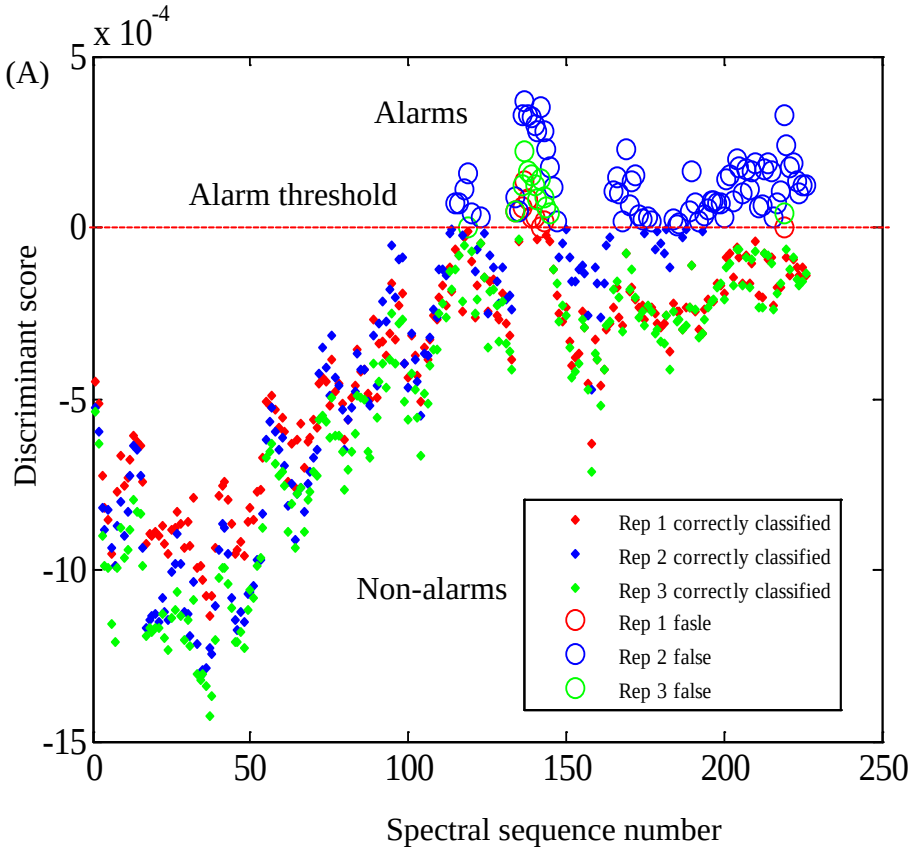


Figure 7.49. For the second prediction day for rat 134, values of RMS noise (in µAU) computed for each consecutive pair of signal-averaged spectra. The 4500-4300 cm^{-1} region was used for the noise calculation. Noise values were computed by fitting the 100% line in the given wavenumber region to a third-order polynomial and computing the noise about the fit. Except for a few extreme cases, the noise values are consistent across the time course of the experiment.


(A)
x 10-4
5
0
-5
-10
-15
Discriminant score
Alarms
Alarm threshold
Non-alarms
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Rep 1 fasle
Rep 2 false
Rep 3 false
0
50
100
150
200
250
Spectral sequence number

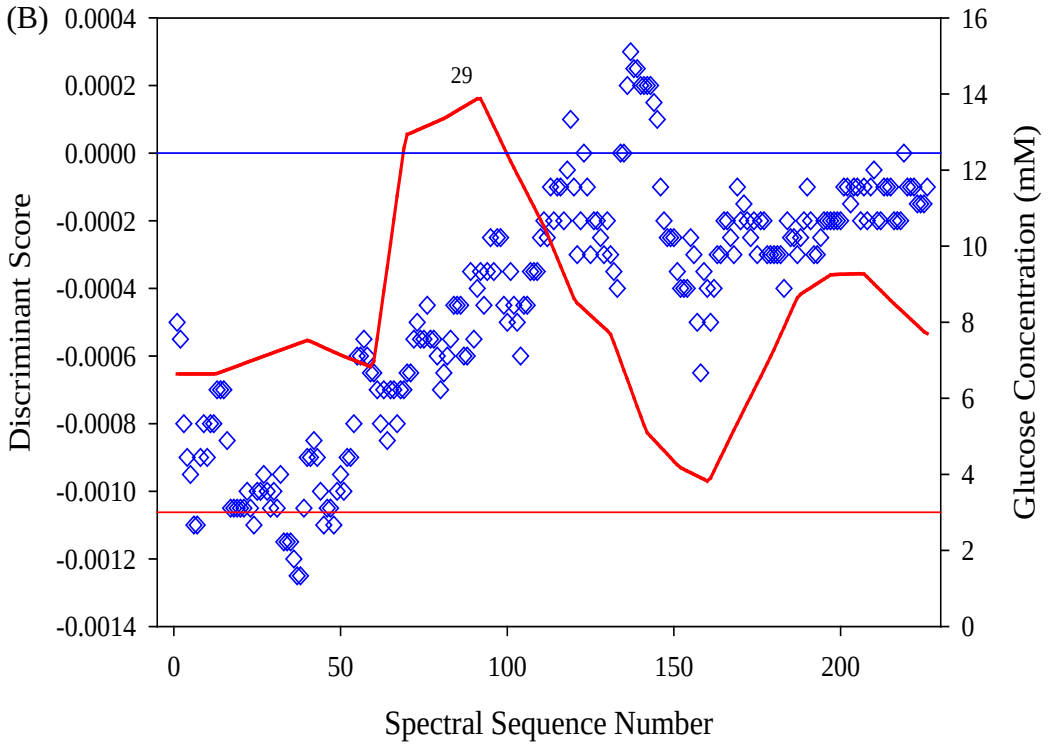


Figure 7.50. A. Discriminant score plot for the second prediction day of rat 134 with three replicate classifiers. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). The glucose concentration profile corresponding to a lag time of 29 minutes is shown in the figure.

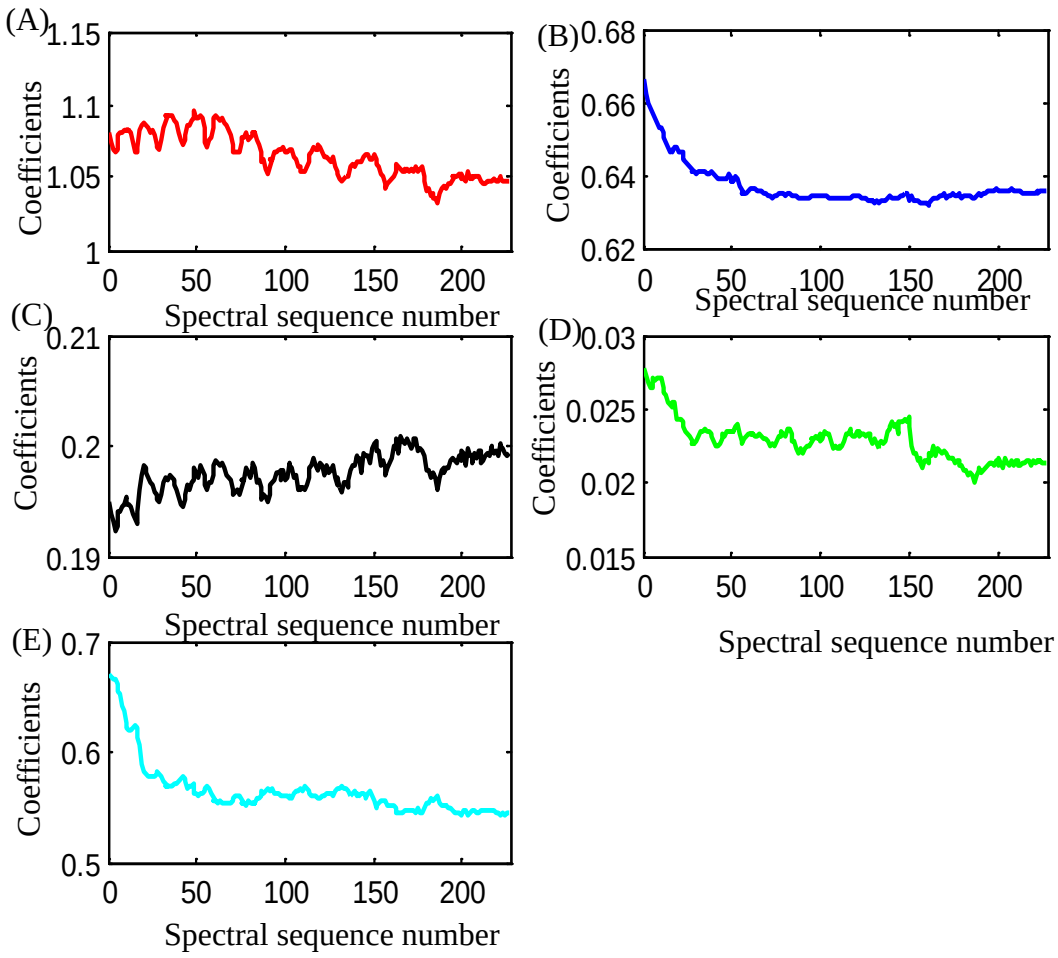


Figure 7.51 Regression coefficients computed for each of the spectra collected for the second prediction day of rat 134. Panels A-E correspond to regression coefficients for collagen, water, keratin, fat, and an intercept term. The regression coefficients represent the thickness of each component in the skin tissue matrix.

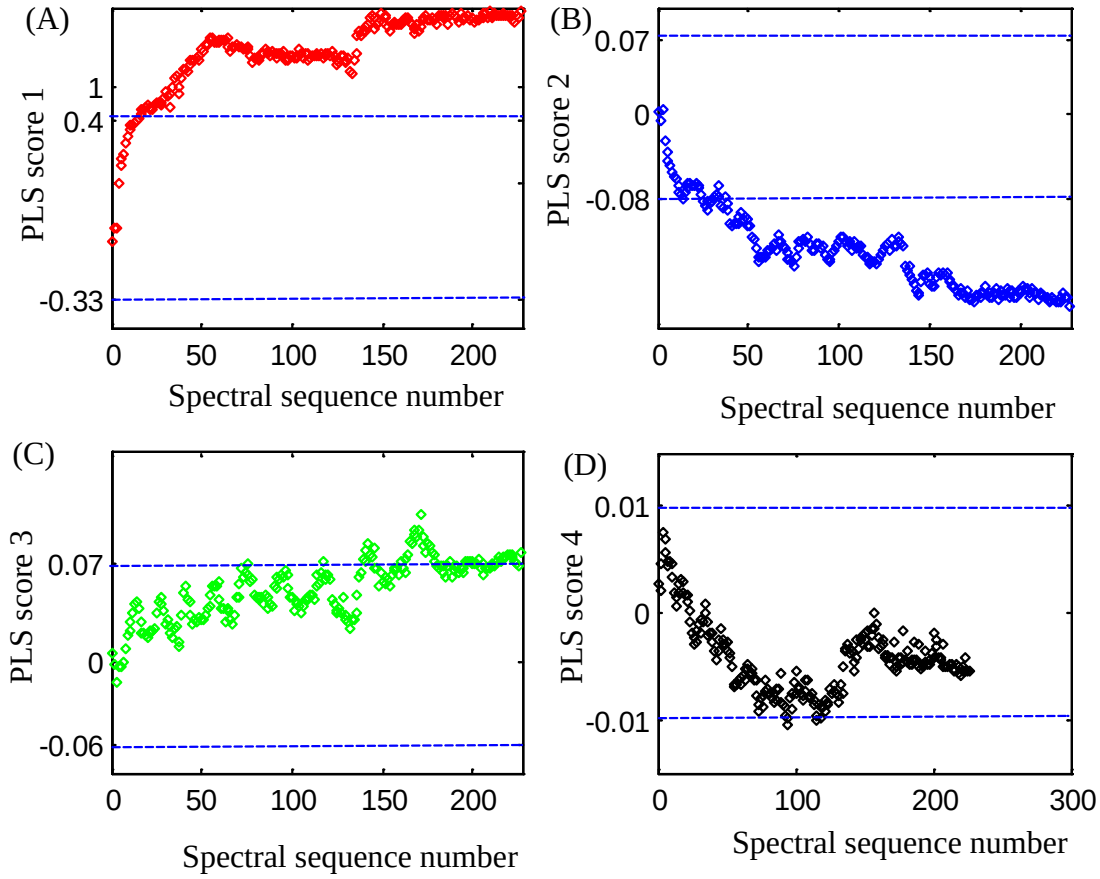


Figure 7.52. Plot of PLS scores vs. spectral sequence number for the prediction differential spectra collected on the second day. The 4850-4250 cm^{-1} wavenumber range was used. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as dashed lines. Panels A, B, C, and D correspond to the scores for latent variables 1-4, respectively.

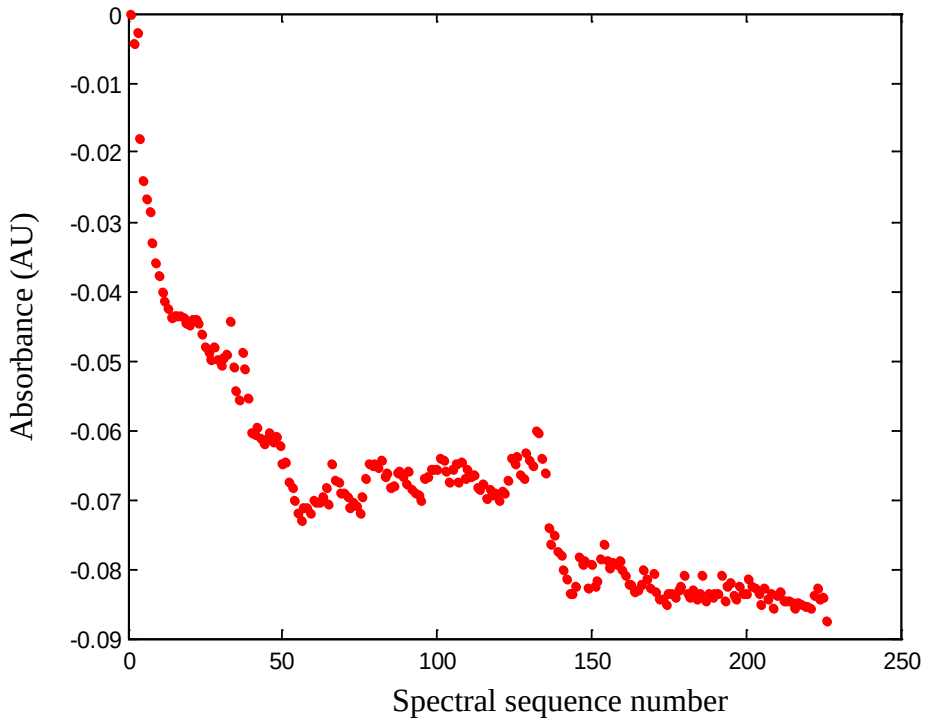


Figure 7.53. Absorbance at 4400 cm^{-1} plotted vs. the spectral sequence number for the differential spectra collected on the second prediction day.

The calibration lag time of 29 minutes was used to assign a glucose concentration to each of the spectra collected on the third prediction day. The data for the third prediction set was collected 17 days after the calibration day. As before, spectra collected during the first 15 minutes were removed from the data prediction. The glucose concentration profile for the third prediction day is shown in Figure 7.54. As with the second day, there were no alarm spectra. The signal-averaged single-beam spectra were used for RMS noise calculations and with 128 scans and a third-order polynomial fit, the average RMS noise value was 38 μAU. A plot of RMS noise values is provided in Figure 7.55.

The first spectrum after removing the first 15 signal-averaged spectra had a concentration of 6.81 mM and was used as the reference spectrum. The alarm threshold was 3.00 mM and the critical concentration was -3.81 mM. The PLS scores computed with the optimal wavenumber-latent variable combination were partitioned into 6842 alarm and 19,954 non-alarm patterns. Three replicate classifiers were computed with the calibration PLS patterns and the first two discriminant functions from each of the replicates were used for the prediction. On average, the three replicate classifiers separated 34 % of the alarm patterns in the calibration set. As observed previously, this is a low degree of separation.

When the alarm decision rule was applied, there were no false alarms out of 263 non-alarm patterns. The discriminant scores plot is shown in Figure 7.56 A and the committee result is displayed in Figure 7.56 B. The value of the committee classifier is clearly observed, as only two of the three replicate classifiers are performing correctly.

The regression coefficients obtained from fitting the pure components to the spectra for day 3 are plotted in Figure 7.57. Variation in the water thickness is again noted relative to the previous days of data collection (Figures 7.45 and 7.51).

Figure 7.58 compares the PLS scores for latent variables 1-4 computed from the prediction data to the corresponding calibration scores. Scores 1-3 match reasonably well to the calibration data, although the scores for the fourth factor show clear deviation. Figure 7.59 displays the absorbance at 4400 cm^{-1} over the course of the experiment. As seen previously, there is indication of drift in the response with time. The overall conclusion for day 3 is that while there are some clear issues of drift in the response and some degree of mismatch to the calibration data, the committee classifier is able to overcome these limitations well enough to assign the spectra correctly to the non-alarm class.

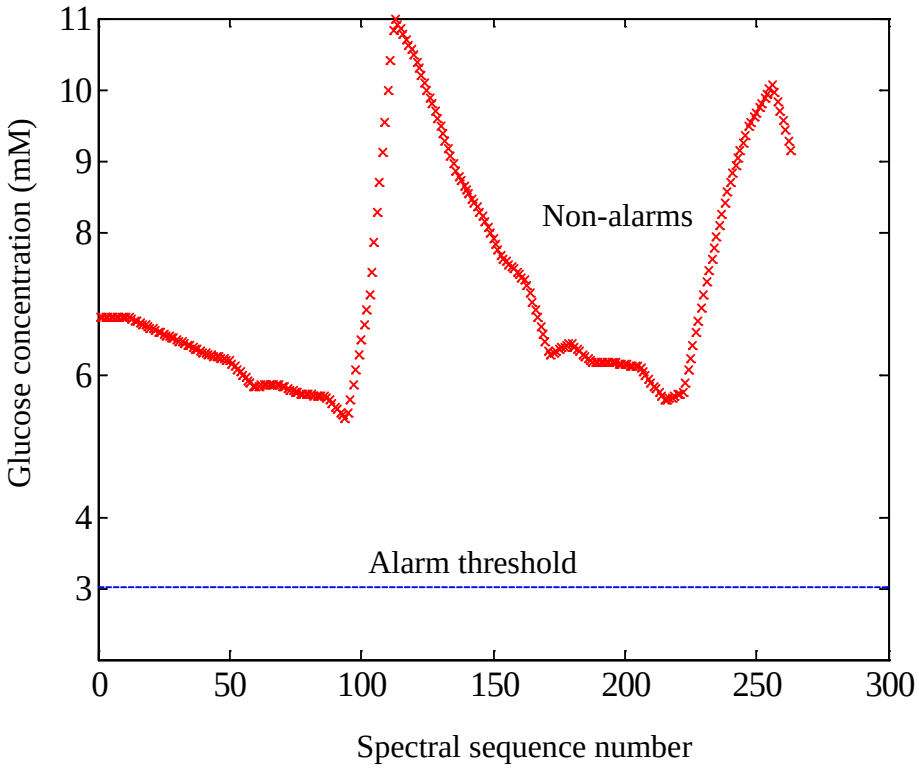


Figure 7.54. Glucose concentration profile for the third prediction day of rat 134. The concentration profile contained 263 total spectra and all had concentration above 3.00 mM, the alarm threshold (horizontal line).

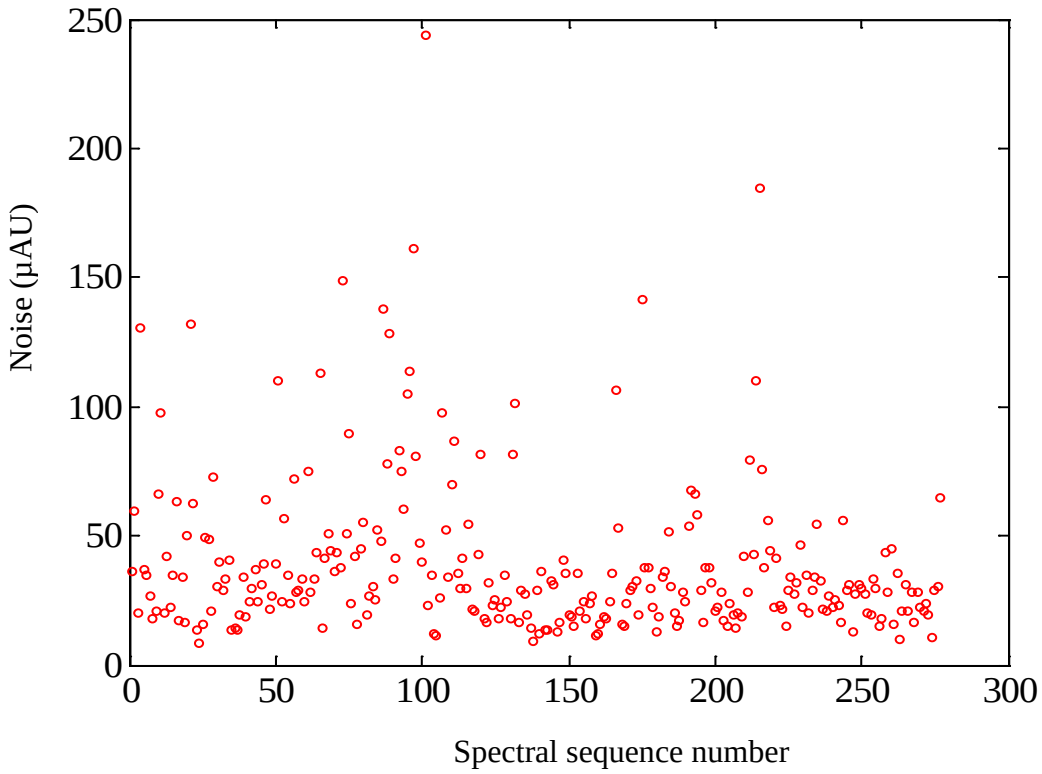
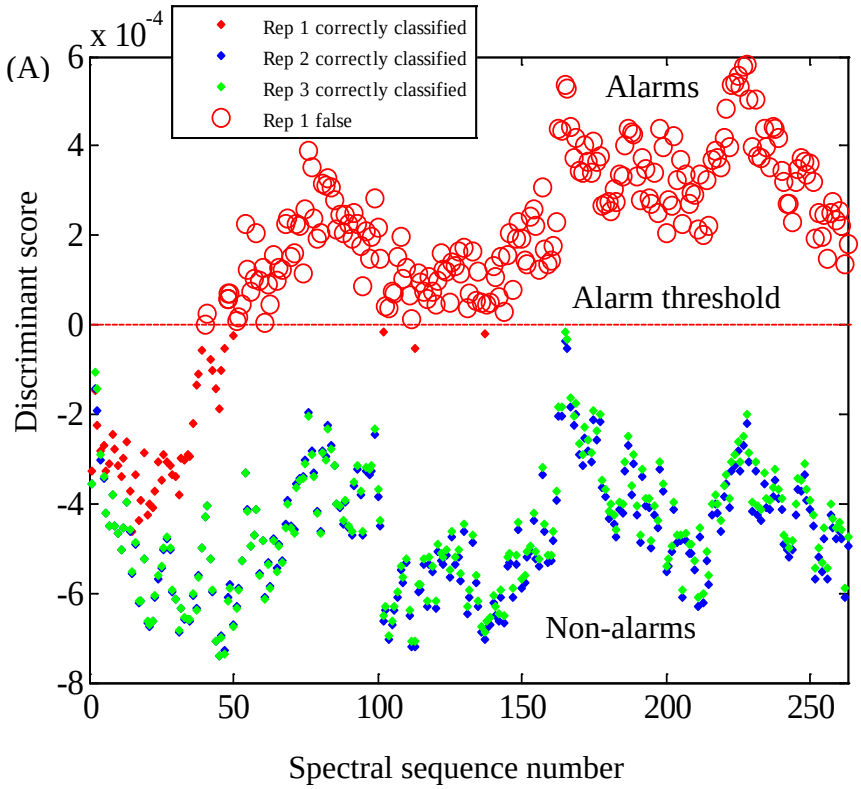


Figure 7.55. Values of RMS noise (in µAU) computed for each signal-averaged spectrum collected on the third prediction day for rat 134. The 4500-4300 cm^{-1} region was used for the calculation. Noise values were computed by fitting the 100% line in the given wavenumber region to a third-order polynomial and computing the noise about the fit. While the average is consistent with the previous days, a greater degree of fluctuation is observed. This could indicate more rat movement during the experiment.


(A)
x 10^-4
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Rep 1 false
Alarms
Alarm threshold
Non-alarms
Discriminant score
Spectral sequence number
6
4
2
0
-2
-4
-6
-8
0
50
100
150
200
250

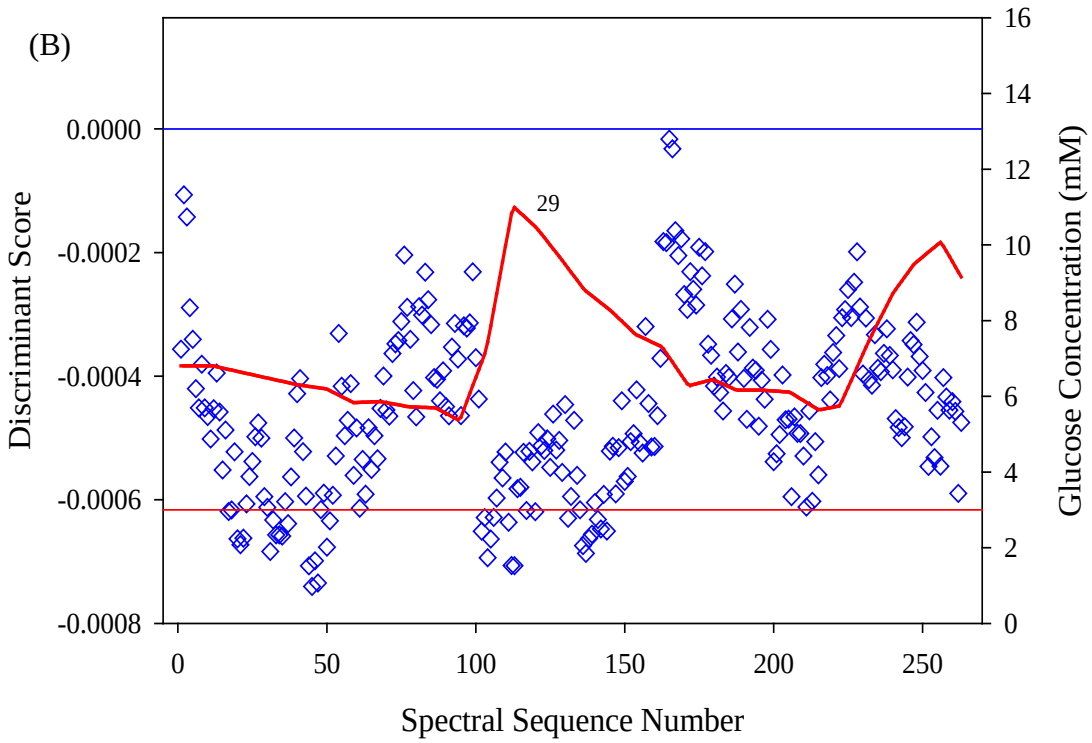


Figure 7.56. A. Discriminant score plot for the third prediction day of rat 134 with three replicate classifiers. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). The displayed glucose concentrations correspond to a lag time of 29 minutes.

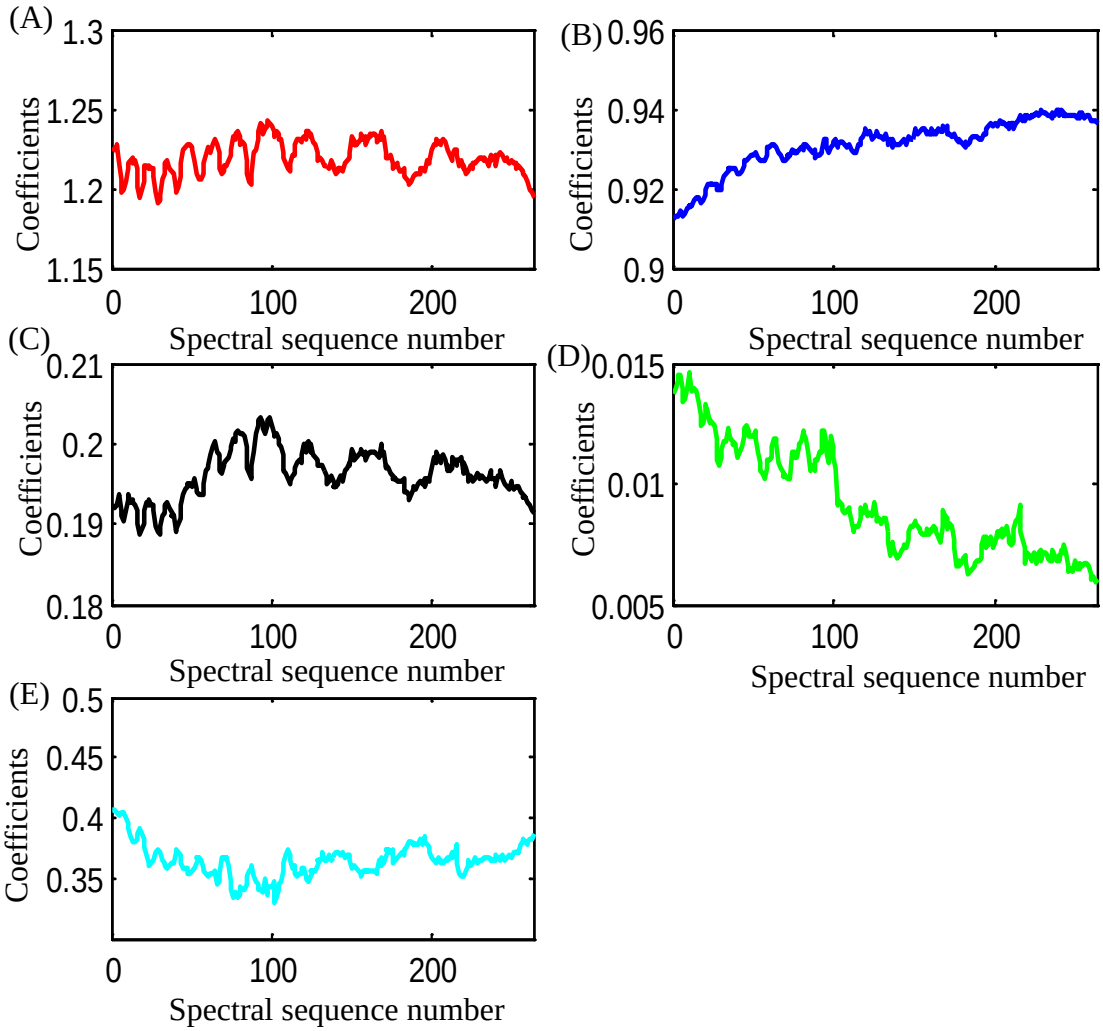


Figure 7.57. Regression coefficients computed for each of the spectra collected for the third prediction day of rat 134. Panels A-E correspond to regression coefficients for collagen, water, keratin, fat, and an intercept term. The regression coefficients represent the thickness of each component in the skin tissue matrix.

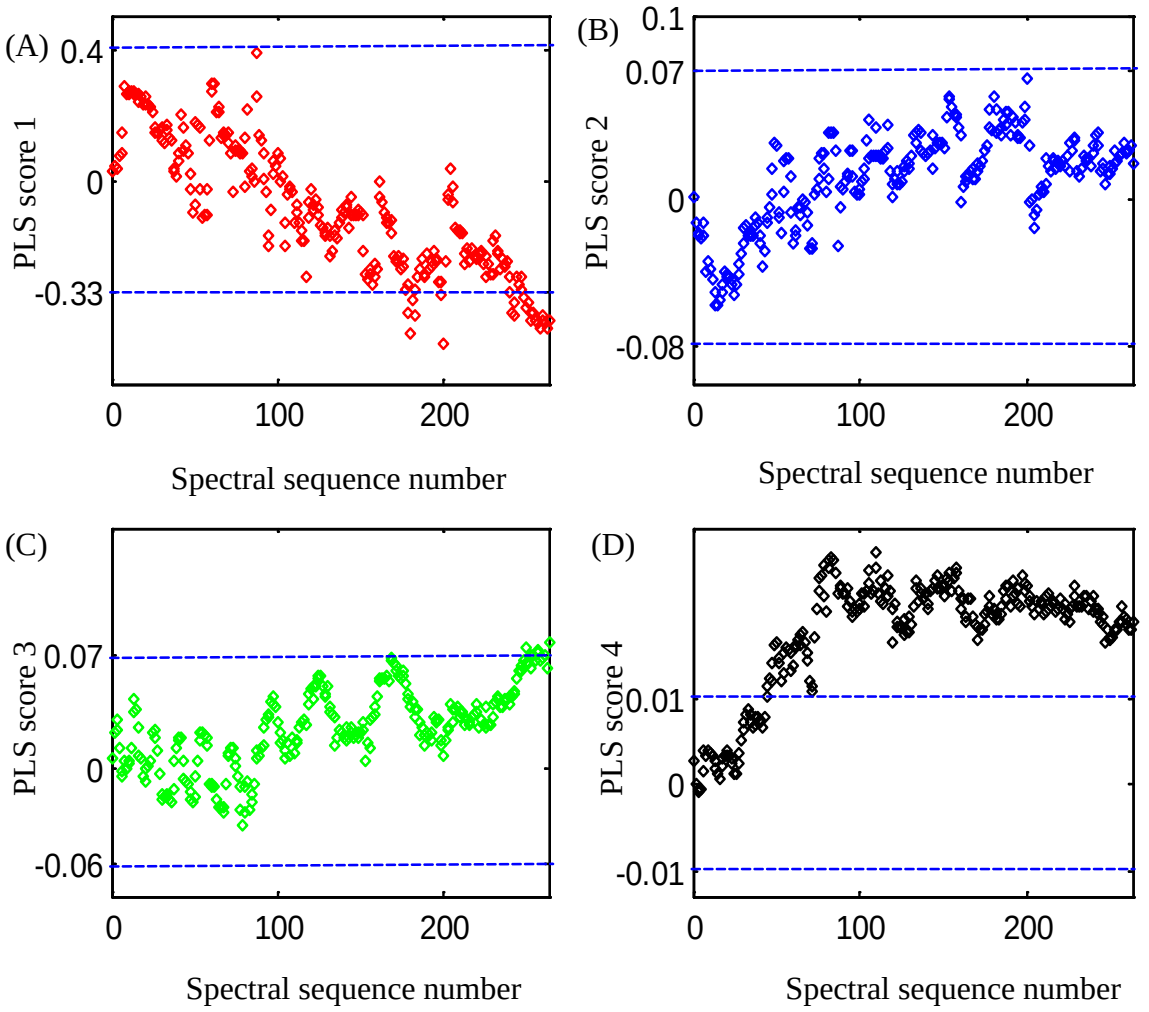


Figure 7.58. Plots of PLS scores vs. spectral sequence number for the prediction differential spectra collected on the third prediction day for rat 134. The 4850-4250 cm^{-1} wavenumber range was used for the PLS calculation. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as dashed lines. Panels A, B, C, and D correspond to latent variables 1, 2, 3, and 4, respectively.

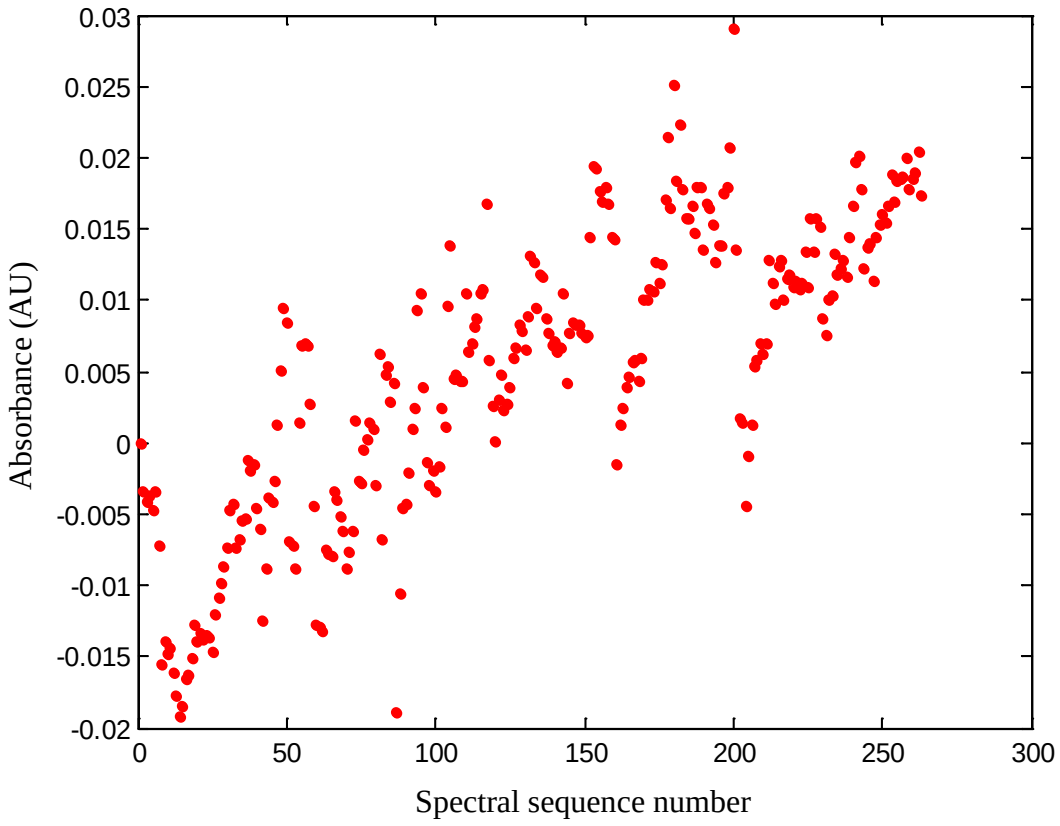


Figure 7.59. Absorbance at 4400 cm^{-1} plotted vs. the spectral sequence number for the differential spectra collected on the third prediction day for rat 134. As with the previous days, drift in the absorbance is noted over the time course of the experiment.

Each of the spectra collected on the fourth prediction day for rat 134 was assigned to a glucose concentration using the calibration lag time of 29 minutes. The same assumption used in predicting the first three prediction days discussed previously was used here for the prediction of the fourth day and the spectra collected during the first 15 minutes were removed from the data prediction. The glucose concentration profile for the fourth prediction day is shown in Figure 7.60. The signal-averaged single-beam spectra were used for the RMS noise calculations and with 128 scans and a third-order polynomial fit, the average RMS noise value was 69 μAU. This is the highest noise level obtained for any of the days for rat 134. A plot of the noise values is provided in Figure 7.61.

The first spectrum after removing the first 15 signal-averaged spectra had a concentration of 8.08 mM and was used as the reference spectrum. The alarm threshold was 3.00 mM and the critical concentration was -5.08 mM. The PLS scores computed with the optimal wavenumber-latent variable combination were partitioned into 2984 alarm and 23,812 non-alarm patterns. Three replicate classifiers were computed with the calibration PLS patterns and the first two discriminants from each of the replicates were used for the prediction. On average, the three replicate classifiers separated 38 % of the alarm patterns in the calibration set. This is again a low degree of separation.

Classification performance declined sharply on the fourth prediction day. Data for the fourth prediction day were collected 21 days after the calibration day. Using the previously optimized 15 latent variables, the committee classifier produced no missed

alarms out of the 53 alarm patterns present. However, there were 233 false alarms out of 256 non-alarm patterns.

Another PLS factor was added to check the possibility of improving the prediction of the glucose transients. The optimum wavenumber range of 4850-4250 cm^{-1} and 16 latent variables were thus used to predict the glucose transients on the fourth prediction day. When the alarm decision rule was applied, there were 23 missed alarms out of 53 alarm patterns and 21 false alarms out of 256 non-alarm patterns. This corresponded to the detection of 56.6 % of the alarm events and 8.2 % false detections. The discriminant scores plot is shown in Figure 7.62A. The committee result for the data prediction is shown in Figure 7.62B.

In addition to the concentration profile based on a lag time of 29 minutes, Figure 7.62B plots concentration profiles based on lag times of 0 and 15 minutes. Comparison of the three concentration profiles reveals that for this day, the lag time of 29 minutes is suspect. For example, when a lag time of 15 minutes was used for the data prediction it was observed to improve the data prediction. The number of missed alarms decreased to 12 out of 52 and the false alarms totaled 11 out of 257 non-alarm patterns. This corresponded to the detection of 77.0 % of the alarm events and 4.3 % false alarms. Imprecision in the determination of the lag time is clearly a key problem with the current methodology.

It was hypothesized that the poorer prediction results for the fourth prediction day might arise from the improper selection of the reference spectrum. The optical path length and the spectral background matrix might change during the data collection and use of the initial spectrum as a reference for predicting all the spectra may not be

sufficiently accurate. To check this, the reference spectrum was updated every 60 minutes. For this work, the lag time remained at 29 minutes as determined originally. The PLS scores used were based on a range of 4850 to 4250 cm^{-1} and 16 scores were used to generate the patterns submitted to PLDA. Each discriminant computed with PLDA was based on two discriminant functions.

The glucose concentration profile for the first hour is shown in Figure 7.63 and the corresponding discriminant score plot is given in Figure 7.64A along with the committee result in Figure 7.64 B. There were no false alarms out of 60 non-alarm patterns. The reference spectrum was then updated from the 61^{st} spectrum. The reference concentration was 7.64 mM. This reference was used for the next two hours as the end of the second hour corresponded to a region of the concentration profile that could not be used as a reference. For example, the 121^{st} spectrum had a reference concentration of 14.2 mM that was too high for use with the calibration database (i.e., the critical concentration was such that there were no alarm patterns in the calibration set). The concentration profile and the corresponding discriminant score plots are shown in Figures 7.65 and 7.66 A, respectively. The committee result is shown in Figure 7.66 B. There were no false alarms out of 120 non-alarm patterns.

The reference spectrum was then updated from the 181^{st} spectrum. The reference concentration was 8.96 mM. The glucose concentration profile collected during the next hour had 20 alarm and 40 non-alarm patterns. The glucose transient is shown in Figure 7.67. The PLDA prediction yielded 6 missed alarms out of 20 alarm patterns and 23 false alarms out of 40 non-alarm patterns as shown in Figure 7.68A. Three discriminants were used for this particular case since the third discriminant also separated a significant

number of patterns when compared to first two discriminant functions. The committee result is shown in Figure 7.68 B.

The 241st spectrum had an alarm concentration and thus could not be used as a reference. The 287th spectrum was then used as the reference which had a concentration of 3.63 mM. There were 34 alarm patterns and 15 non-alarm patterns within this time period. Applying PLDA resulted in 1 missed alarm (97.1 % successful detection) and 5 false alarms (occurrence of 33.3 %). The corresponding glucose concentration profile and the discriminant score plots are shown in Figures 7.69 and 7.70 A, respectively. The corresponding committee result is shown in Figure 7.70 B.

The overall committee result for the entire run after applying the reference updating strategy is shown in Figure 7.71. Using the original lag time of 29 minutes, the combined result exhibited 7 missed alarms out of 53 alarm patterns and 27 false alarms out of 256 non-alarm patterns. This corresponded to the detection of 86.8 % of the alarm events with a false alarm occurrence of 10.5 %.

Inspection of Figure 7.71 again suggests the originally determined lag time of 29 minutes may be too long for the fourth prediction day. This is especially true in the location of the initial transition into the alarm state just past spectrum 200. The situation is less clear at the end of the run, however. Here the concentration profile hovers near the alarm threshold and precise assignment of alarm and non-alarm designations is impossible. If the lag time were reduced to 15 minutes, the number of missed alarms would be 7 out of 52 alarm patterns (detection of 86.5 % of alarm events). There would be 30 false alarms out of 257 non-alarm patterns (occurrence of 12.6 %).

Overall, the reference updating strategy improved the data prediction significantly, especially the number of missed alarm patterns. This confirms that the current methodology remains dependent on the assumption that the reference spectrum characterizes the background adequately. Going forward, an automated ability to determine when the reference needs to be updated is clearly needed.

The regression coefficients obtained from the fit of the pure components to each of the spectra collected for the fourth prediction day of rat 134 are shown in Figure 7.72. This figure shows the thickness variations of each individual component during the data collection. Nothing remarkable is observed for the fourth prediction day. There is some oscillation in the regression coefficients but no dramatic changes during the course of the run.

The first four prediction PLS scores for the fourth prediction day are plotted with respect to the spectral sequence numbers in Figure 7.73. Scores 3 and 4 fall within the range defined by the corresponding calibration scores. For the plots of PLS scores 1 and 2, it can be observed that some prediction PLS scores occupy a space outside the range defined by the calibration scores, especially the missed alarm patterns at spectral sequence numbers 200-22.

The differential absorbance value at 4400 cm^{-1} for the fourth prediction day was plotted vs. the spectral sequence number to check for spectral variance at the most important glucose absorption band at 4400 cm^{-1}. As noted previously for the other prediction days, it can be clearly seen from Figure 7.74 that the absorbance value shows a sequential drift with time.

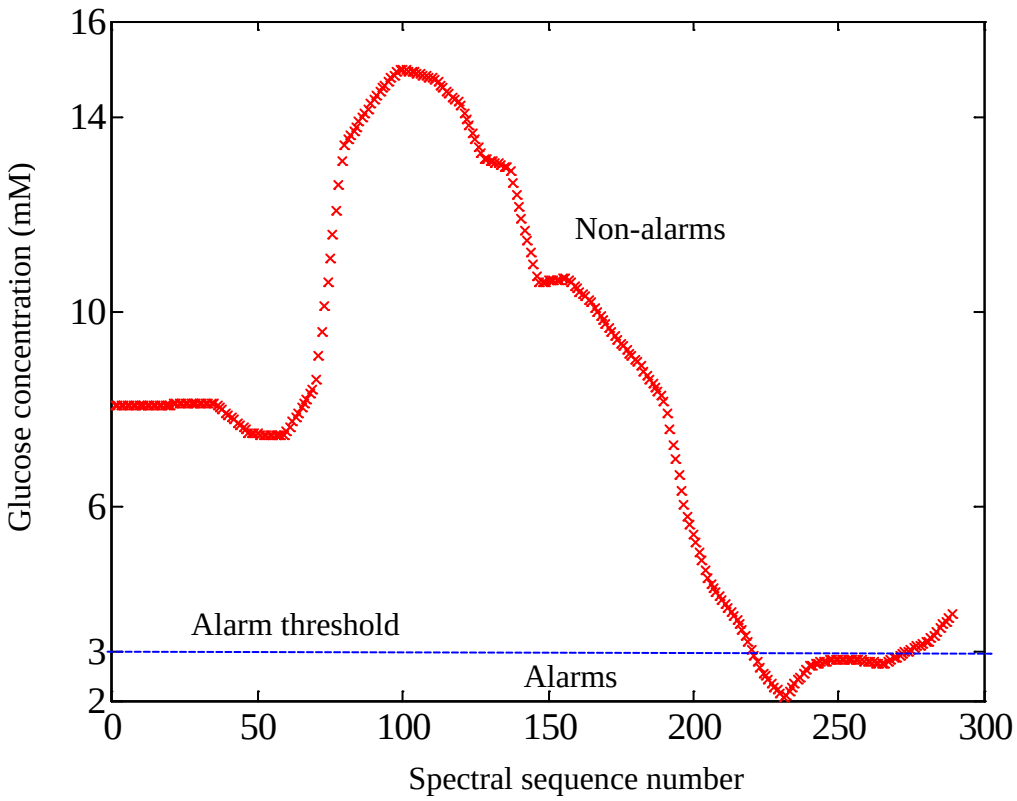


Figure 7.60. Glucose concentration profile for the fourth prediction day of rat 134. The concentration profile contained 289 total spectra, 53 alarms and 256 non-alarms for a 3.00 mM alarm threshold (horizontal line).

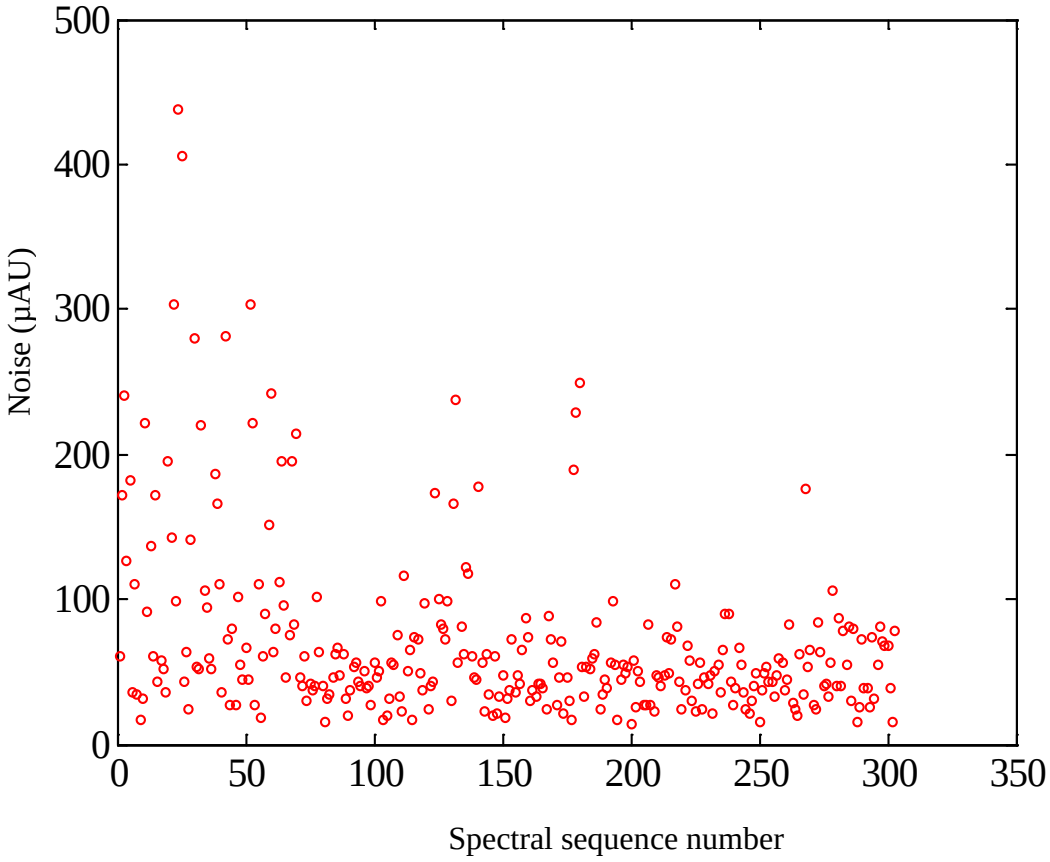
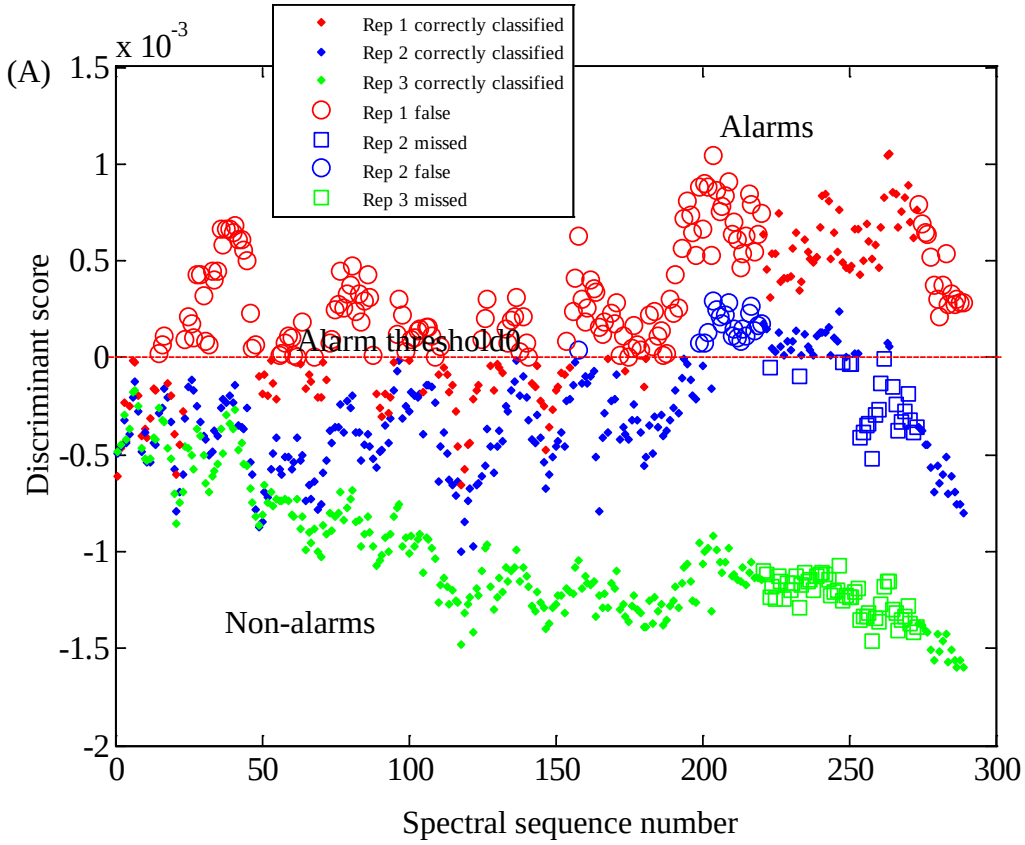


Figure 7.61. Plot of RMS noise values (in µAU) computed for each signal-averaged spectrum collected on the fourth prediction day for rat 134. The 4500-4300 cm^{-1} region was used for the noise calculation. Noise values were computed by fitting the 100% line in the given wavenumber region to a third-order polynomial and computing the noise about the fit.


(A)
x 10^-3
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Rep 1 false
Rep 2 missed
Rep 2 false
Rep 3 missed
Alarms
Alarm threshold0
Non-alarms
Discriminant score
Spectral sequence number
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
50
100
150
200
250
300

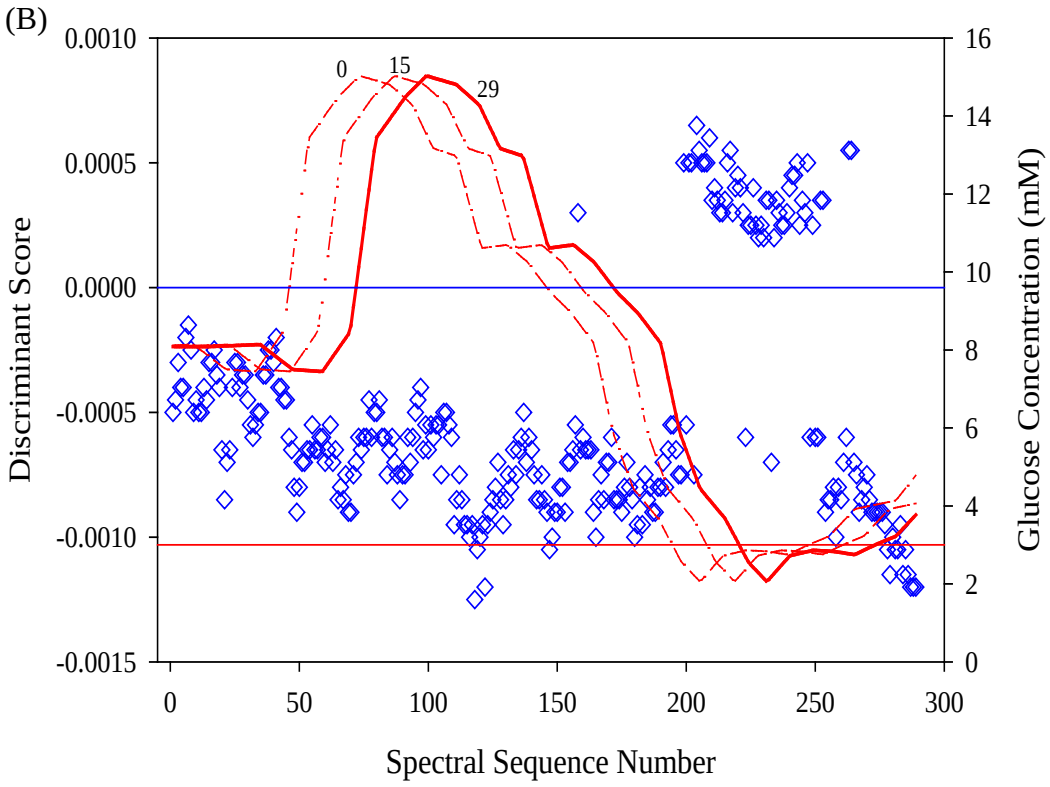


Figure 7.62. A. Discriminant score plot for the fourth prediction day of rat 134 with three replicate classifiers. The first spectrum was used as the reference. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). Glucose concentration profiles corresponding to lag times 0, 15 and 29 minutes are shown in the figure. From these results, a lag time closer to 15 minutes appears to improve the classification results.

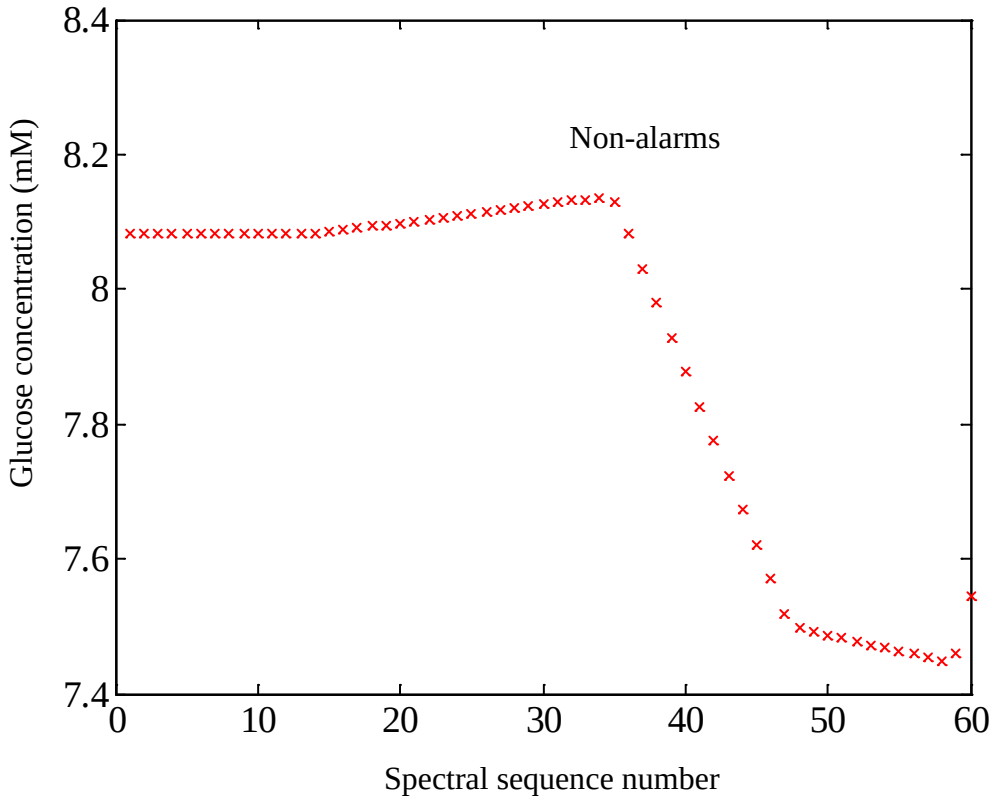
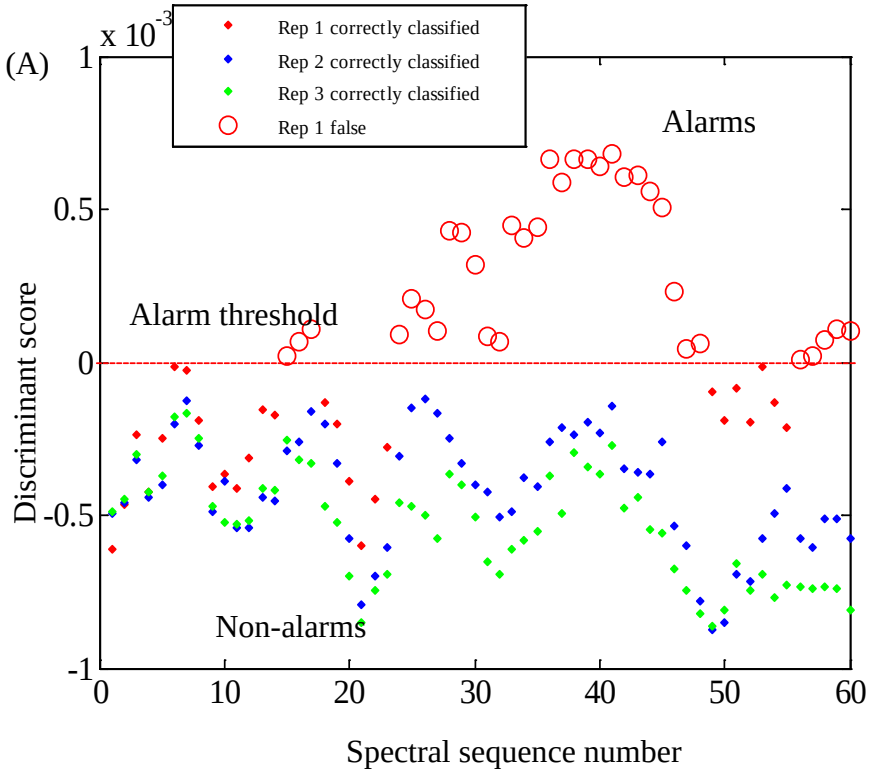


Figure 7.63. Glucose concentration profile for the first hour of the data collection of the fourth prediction day of rat 134. The concentration profile contained 60 total spectra and all were non-alarm patterns.


(A)
x 10^-3
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Rep 1 false
Alarms
Alarm threshold
Non-alarms
Discriminant score
Spectral sequence number
1
0.5
0
-0.5
-1
0
10
20
30
40
50
60

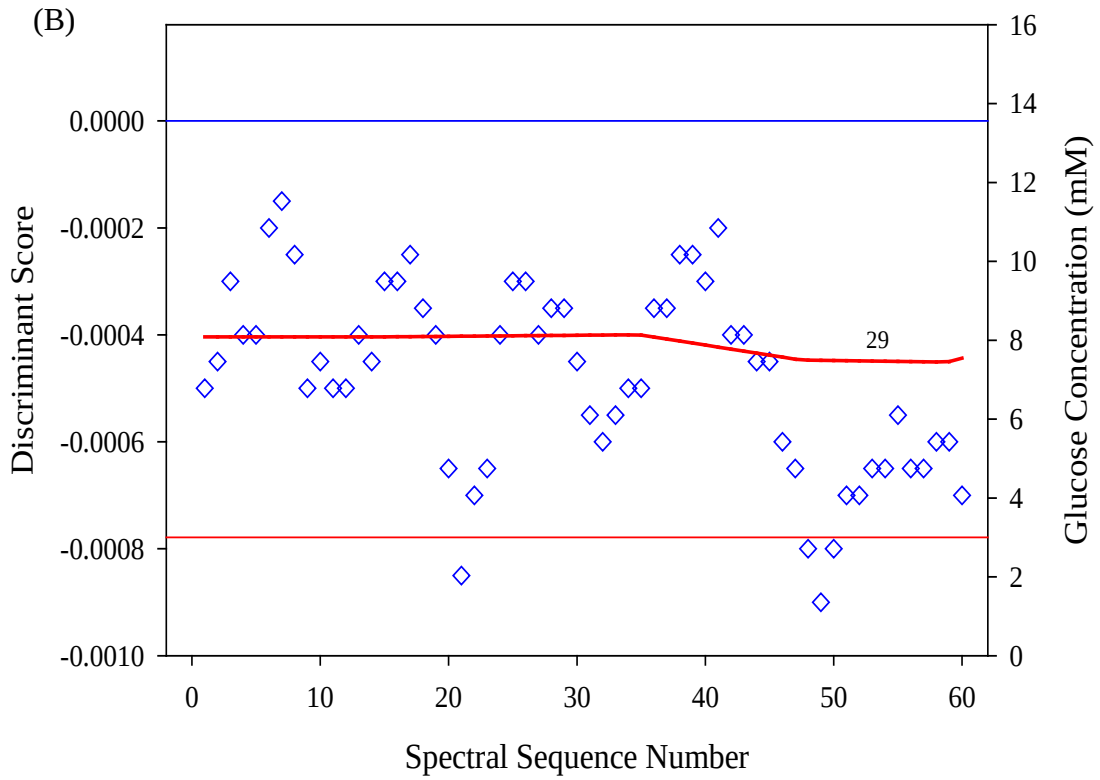


Figure 7.64. A. Discriminant score plot for the first hour of data collection on the fourth prediction day of rat 134 with three replicate classifiers. The first spectrum was used as the reference. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). The glucose concentration profile is based on a lag time of 29 minutes.

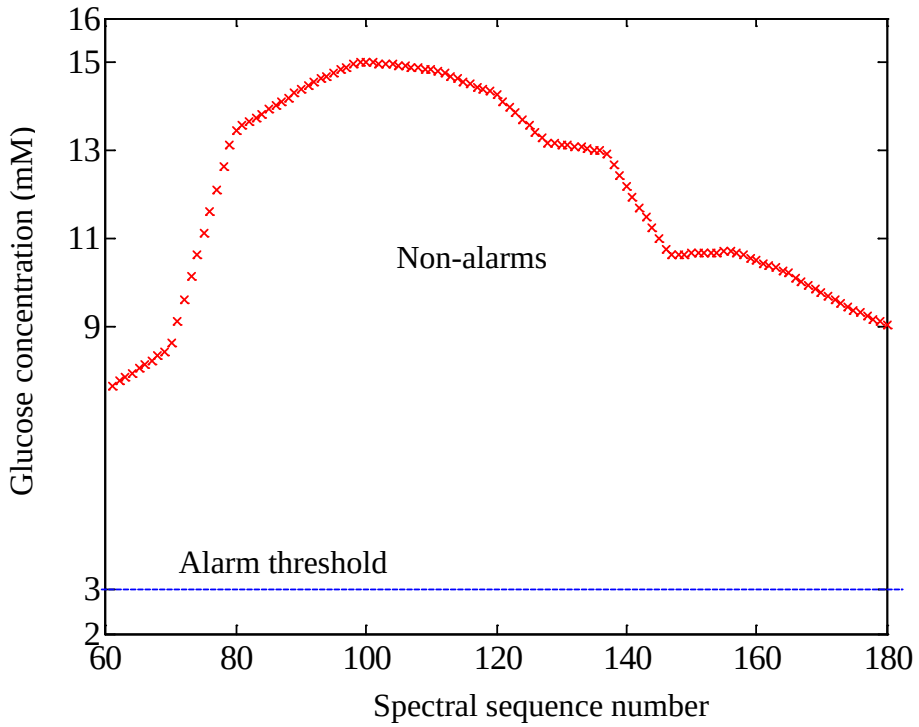
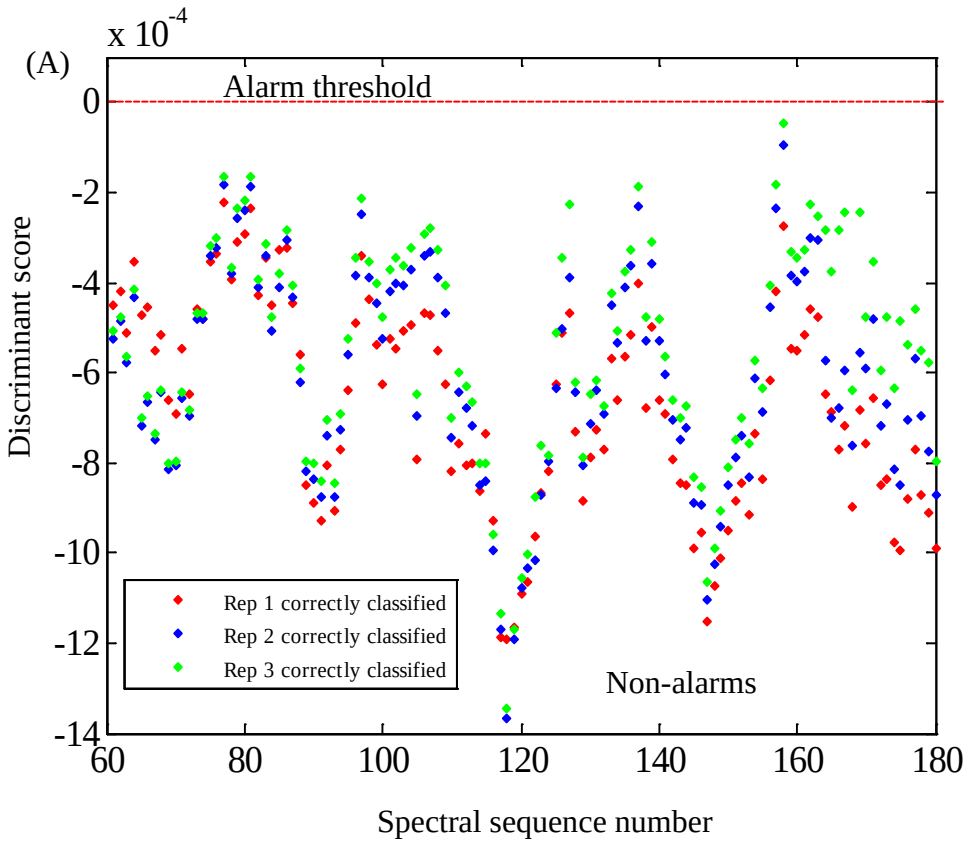


Figure 7.65. Glucose concentration profile for the data collected during 61-180 minutes of the fourth prediction day of rat 134. The concentration profile contained 120 total spectra and all were non-alarm patterns. The horizontal line marks the alarm threshold of 3.0 mM.


x 10^-4
(A)
Alarm threshold
0
-2
-4
-6
-8
-10
-12
-14
Discriminant score
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Non-alarms
60
80
100
120
140
160
180
Spectral sequence number

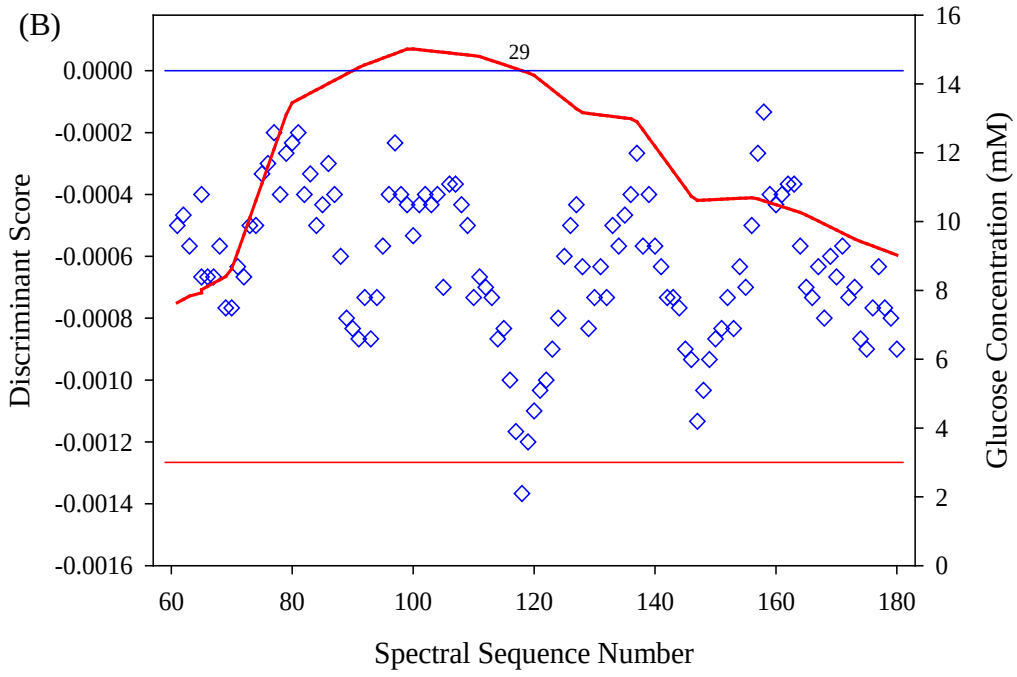


Figure 7.66. A. Discriminant score plot for 61-180 minutes of the fourth prediction day of rat 134 with three replicate classifiers. The spectrum at 61 minutes was used as the reference. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). The glucose concentration profile corresponds to a lag time of 29 minutes.

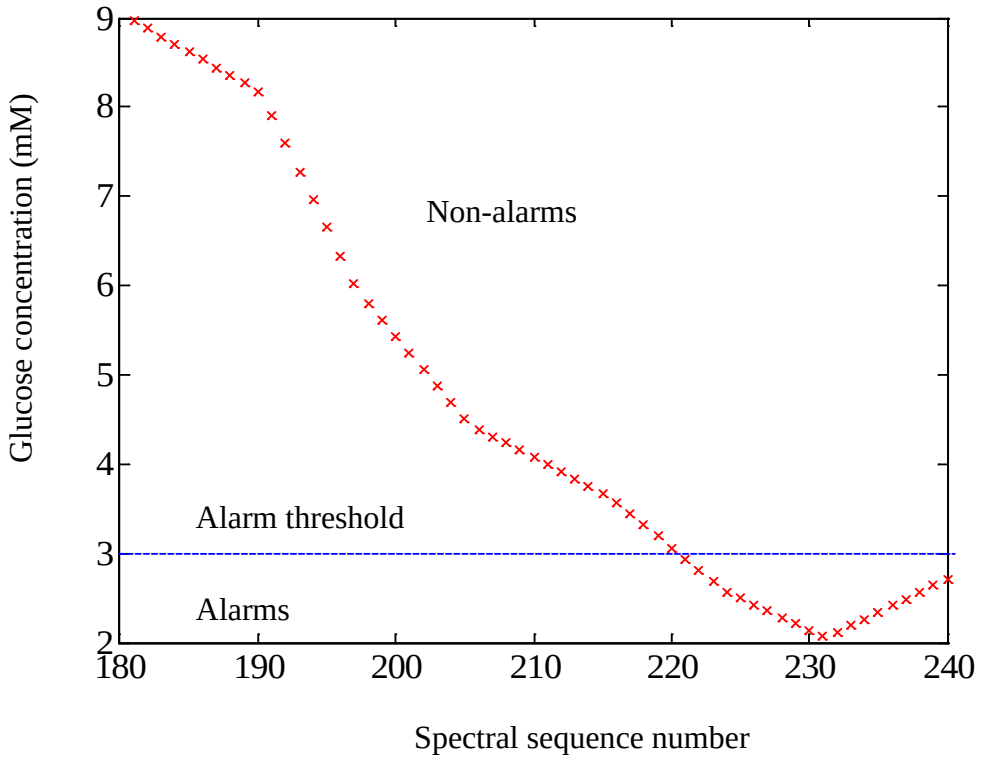
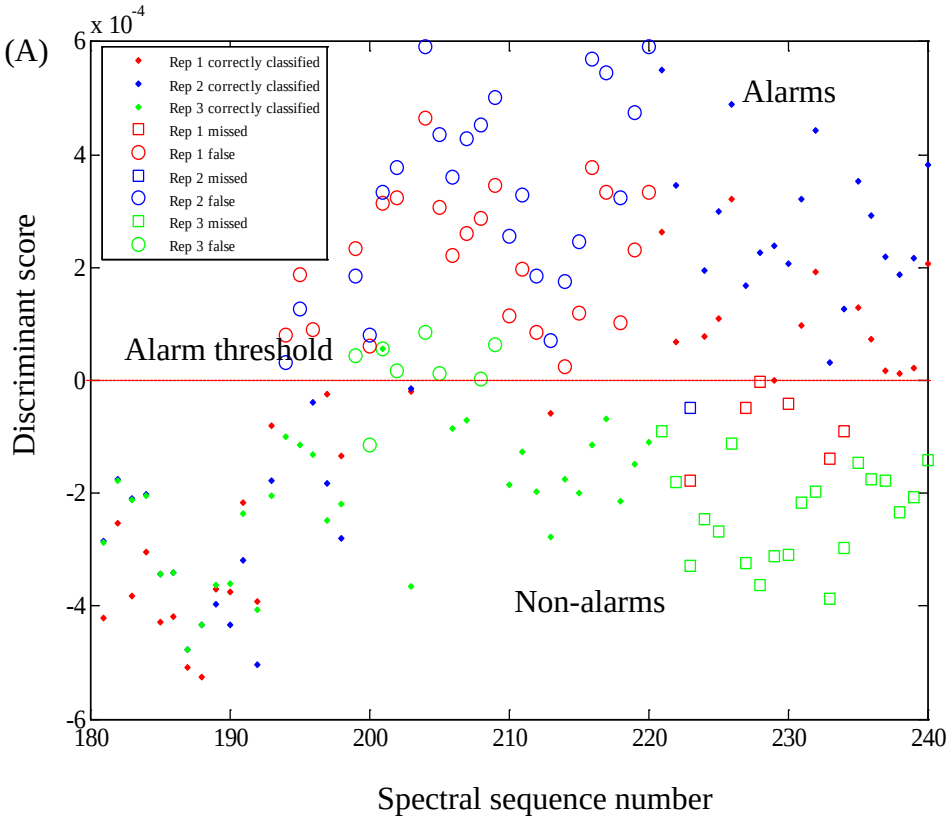


Figure 7.67. Glucose concentration profile for the data collected during 181-240 minutes of the fourth prediction day of rat 134. The concentration profile contained 60 total spectra, 20 alarm and 40 non-alarm patterns. The horizontal line specifies the alarm threshold of 3.0 mM.


(A)
x 10^-4
6
4
2
0
-2
-4
-6
180
190
200
210
220
230
240
Discriminant score
Spectral sequence number
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Rep 1 missed
Rep 1 false
Rep 2 missed
Rep 2 false
Rep 3 missed
Rep 3 false
Alarms
Alarm threshold
Non-alarms

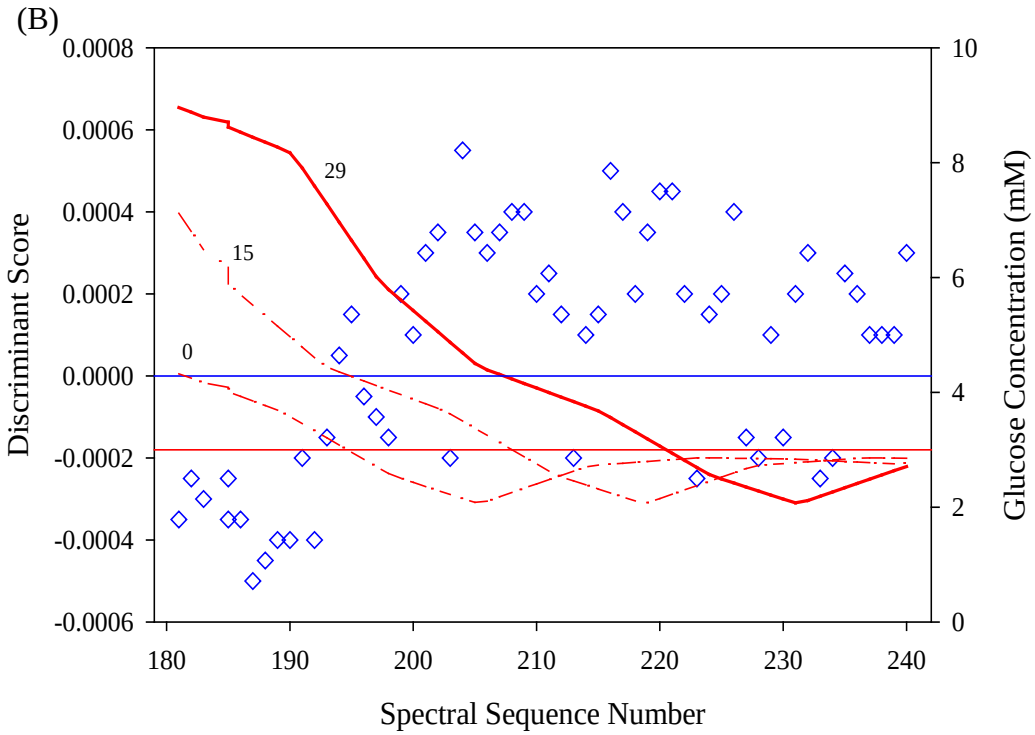


Figure 7.68. A. Discriminant score plot for 181-240 minutes on the fourth prediction day of rat 134 with three replicate classifiers. The spectrum collected at 181 minutes was used as the reference. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). Glucose concentration profiles are shown for lag times 0, 15 and 29 minutes. A lag time of 15 minutes or less appears to improve the classification results.

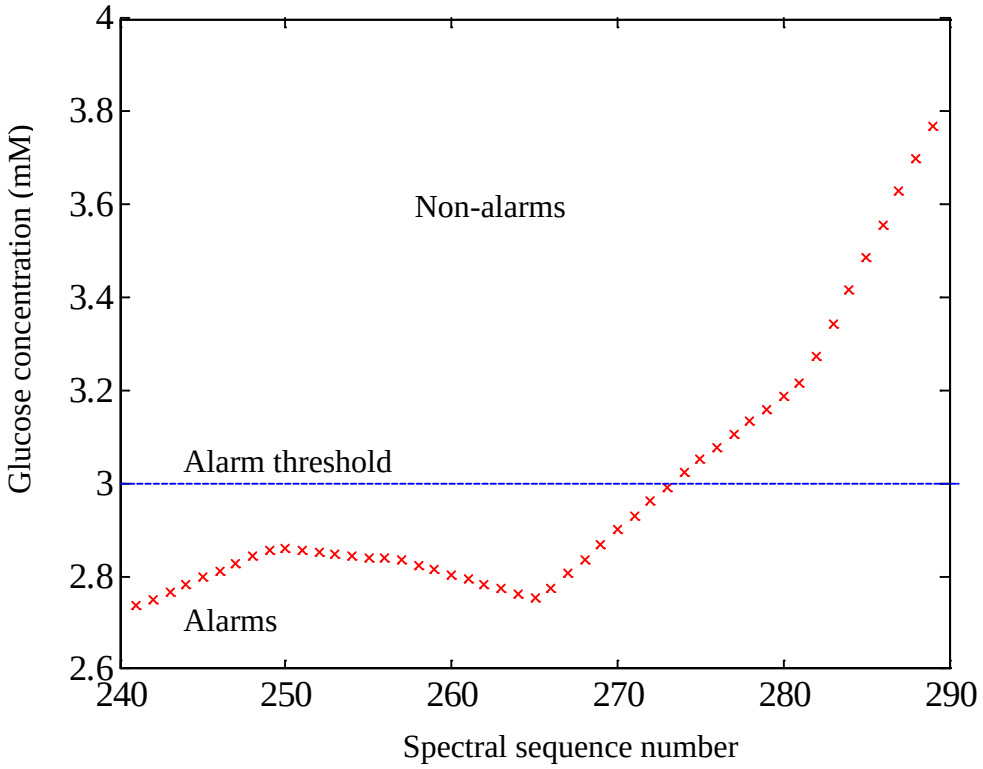
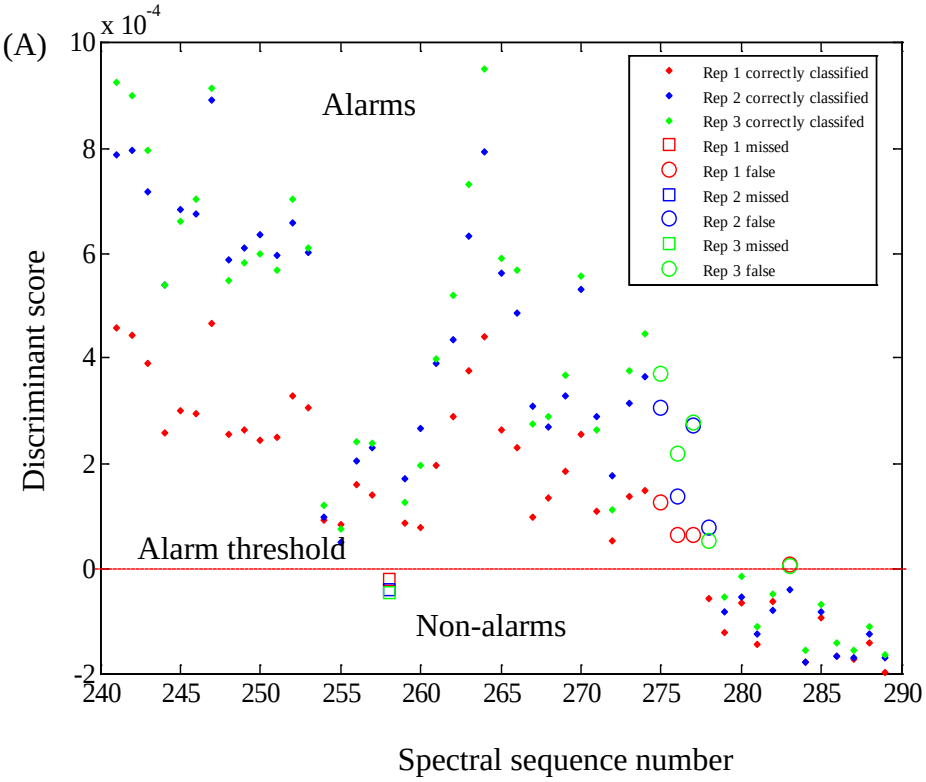


Figure 7.69. Glucose concentration profile for the data collected during 241-289 minutes of the fourth prediction day for rat 134. The concentration profile contained 49 total spectra, 34 alarm and 15 non-alarm patterns. The horizontal line identifies the alarm threshold of 3.0 mM.


(A)
x 10^-4
10
8
6
4
2
0
-2
Discriminant score
Alarms
Alarm threshold
Non-alarms
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classifed
Rep 1 missed
Rep 1 false
Rep 2 missed
Rep 2 false
Rep 3 missed
Rep 3 false
240
245
250
255
260
265
270
275
280
285
290
Spectral sequence number

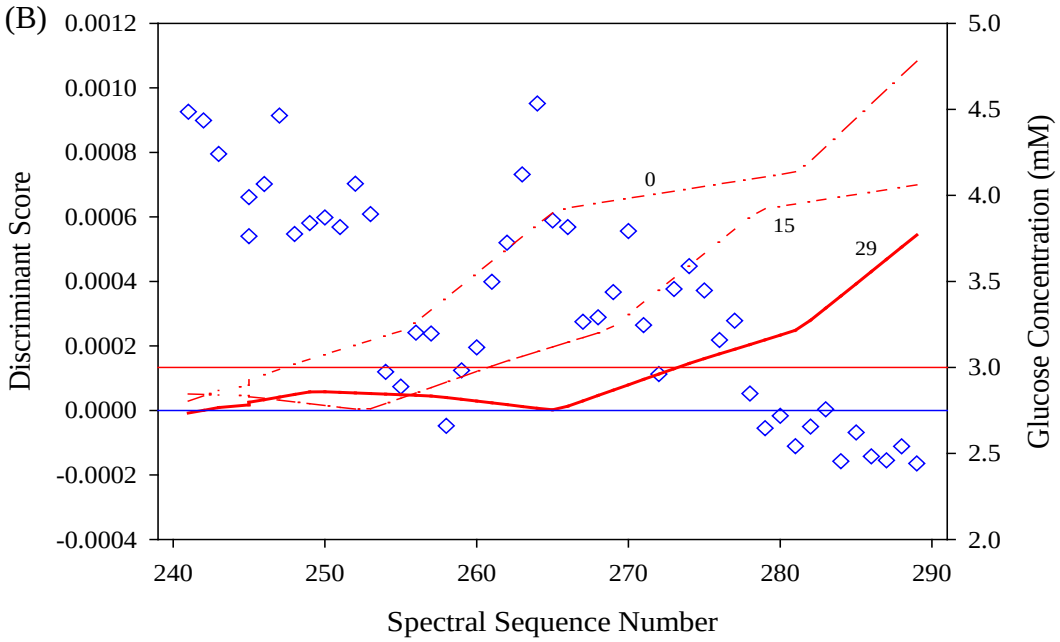


Figure 7.70. A. Discriminant score plot for 241-289 minutes on the fourth prediction day of rat 134 with three replicate classifiers. The spectrum collected at 287 minutes was used as the reference. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). Glucose concentration profiles for lag times of 0, 15 and 29 minutes are shown in the figure.

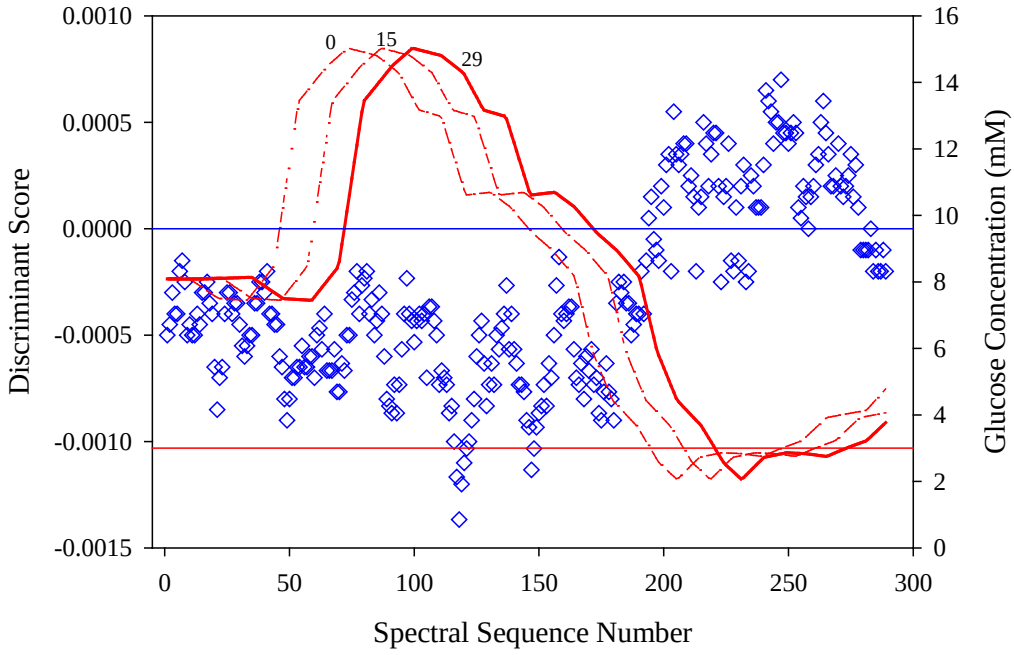


Figure 7.71. The combined discriminant scores corresponding to the committee results that incorporated reference updating are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). Glucose concentration profiles corresponding to lag times of 0, 15 and 29 minutes are shown in the figure.

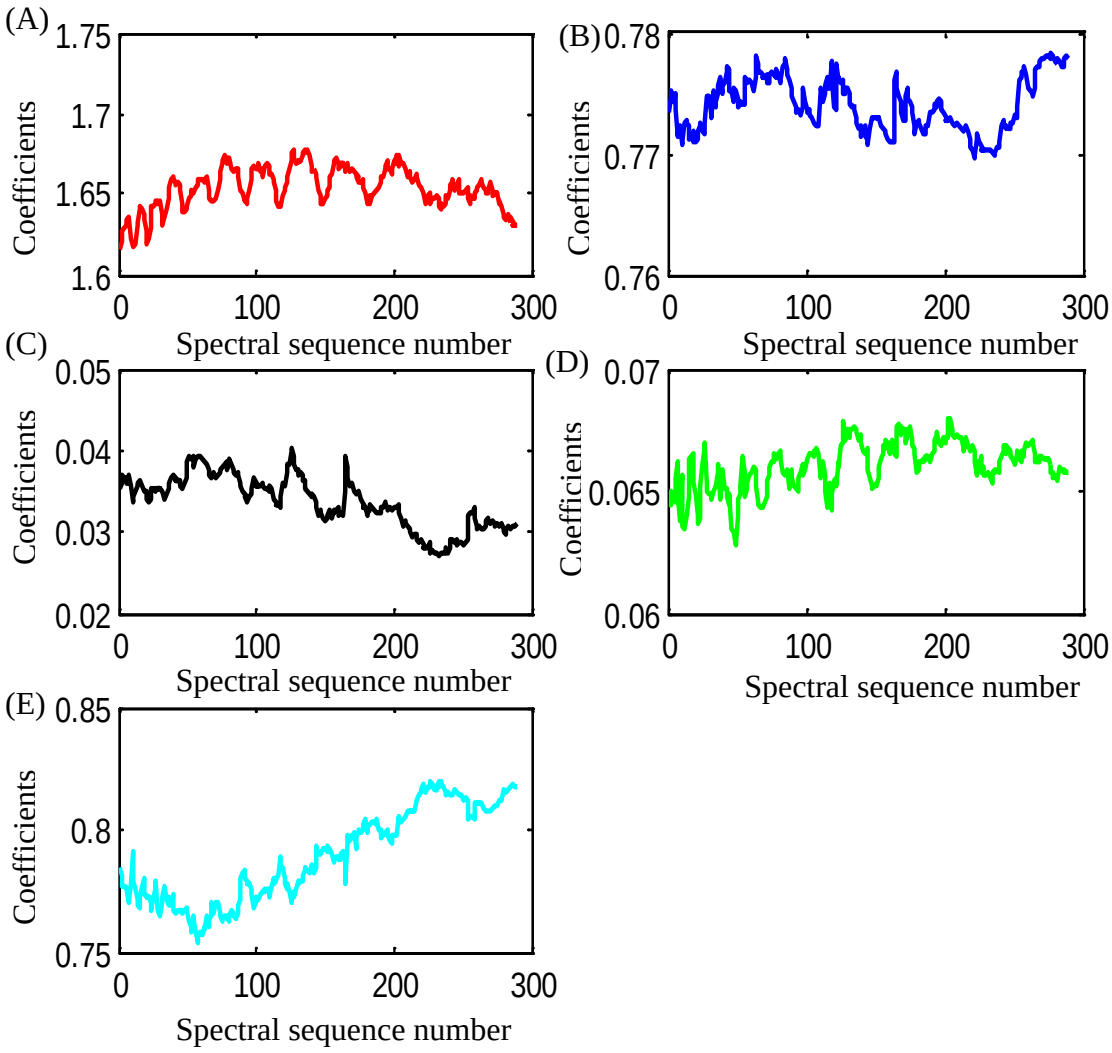


Figure 7.72. Regression coefficients computed for each of the spectra collected for the fourth prediction day of rat 134. Panels A-E correspond to regression coefficients for collagen, water, keratin, fat, and an intercept term. The regression coefficients represent the thickness of each component in the skin tissue matrix.

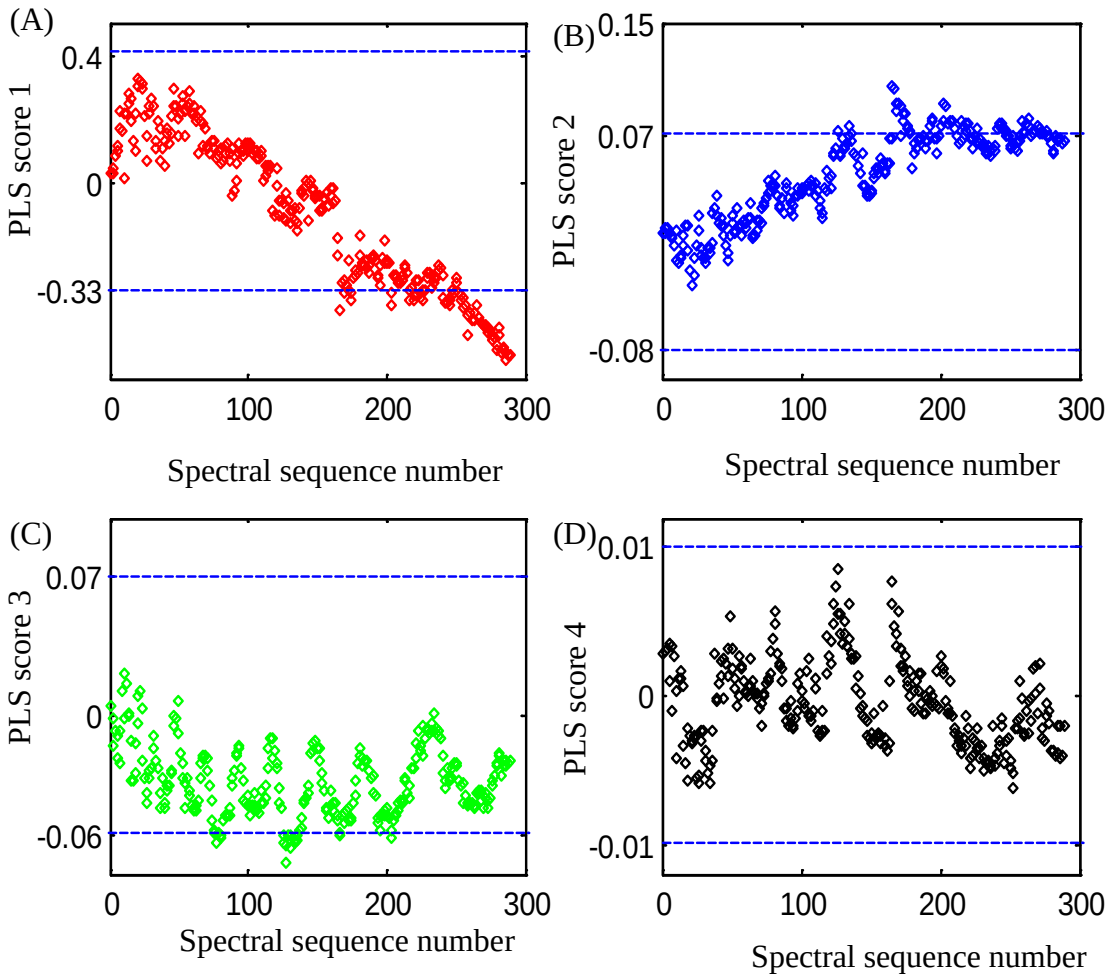


Figure 7.73. Plots of PLS scores vs. spectral sequence number for the prediction differential spectra collected on the fourth day for rat 134. The 4850-4250 cm^{-1} wavenumber range was used for the PLS calculation. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as dashed lines. Panels A, B, C, and D correspond to latent variables 1, 2, 3, and 4, respectively.

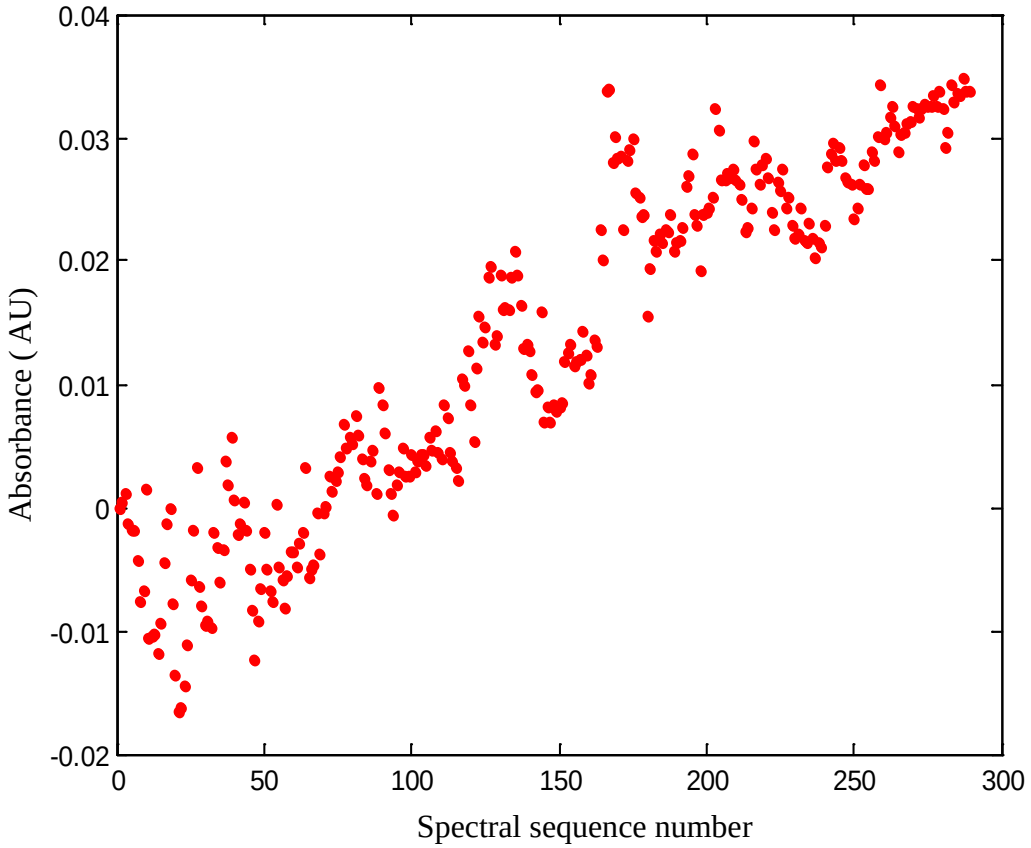
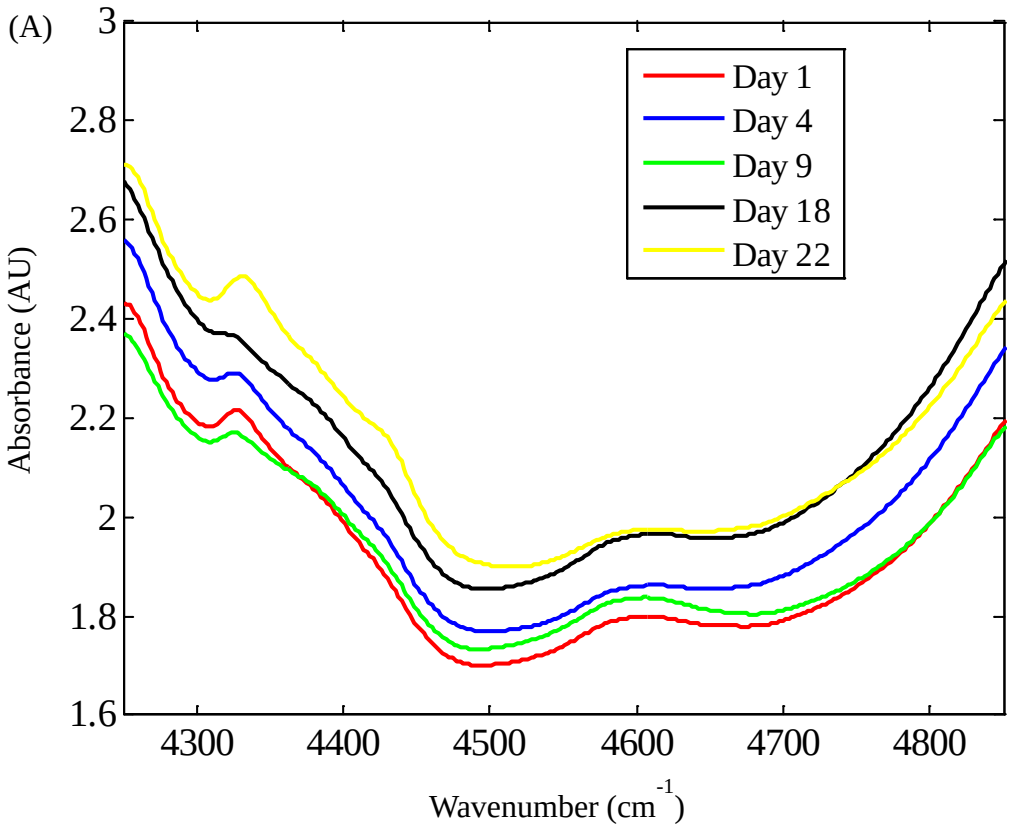


Figure 7.74. Absorbance at 4400 cm^{-1} plotted vs. the spectral sequence number for the differential spectra collected on the fourth prediction day for rat 134.

The mean absorbance spectra collected on each day of data collection for rat 134 are plotted as in Figure 7.75 A. The corresponding mean differential spectra are shown in Figure 7.75 B. It is clear that there are significant variations in the absorbance from day to day. These might be due to the specific skin tissue locations used or some other variation related to the day. These variations may the reason for the comparatively higher errors in data prediction for rat 134.

A PCA score plot produced by combining the spectra collected on all five days is shown in Figure 7.76. The spectra were mean-centered using the overall mean of the spectra and the wavenumber range of 4900 to 4200 cm^{-1} was submitted to the PCA calculation. The different clustering in the principal component scores corresponding to each day of data collection can clearly be observed.

A corresponding score plot computed from the differential spectra is shown in Figure 7.77. While the days are still distinct, it can be seen that the scores corresponding to the differential spectra are clustered together better than in Figure 7.76. The spectra corresponding to the third day of data collection are the most different from the rest in this representation. These results confirm the utility of differential spectra in helping to remove spectral variation.


(A)
Absorbance (AU)
Wavenumber (cm-1)
Day 1
Day 4
Day 9
Day 18
Day 22
3
2.8
2.6
2.4
2.2
2
1.8
1.6
4300
4400
4500
4600
4700
4800

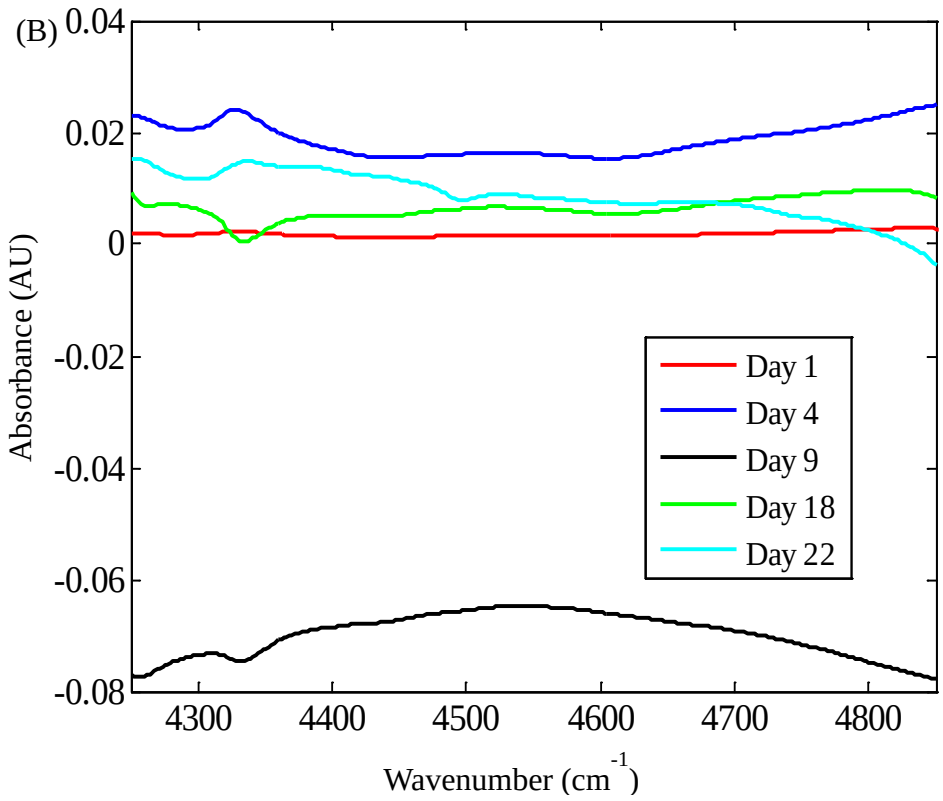


Figure 7.75. A. Mean absorbance spectra for each day of data collection for rat 134. B. Corresponding mean differential spectra.

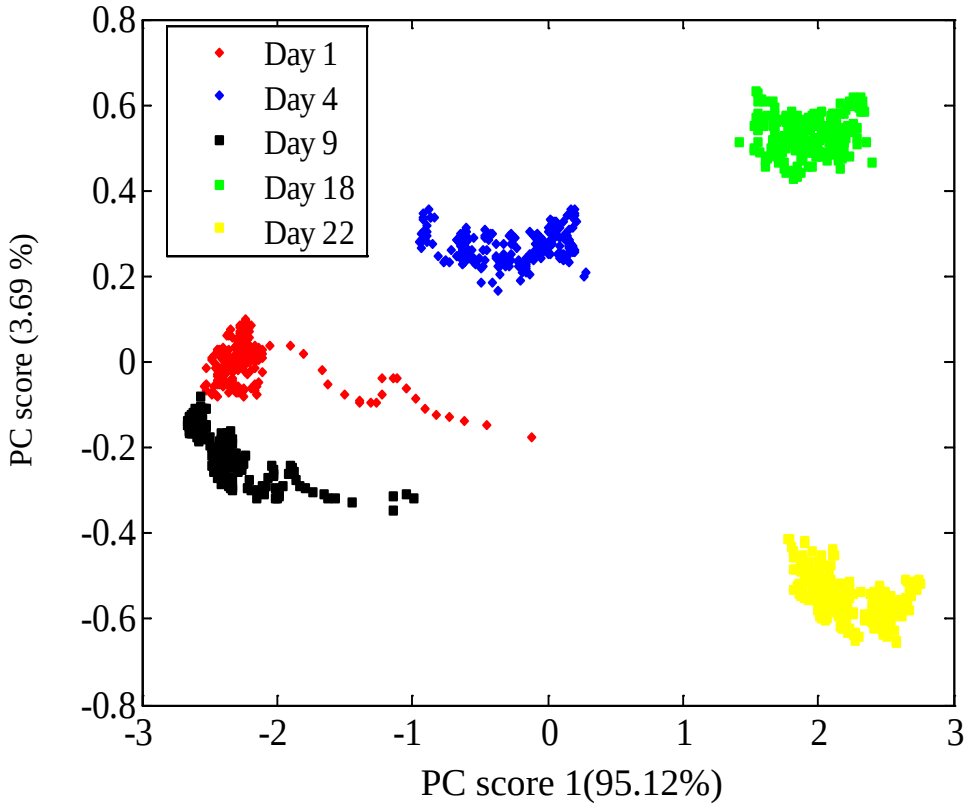


Figure 7.76. Principal component score plot for the five days of data collection for rat 134. Days are numbered relative to the calibration day (day 1). The spectra were mean-centered using the overall mean of the spectra and the region of 4900 to 4200 cm^{-1} was used. Clear clustering is observed related to the day of data collection.

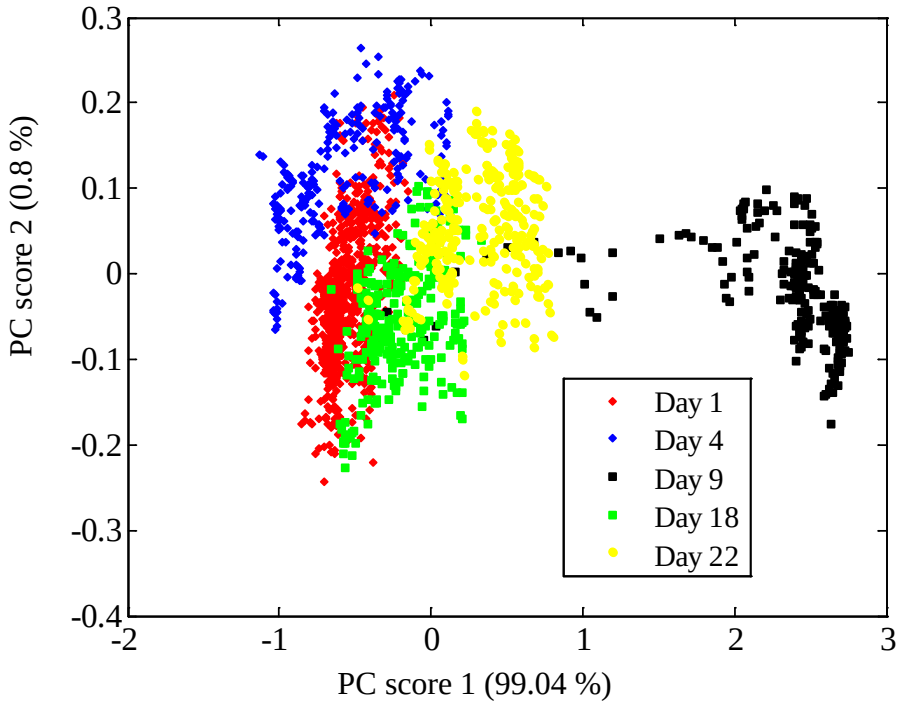


Figure 7.77. Principal components score plot for the five days of data collection for rat 134. The differential spectra computed for each day were used in computing the principal component scores. Days are numbered relative to the calibration day (day 1).

Analysis of Rat 169

The glucose transient studies for the fourth rat were performed over multiple days. The fourth rat used for data collection was supplied with ports inserted by Charles River Laboratories International, Inc. The results discussed here were acquired over three different days and the first day of data collection was used as the calibration day. The rat was awake during the data collection.

As before, spectra were collected on the basis of averaging 32 scans. For the results presented here, four consecutive single-beam spectra were averaged together to increase the SNR. The RMS noise calculated with a third-order polynomial fit for the

calibration data was ~ 97 μAU at the wavenumber region of 4500-4300 $cm^{-1}$and the corresponding noise value plot is shown in Figure 7.78. These noise values are the highest observed for any of the four rats studied.

As before, PCA was used to assess the data consistency. The first two principal component scores computed from the absorbance spectra collected on the calibration day are plotted together in Figure 7.79. The 4900 to 4200 cm^{-1} range was used, and the data matrix was mean-centered before submission to PCA. The first four spectra are separated from the rest, indicating the rat is still adjusting to the interface. These spectra were removed from building the calibration database. The remainder of the data was divided into two distinct clusters on the basis of time. Figure 7.80 plots Q vs T^2 for the calibration spectra. Spectra 1-4 are clear outliers, confirming the decision to remove them.

A cross-validation was performed with the rat skin tissue absorbance spectra to find the best lag time. Figure 7.81 plots SECV vs. the number of latent variables for each of the lag times considered. These results indicate that no lag time is needed for rat 169. This is clearly a suspicious result and must be taken into account when considering the further results obtained with rat 169. The same computation for the first and second prediction days produced lag times of 13 and 2 minutes, respectively. It was decided to proceed with the protocol as established for the three previous rats and evaluate the issue of lag time in the context of the prediction results.

The glucose concentration profile for the calibration day is shown in Figure 7.82. As before, the concentration profile was partitioned into training and monitoring sets as shown in the figure. The training and monitoring sets contained 179 and 67 signal-averaged single-beam spectra, respectively. A non-contiguous monitoring set was used to

balance the number of alarm and non-alarm patterns and to have a more standard reference concentration. A total of 15,931 differential spectra were obtained for the training set.

A grid search analysis was then performed and the top four wavenumber ranges were identified as shown in Table 7.7. The same procedures as described previously were used. Figure 7.83 shows a plot of SECV vs. the number of latent variables for the top wavenumber range that resulted from the grid search analysis. These top four wavenumber ranges and latent variables 12-16 were then used to build classification models with PLDA and applied to the monitoring data.

The reference concentration, alarm threshold concentration and the critical concentration for the monitoring set were 5.05 mM, 3.00 mM and -2.05 mM, respectively. There were 2723 alarm and 13,208 non-alarm patterns in the training set. Table 7.8 summarizes the percentage (average ± standard deviation) of missed and false alarms when PLDA was employed. Three replicate discriminants were used, and each classifier was based on three discriminant functions.

The wavenumber range and corresponding smallest number of PLS factors that gave the minimal missed and false alarm percentages was taken as the optimal combination for subsequent application to the prediction days. For rat 169, these parameters were 4850-4250 cm^{-1} and 13 PLS factors.

Inspection of the number of alarm patterns separated in the calculation of the discriminants revealed great difficulty in separating the alarm and non-alarm patterns. It was observed that most of the alarm patterns were very close to the critical concentration

of -2.05 mM. There were 1861 patterns out of 2723 within the -2.05 to -2.5 mM range. This is most likely the reason behind the poor data separation observed with this rat.

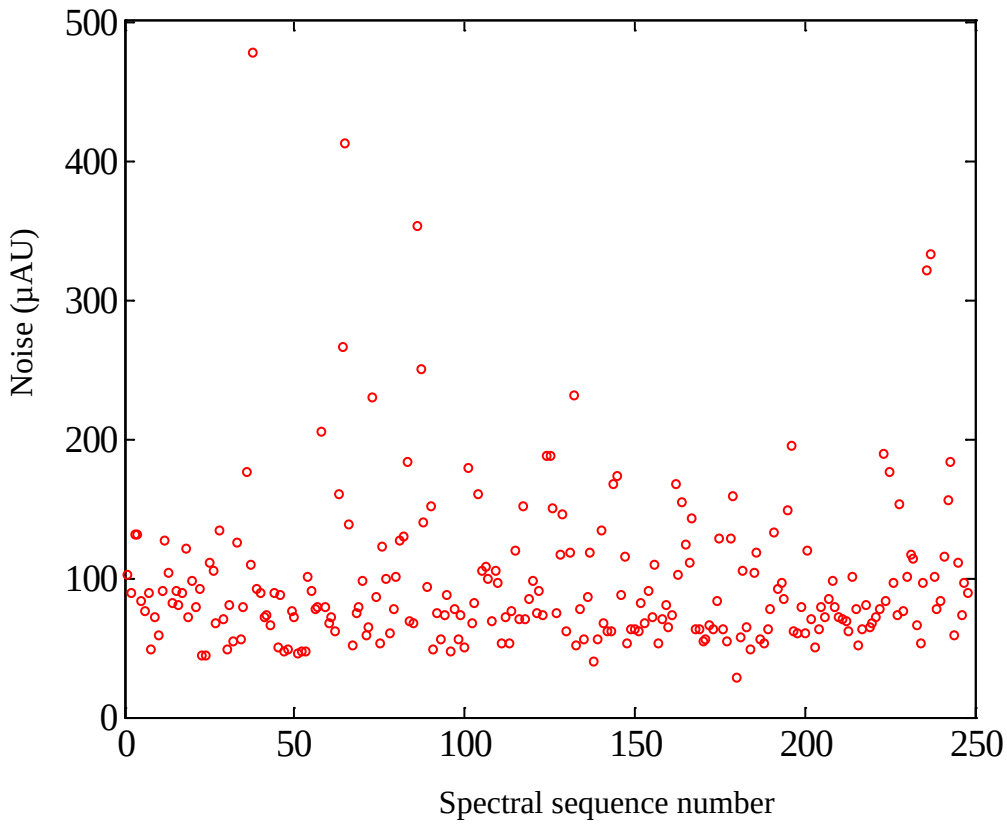


Figure 7.78. Plot of RMS noise values (in µAU) computed for each consecutive pair of signal-averaged spectra collected on the calibration day of rat 169. The 4500-4300 cm^{-1} region was used for the calculation. Noise values were computed by fitting the 100% line in the given wavenumber region to a third-order polynomial and computing the noise about the fit. These noise levels are the highest observed for the four rats studied.

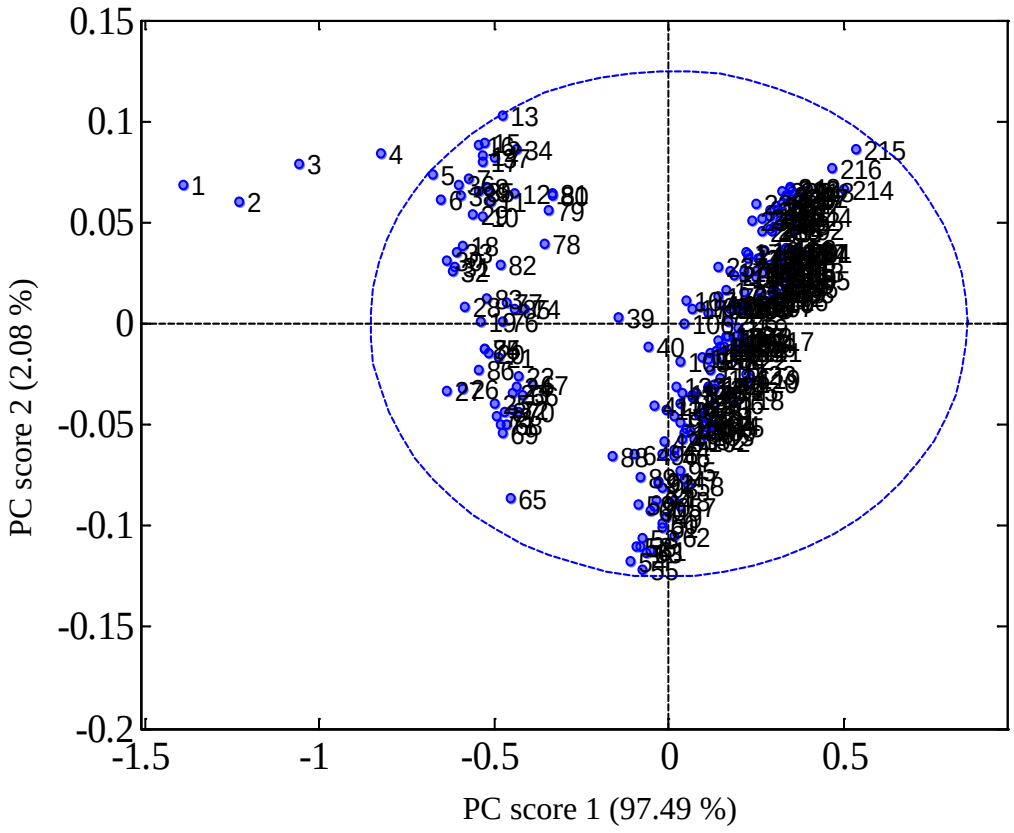


Figure 7.79. Principal component score plot for rat 169 for the calibration day. Labels correspond to the spectral sequence number. A different data clustering was observed for the first four averaged spectra. In addition, the remainder of the data occupy two distinct clusters that are grouped by time of data collection. The ellipse shows the 95 % confidence interval about the scores.

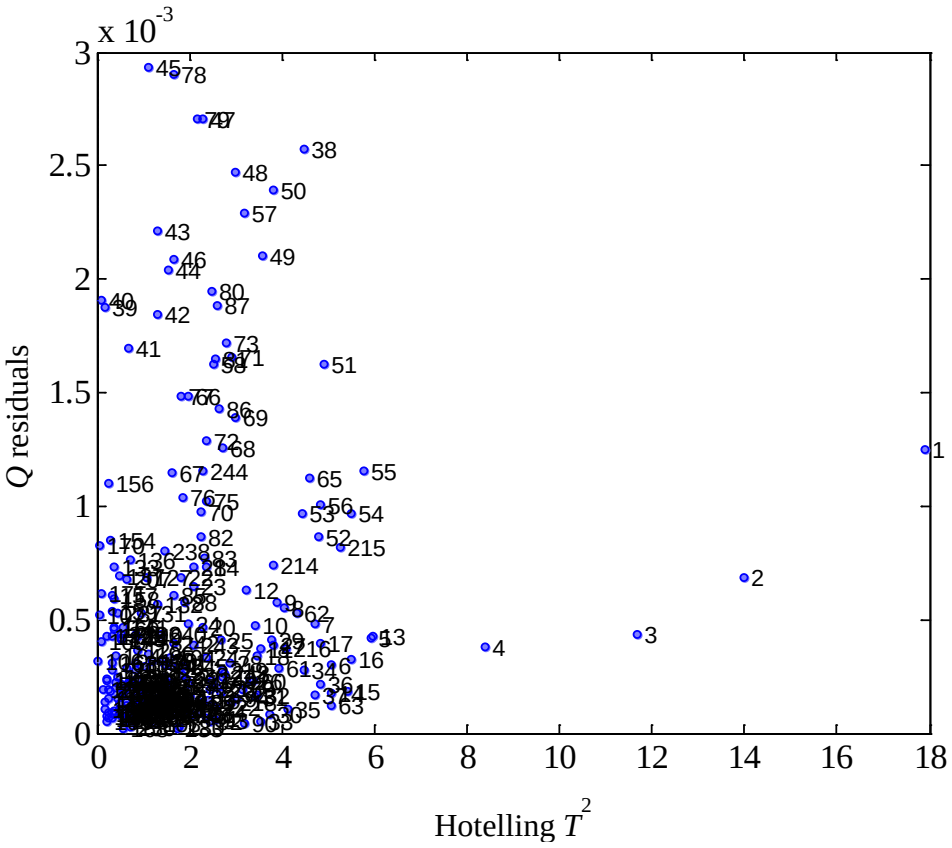


Figure 7.80. Plot of Q vs. Hotelling's T^2 for the absorbance spectra collected on the calibration day for rat 169. The first two principal components were used for the analysis. Data labels indicate the spectral sequence number. Spectra 1-4 are clear outliers and were removed from further analysis.

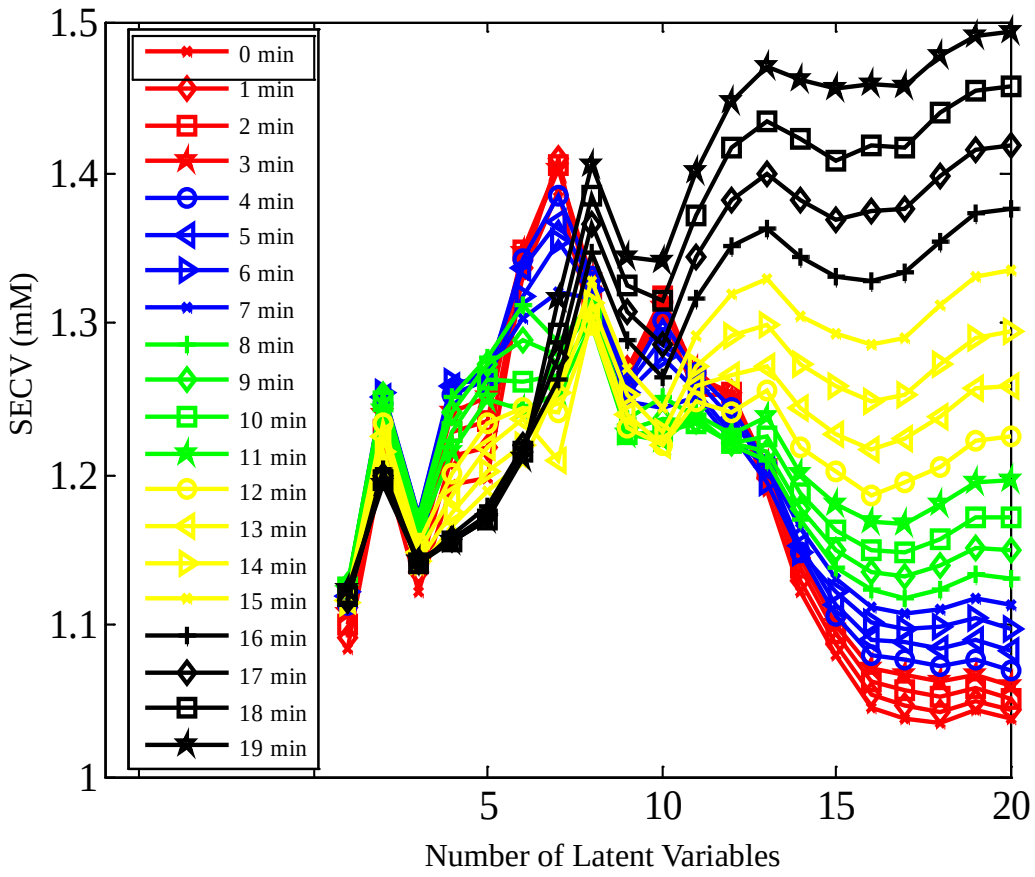


Figure 7.81. Plot of SECV vs. latent variables for the calibration data for rat 169. These results suggest that no lag time is needed.

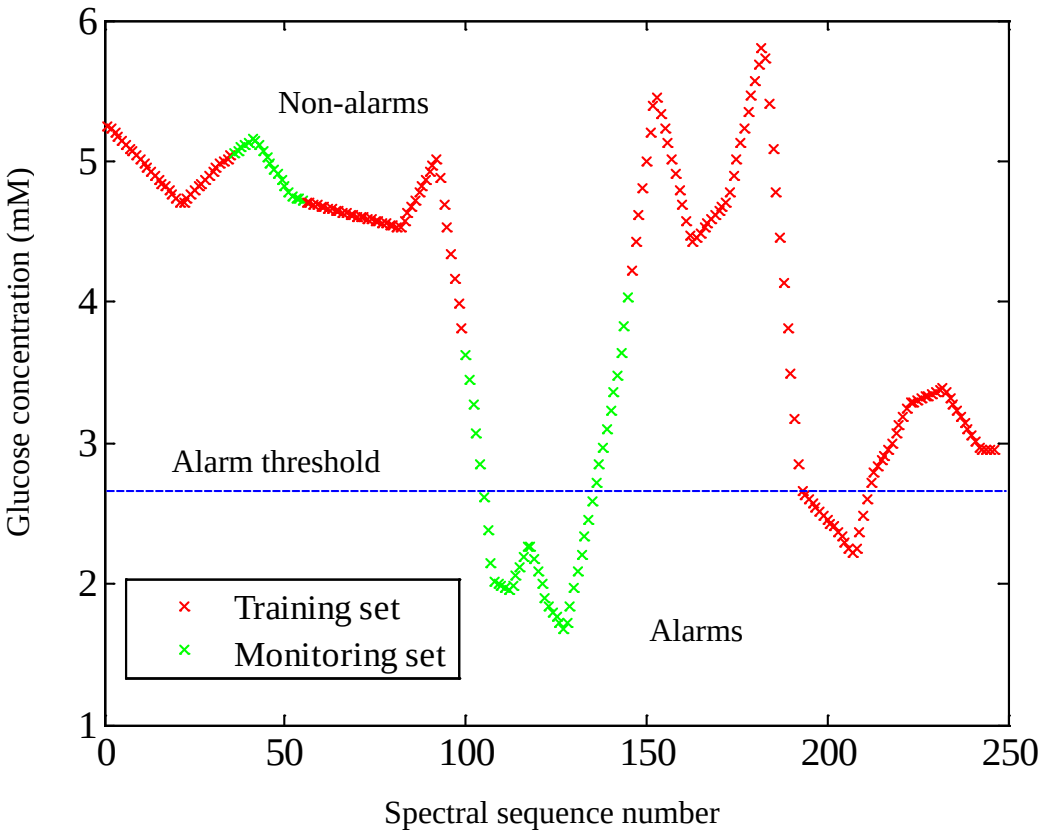


Figure 7.82. Glucose concentration profile for the calibration day for rat 169. The concentration profile was partitioned into a training set which contained 180 spectra and a monitoring set which had 66 spectra. A non-contiguous monitoring set was used to balance the number of alarm and non-alarm patterns and to have a more standard reference concentration. The horizontal line specifies the alarm threshold of 3.0 mM.

Table 7.7. Top four wavenumber-latent variable combinations based on minimum SECV value from the grid search analysis of training differential spectra

Wavenumber (cm^{-1})	Latent variables	SECV (mM)
4900-4250	16	0.51_1
4900-4200	16	0.51_8
4900-4300	16	0.52_8
4850-4250	16	0.54_1

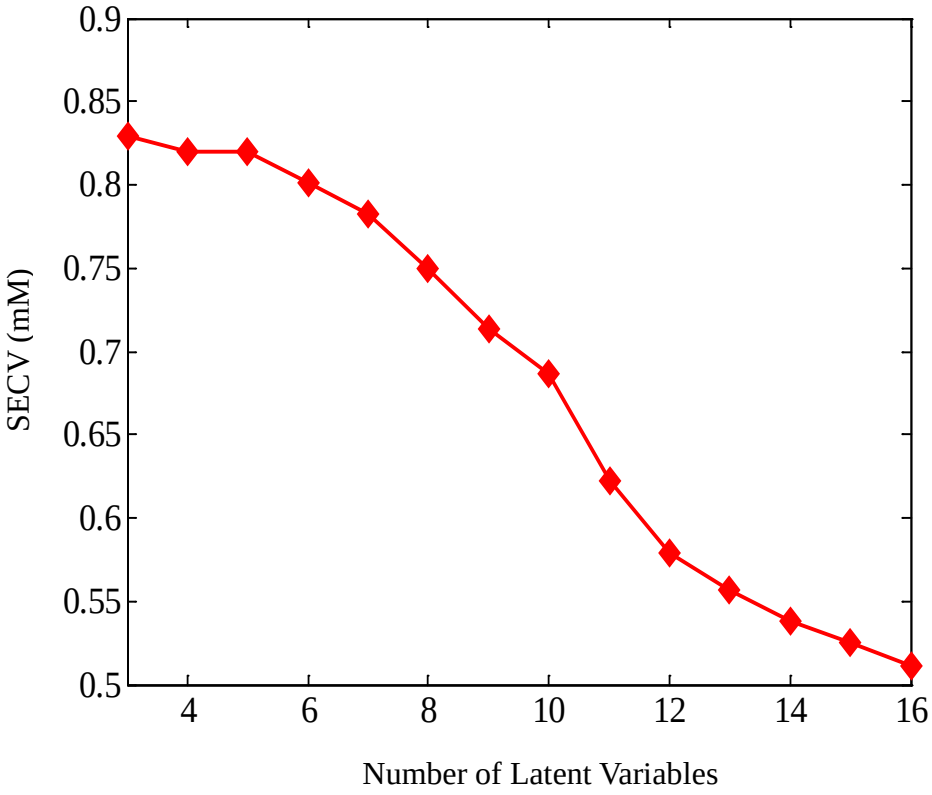


Figure 7.83. Plot of SECV vs. latent variable plot for the top wavenumber range resulting from the grid search. The minimum SECV was 1.03 mM at 16 PLS factors.

Table 7.8. Average percentage of missed and false alarms for the monitoring Set

Spectral range, cm^{-1}		PLS factors 12	13	14	15	16
4900-4250	AM (%)[a] ± SD	53.3 ± 40.0	75.2 ± 42.9	59.0 ± 27.5	40.0 ± 23.4	35.2 ± 23.0
	AF (%)[b] ± SD	9.4 ± 8.3	5.2 ± 9.0	7.3 ± 6.5	16.6 ± 7.9	12.5 ± 6.2
	DS (%)[b] ± SD	12.0 ± 1.6	14.8 ± 2.4	19.3 ± 1.6	19.6 ± 3.2	21.3 ± 1.6
4850-4250	AM (%)[a] ± SD	7.6 ± 5.9	3.81 ± 4.4	47.6 ± 26.6	28.6 ± 2.9	20.0±10.3
	AF (%)[b] ± SD	34.4 ± 3.1	33.3 ± 3.6	10.4 ± 6.5	14.6 ± 1.8	22.9 ± 9.0
	DS (%)[b] ± SD	9.4 ± 0.6	13.0 ± 1.9	11.1 ± 4.0	15.4 ± 1.3	14.8 ± 2.1.
4900-4200	AM (%)[a] ± SD	94.3 ± 9.90	94.3 ± 9.9	99.0 ± 1.6	100.0 ± 0.0	100.0 ± 0.0
	AF (%)[b] ± SD	5.2 ± 9.	5.2 ± 9.0	1.04 ± 1.8	0.0 ± 0.0	0.0 ± 0.0
	DS (%)[b] ± SD	12.3 ± 2.7	14.7 ± 0.9	18.8 ± 0.9	19.1 ± 3.4	22.6 ± 1.9
4900-4300	AM (%)[a] ± SD	70.4± 11.5	37.1± 35.8	53.3 ± 19.4	79.0 ± 8.7	100.0 ± 0.0
	AF (%)[b] ± SD	8.3 ± 4.8	15.6 ± 11.3	10.4 ± 4.8	8.3 ± 9.31	0.0 ± 0.0
	DS (%)[b] ± SD	10.8 ± 2.6	11.2 ± 2.4	10.4 ± 0.8	12.5 ± 0.8	16.5 ± 1.9

[a]AM (%) ± SD: Average percentage of missed alarms ± standard deviation.

[b]AF (%) ± SD: Average percentage of false alarms ± standard deviation.

[c]DS (%)± SD Average percentage of separated alarm patterns with three discriminants (DS) ± standard deviation.

The first prediction day was one day removed from the calibration. Each of the single-beam spectra collected was assigned a glucose concentration on the basis of the interpolated reference measurements and no lag was applied. Using the same protocol as discussed for rat 134, the data collected during the first 15 minutes (15 signal-averaged single-beam spectra) were removed from the analysis. The average RMS noise value for the first prediction day was 141 µAU. The plot of RMS noise values is shown in Figure 7.84. These noise values are higher than those obtained for the calibration day.

The glucose concentration profile for the first prediction day is shown in Figure 7.85. The first spectrum after removing the first 15 signal-averaged spectra had a concentration of 4.12 mM and was used as the reference spectrum. The alarm threshold was 3.00 mM and the critical concentration was -1.12 mM. All the data collected during the calibration day were put into a single block to compute the differential spectra. This yielded 30,129 total spectra. The PLS scores computed with the optimal wavenumber-latent variable combination were partitioned into 15,128 alarm and 14,961 non-alarm patterns. Three replicate classifiers were computed with the calibration PLS patterns and the first three discriminant functions from each of the replicates were used for the prediction. Across the three replicate classifiers, an average of 22 % of the alarm patterns in the calibration set were separated. This low degree of separation again reflects the large number of training patterns near the critical concentration (i.e., near the decision boundary of the discriminant).

When the alarm decision rule was applied, there were five missed alarms out of 87 alarm patterns and 62 false alarms out of 117 non-alarm patterns. The same analysis with a lag time of 10 minutes produced the same classification results. The discriminant scores plots for the individual classifiers and the committee result are shown in Figures 7.86 A and B, respectively. The classifier works well at the beginning of the data collection, correctly going into alarm near spectrum 60. The glucose transient out of the alarm condition near spectrum 100 is missed, however, as the classifier remains in the alarm state from that point forward.

The regression coefficients resulting from the fit to the spectra of the pure components for each of the spectra collected on the first prediction day of rat 169 are

shown in Figure 7.87. The traces with respect to time are reasonably stable. Examination of panel B reveals a water thickness approaching 1.5 mm. This is most likely the reason for the increased noise level for this prediction day.

The first four prediction PLS scores are shown in Figure 7.88. Most of the scores except for score 4 are within the space defined by the calibration scores. Still, each of the traces shows significant fluctuations with time. This might be the reason for the large number of false alarms observed from the classifier.

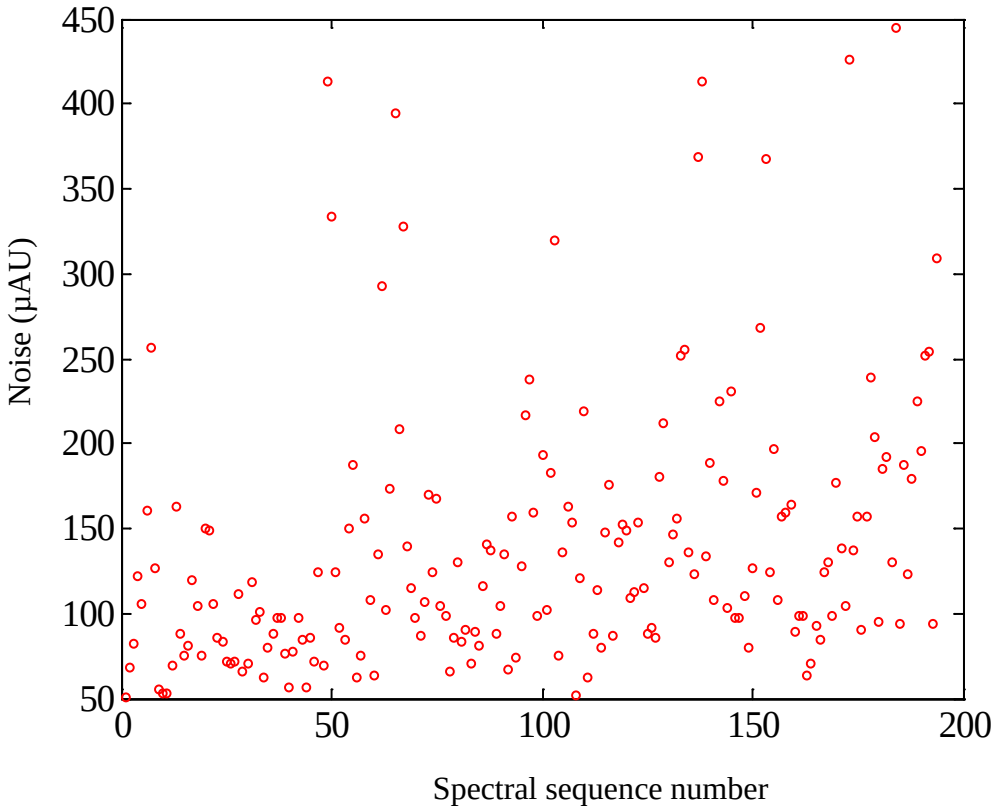


Figure 7.84. Plot of RMS noise values (in µAU) for consecutive pairs of signal-averaged spectra collected on the first prediction day of rat 169. The 4500-4300 cm^{-1} region was used for the noise calculation. Noise values were computed by fitting the 100% line in the given wavenumber region to a third-order polynomial and calculating the noise about the fit. These noise values are higher than those observed for the calibration data.

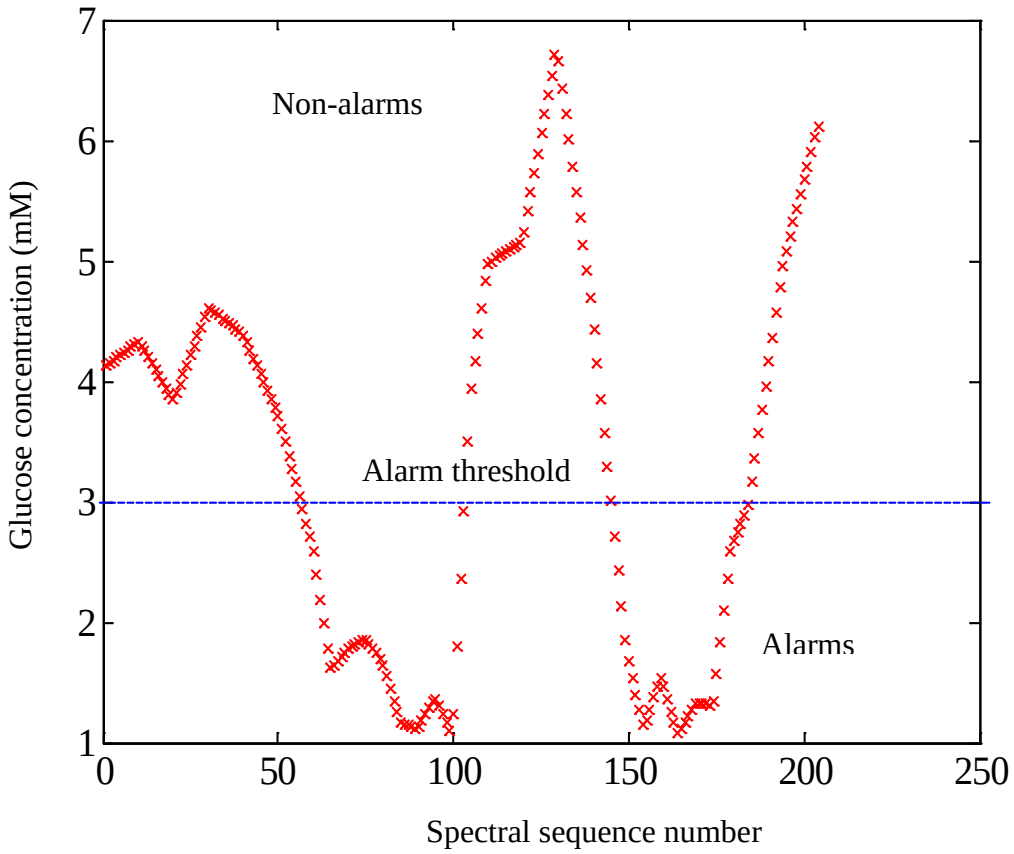
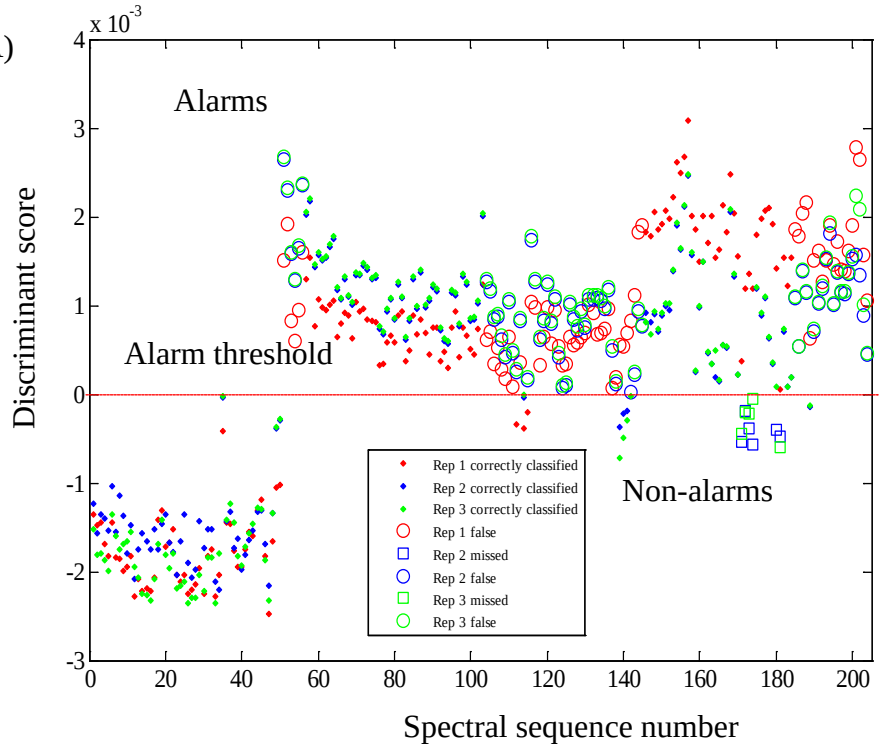


Figure 7.85. Glucose concentration profile for the first prediction day of rat 169. The concentration profile contained 204 total spectra, 117 non-alarm concentrations and 87 alarm concentrations. The horizontal line denotes the alarm threshold of 3.0 mM.

(A)


x 10^-3
Alarms
Alarm threshold
Non-alarms
Discriminant score
Spectral sequence number
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Rep 1 false
Rep 2 missed
Rep 2 false
Rep 3 missed
Rep 3 false

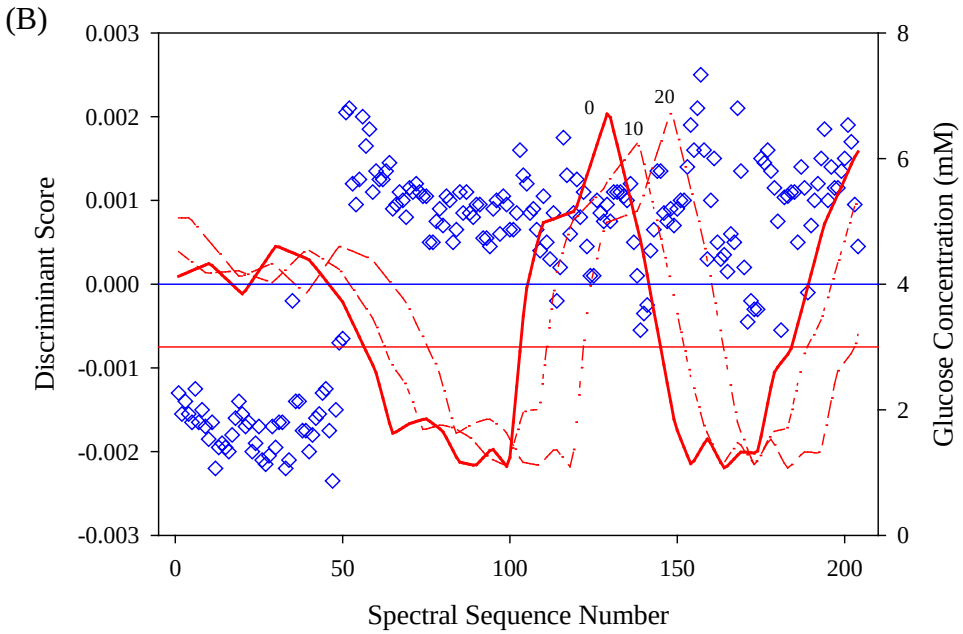


Figure 7.86. A. Discriminant score plot for the first prediction day of rat 169 with three replicate classifiers. The first spectrum was used as the reference. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). Glucose concentration profiles are shown for lag times of 0, 10 and 20 minutes.

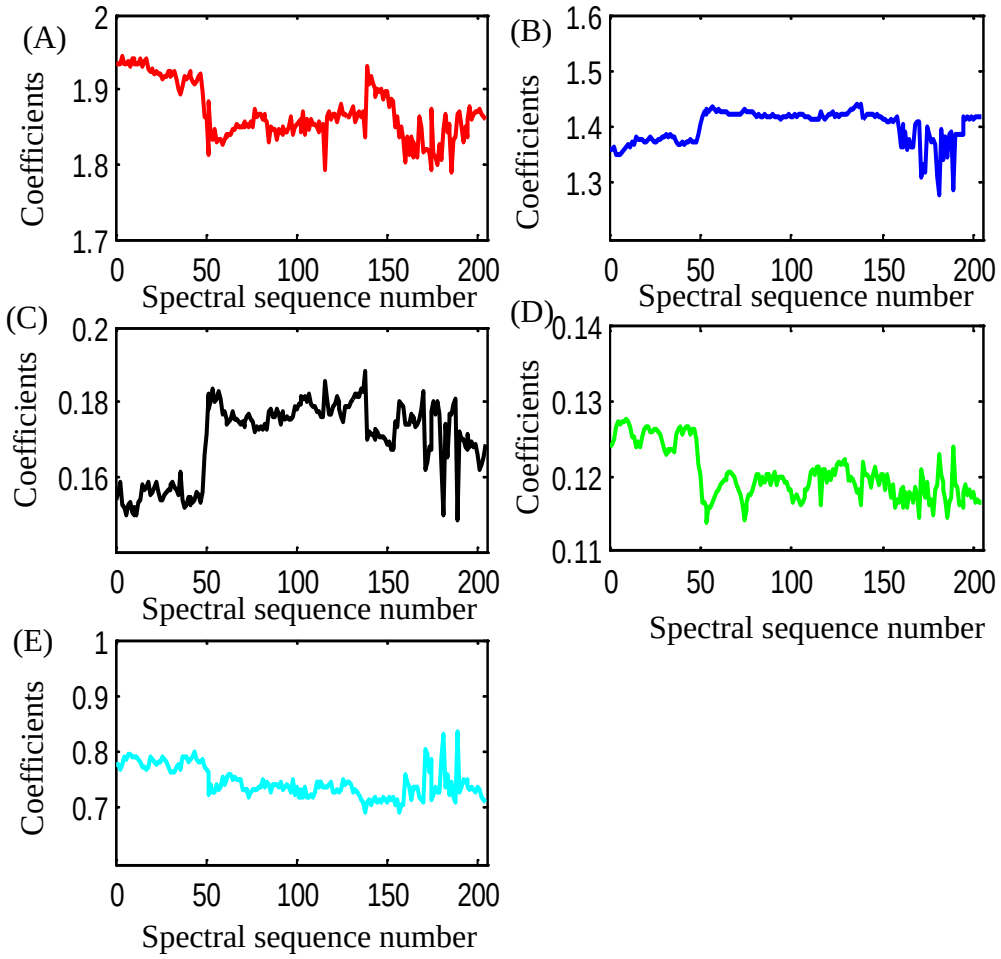


Figure 7.87. Regression coefficients computed for each of the spectra collected for the first prediction day of rat 169. Panels A-E correspond to regression coefficients for collagen, water, keratin, fat, and an intercept term. The regression coefficients represent the thickness of each component in the skin tissue matrix. The water coefficient in panel B is thicker than ideal for obtaining spectra with low noise levels.

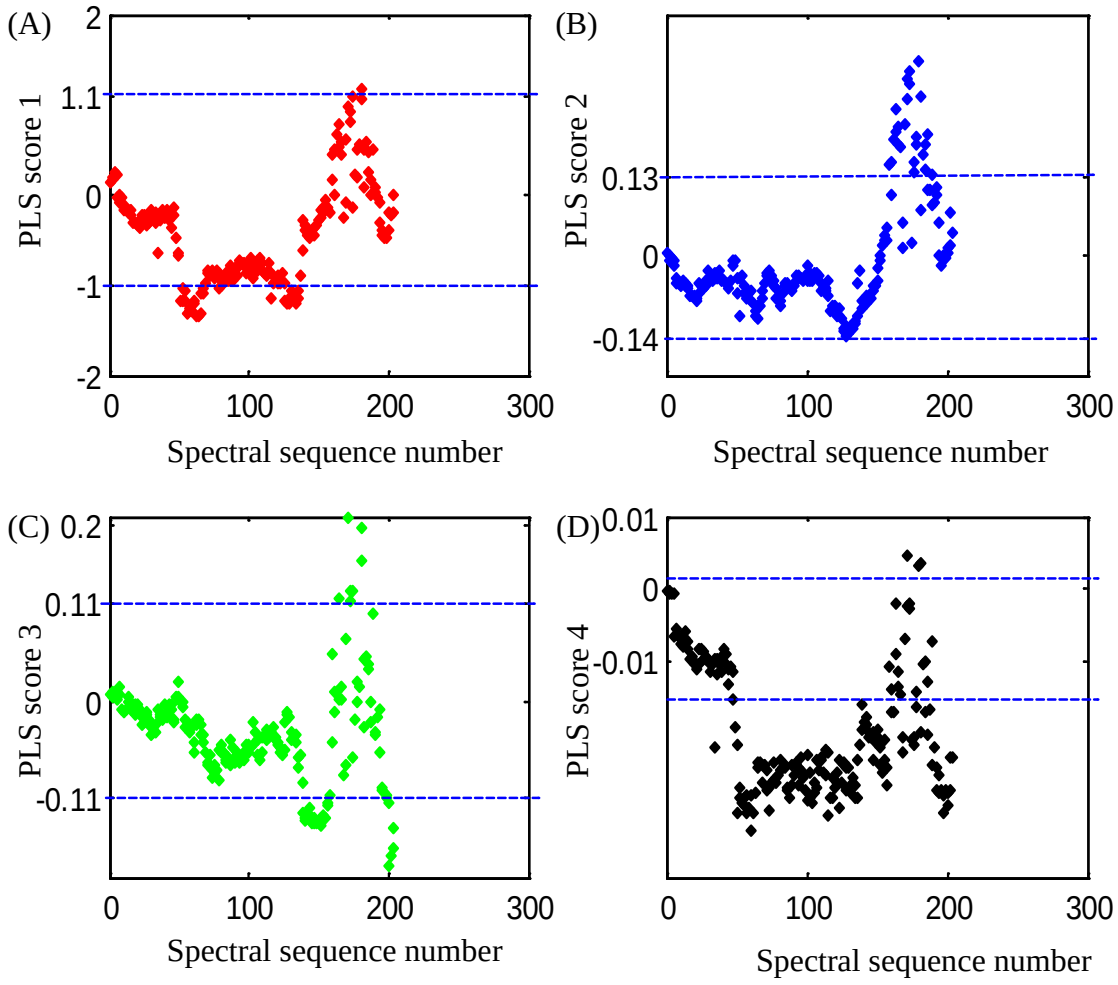


Figure 7.88. Plots of PLS scores vs. spectral sequence number for the prediction differential spectra collected on the first prediction day for rat 169. The 4850-4250 cm^{-1} wavenumber range was used for the PLS calculation. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as dashed lines. Panels A, B, C, and D correspond to latent variables 1, 2, 3, and 4, respectively.

Similar to the procedure used with rat 134, a reference updating strategy was explored to check if this step would improve the classification performance. All other parameters used were as described above. A strategy of updating every hour was applied when it was possible to do so. The first spectrum of the data set was used as a reference for the first two hours of data collection. It was not possible to update the reference spectrum after one hour since the glucose concentration was in the alarm state at that time. There were no missed alarms out of 47 alarm patterns and 16 false alarms out of 73 non-alarm patterns.

The reference spectrum was then updated from the 121st spectrum, two hours after the starting point of the data collection. The reference concentration for the corresponding spectrum was 5.40 mM. The concentration profile and the corresponding discriminant score plot are shown in Figures 7.89 and 7.90, respectively. There were 31 missed alarms out of 40 alarm patterns and 15 false alarms out of 40 non-alarm patterns. After updating, the classification performance tracks the concentration profile until approximately spectrum 170. The missed and false alarms come primarily from the last segment of the experiment.

The overall committee result obtained with updating is provided in Figure 7.91. For a lag time of zero, combining the two sets of results obtained with updating produced 31 missed alarms and 31 false alarms in total. The reference updating strategy improved the data prediction significantly for the non- alarm patterns (i.e., the number of false alarms decreased by half from 62 to 31), but the number of missed alarms increased from 5 to 31. Applying a lag time of 10 minutes to the results obtained with reference updating, there were 25 missed alarms and 34 false alarms. The overall conclusion from

this prediction day is that the noise levels are too high for the alarm algorithm to work well.

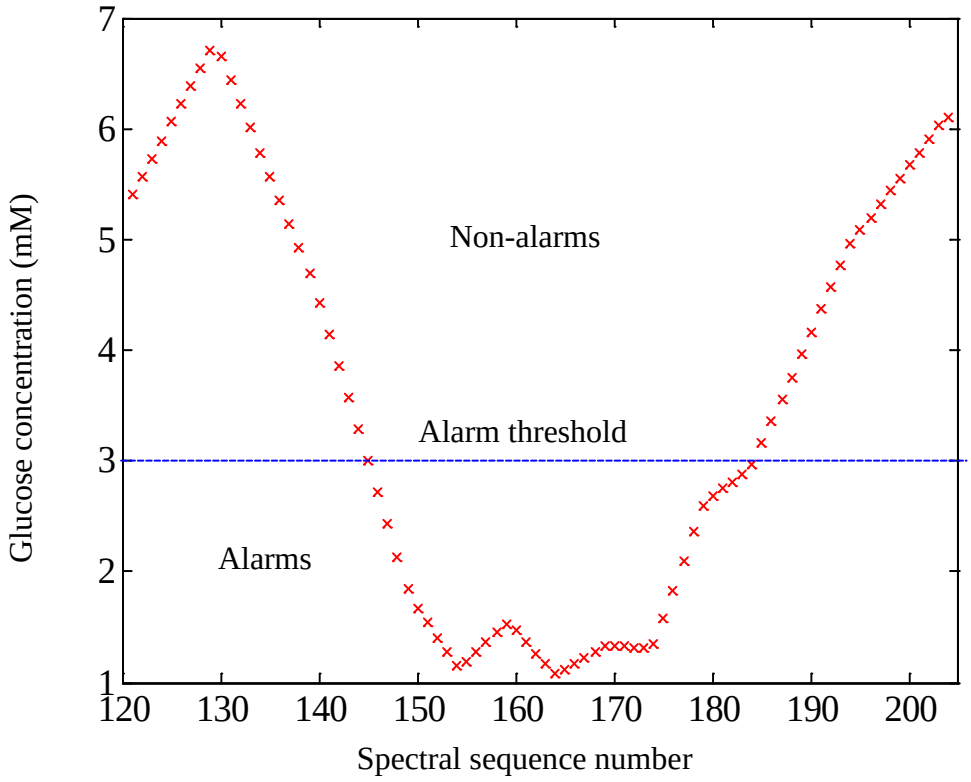
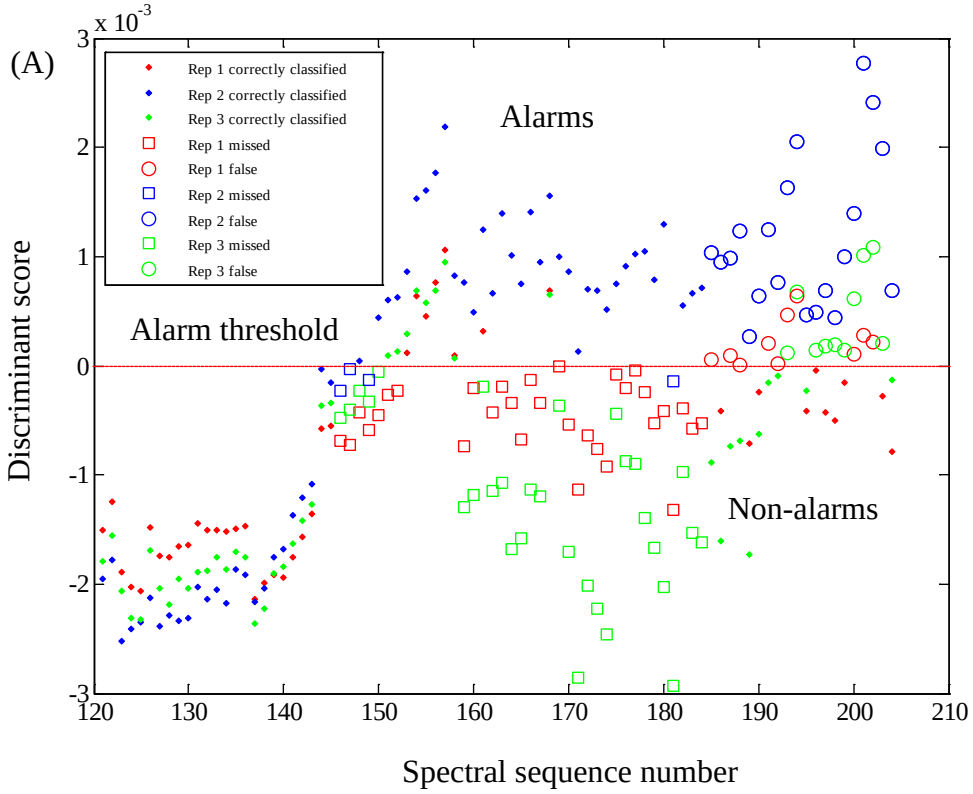


Figure 7.89. Glucose concentration profile for 121-204 minutes of the first prediction day for rat 169. The concentration profile contained 84 total spectra, 44 non-alarm concentrations and 40 alarm concentrations. The alarm threshold of 3.0 mM is denoted by the horizontal line.


(A)
x 10-3
3
2
1
0
-1
-2
-3
Discriminant score
120
130
140
150
160
170
180
190
200
210
Spectral sequence number
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Rep 1 missed
Rep 1 false
Rep 2 missed
Rep 2 false
Rep 3 missed
Rep 3 false
Alarms
Alarm threshold
Non-alarms

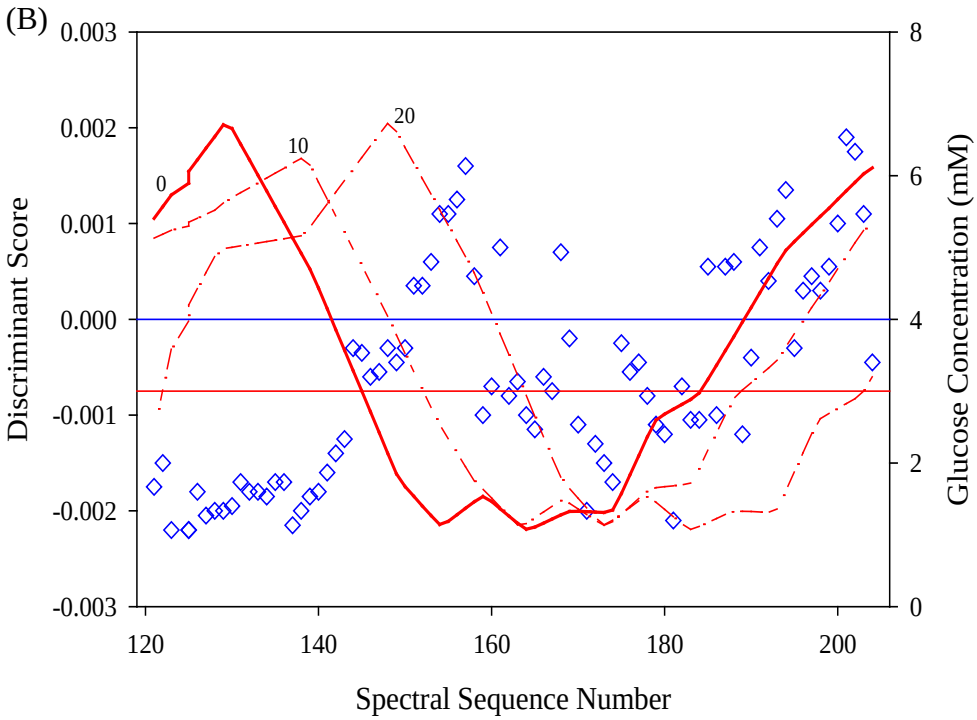


Figure 7.90. A. Discriminant score plot for 121-204 minutes of the first prediction day of rat 169 with three replicate classifiers. The spectrum at 121 minutes was used as the reference. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). Glucose concentration profiles corresponding to lag times of 0, 10 and 20 minutes are shown in the figure.

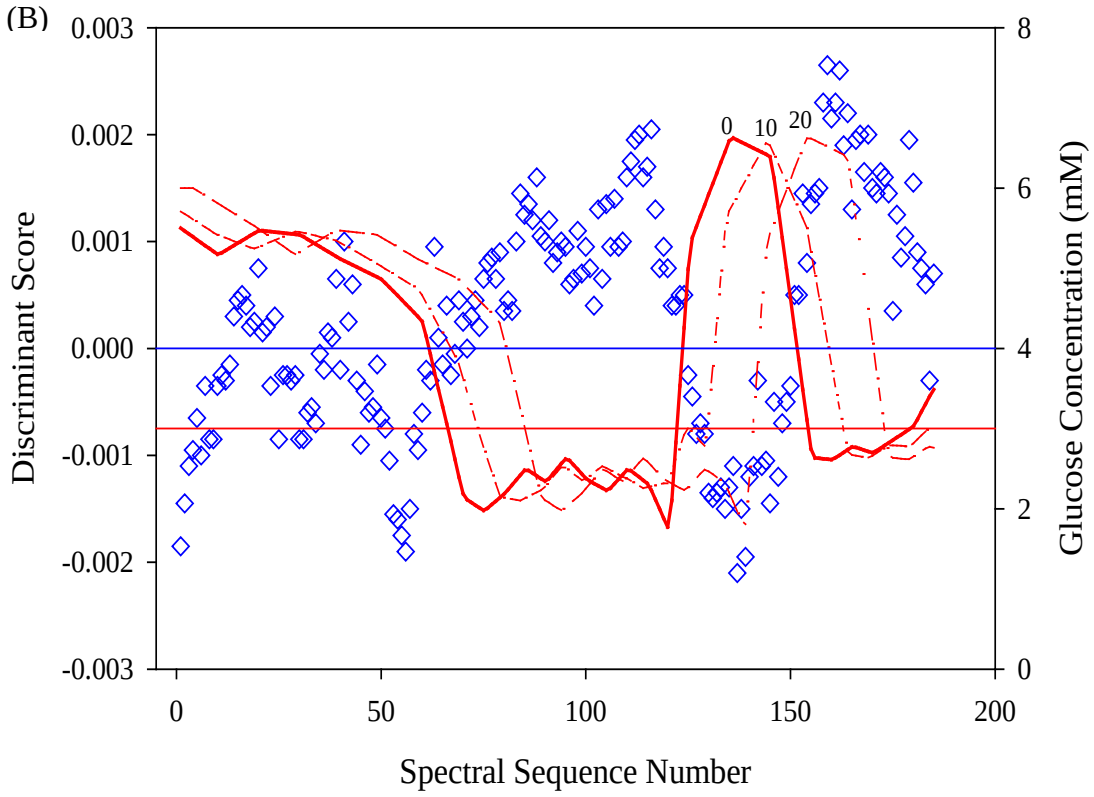


Figure 7.91. Committee result for the first prediction day of rat 169 after updating the reference at spectrum 121. The classification performance tracks the concentration profile until approximately spectrum 170. Glucose concentration profiles corresponding to lag times of 0, 10 and 20 minutes are shown in the figure.

Data for the second prediction set were collected two days after the calibration day. Each of the single-beam spectra was assigned to an interpolated glucose concentration and no lag correction was applied. Using the same procedures as described previously, the average RMS noise value for the second prediction day was 102 μAU. A plot of the computed noise values is provided in Figure 7.92. While lower than the first prediction day, this value still represents a high noise level.

The first spectrum after removing the first 15 signal-averaged spectra had a concentration of 5.50 mM and was used as the reference spectrum. The alarm threshold was 3.00 mM and the critical concentration was -2.50 mM. The PLS scores computed with the optimal wavenumber-latent variable combination were partitioned using the critical concentration into 4480 alarm and 25,649 non-alarm patterns. Three replicate classifiers were computed with the calibration PLS patterns and the first three discriminants from each of the replicates were used for the prediction. Across the three replicate classifiers, an average of 5 % of the alarm patterns in the training set were separated. This value is extremely low and indicates extreme difficulty in establishing a separating boundary between the alarm and non-alarm data classes.

When the alarm decision rule was applied, there were four missed alarms out of 80 alarm patterns and 30 false alarms out of 100 non-alarm patterns. This corresponds to a successful detection of 95 % of the alarm events, but with a problematic 30 % occurrence of false alarms. The corresponding glucose concentration profile and the discriminant scores plot are shown in Figures 7.93 and 7.94, respectively. Figure 7.94 shows the results for the individual replicate classifiers and the committee result. The majority of the false alarms occur in the first 40 minutes of data collection. The

classification response then tracks the concentration profile well for the remainder of the data collection. No reference updating was attempted for the second prediction day as the results did not exhibit a decline in performance over time. Use of a lag time of 10 minutes did not help to improve the data prediction, giving four missed alarms and 36 false alarms.

The regression coefficients corresponding to the fit of the pure-component spectra to each of the spectra collected on the second prediction day are shown in Figure 7.95. The regression coefficients are quite stable over the course of the experiment. The water thickness is again quite large (> 1.5 mm) and is likely responsible for the higher than desired RMS noise levels.

The first four prediction PLS scores for the second prediction day at the wavenumber range of 4850-4250 cm^{-1} are plotted against the spectral sequence numbers in Figure 7.96. The first two prediction PLS scores are within the space defined by the calibration PLS scores. The PLS scores do not show much fluctuation with time. This may be a contributing reason for the overall better data prediction on the second prediction day.

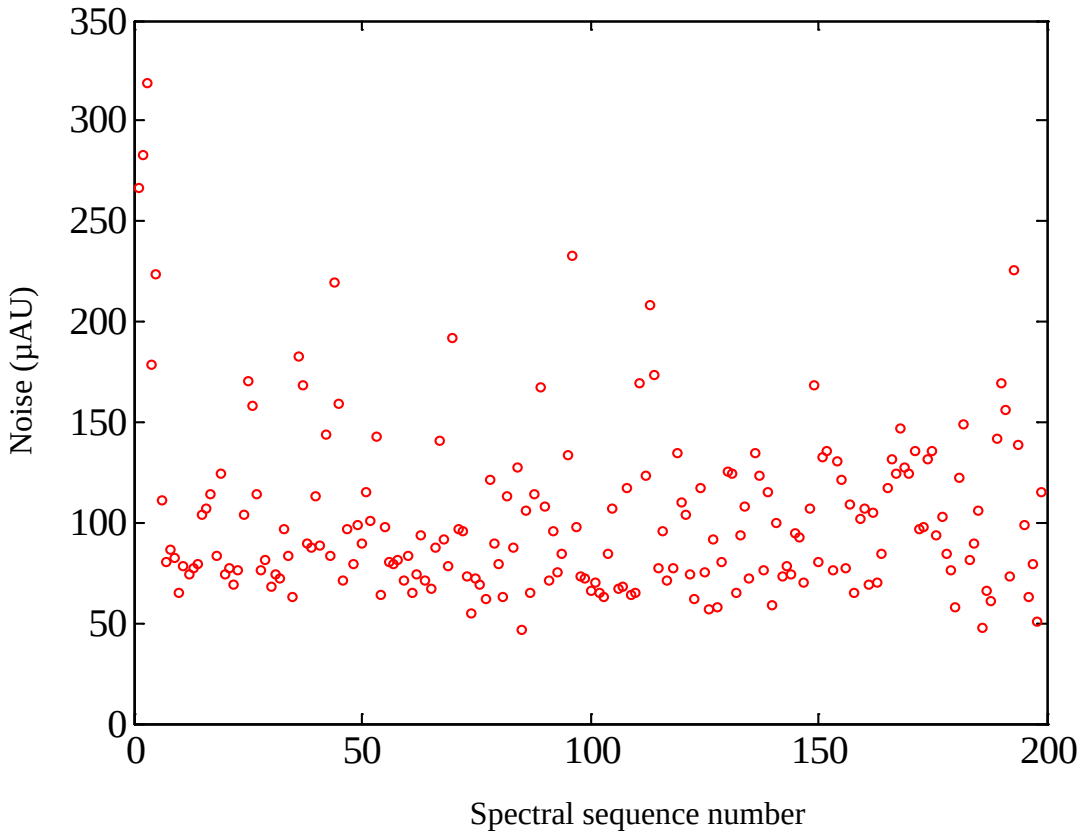


Figure 7.92. Plot of RMS noise values (in µAU) computed for each signal-averaged spectrum collected on the second prediction day for rat 169. The 4500-4300 cm^{-1} region was used. Noise values were computed by fitting the 100% line in the given wavenumber region to a third-order polynomial and computing the noise about the fit.

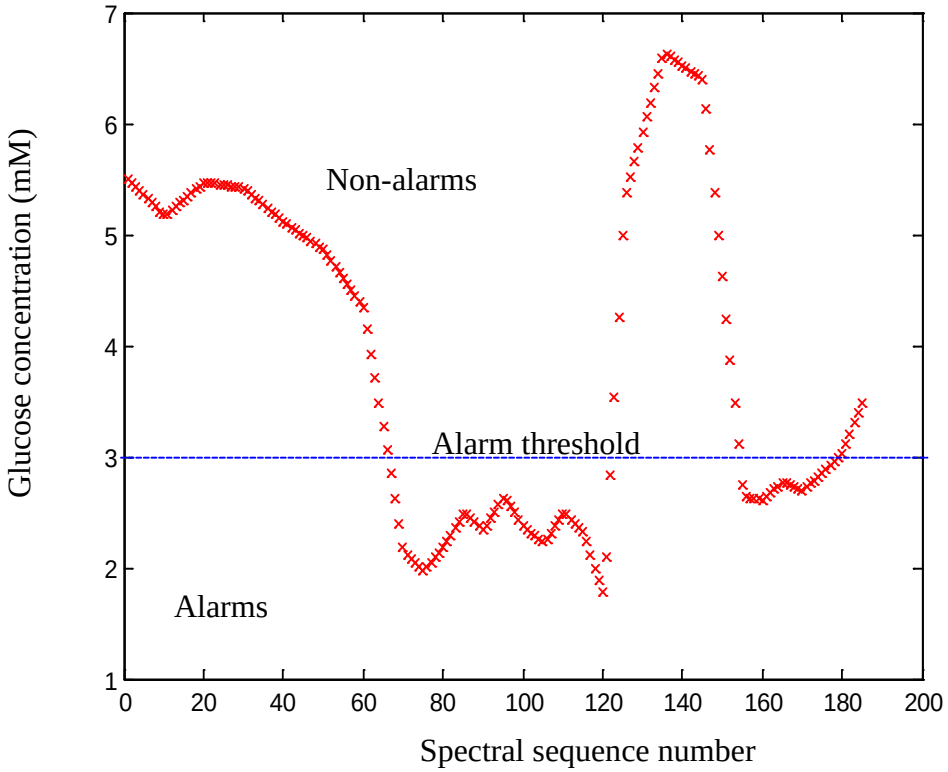
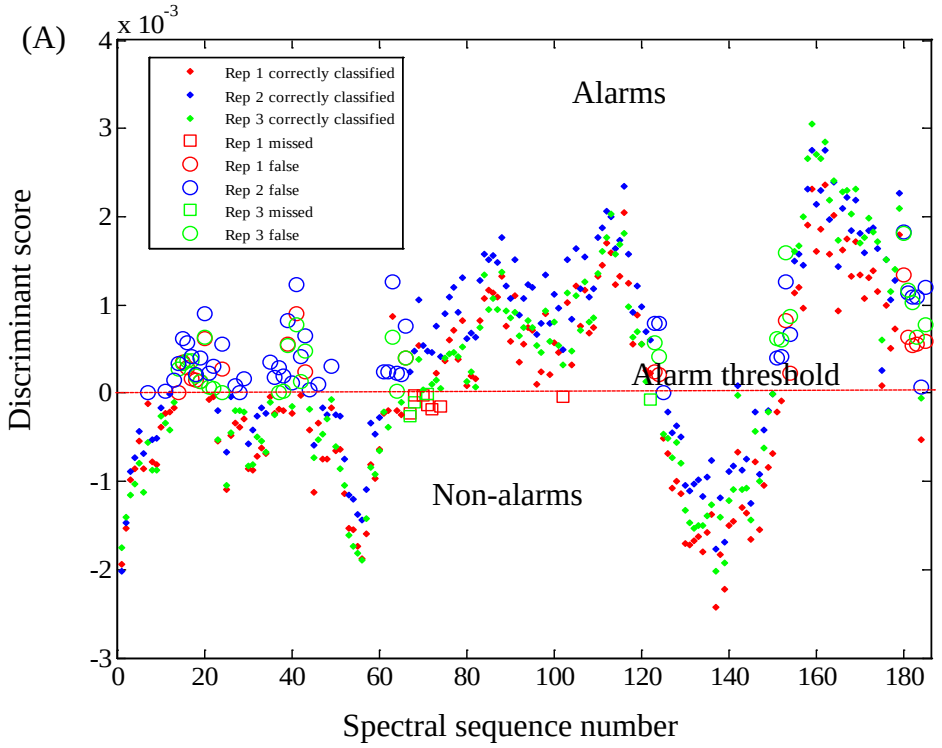


Figure 7.93. Glucose concentration profile for the second prediction day of rat 169. There were 80 alarm and 100 non-alarm spectra collected on the basis of an alarm threshold of 3.0 mM (horizontal line).


(A)
x 10-3
Discriminant score
Spectral sequence number
Rep 1 correctly classified
Rep 2 correctly classified
Rep 3 correctly classified
Rep 1 missed
Rep 1 false
Rep 2 false
Rep 3 missed
Rep 3 false
Alarms
Alarm threshold
Non-alarms
4
3
2
1
0
-1
-2
-3
0
20
40
60
80
100
120
140
160
180

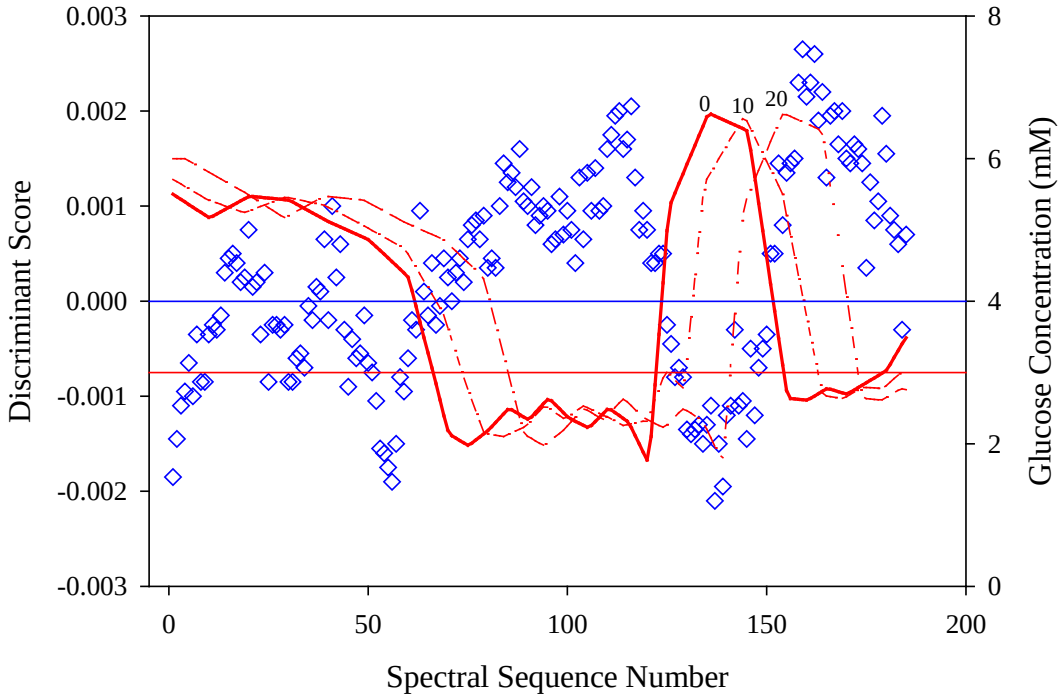


Figure 7.94. A. Discriminant score plot for the second prediction day of rat 169 with three replicate classifiers. The first spectrum was used as the reference. B. The discriminant scores corresponding to the committee result are shown (left y-axis) with the lag-corrected reference glucose concentrations superimposed (right y-axis). The horizontal lines correspond to the alarm/non-alarm thresholds for the discriminant scores (0.0) and reference concentrations (3.0 mM). Glucose concentration profiles are shown for lag times of 0, 10 and 20 minutes.

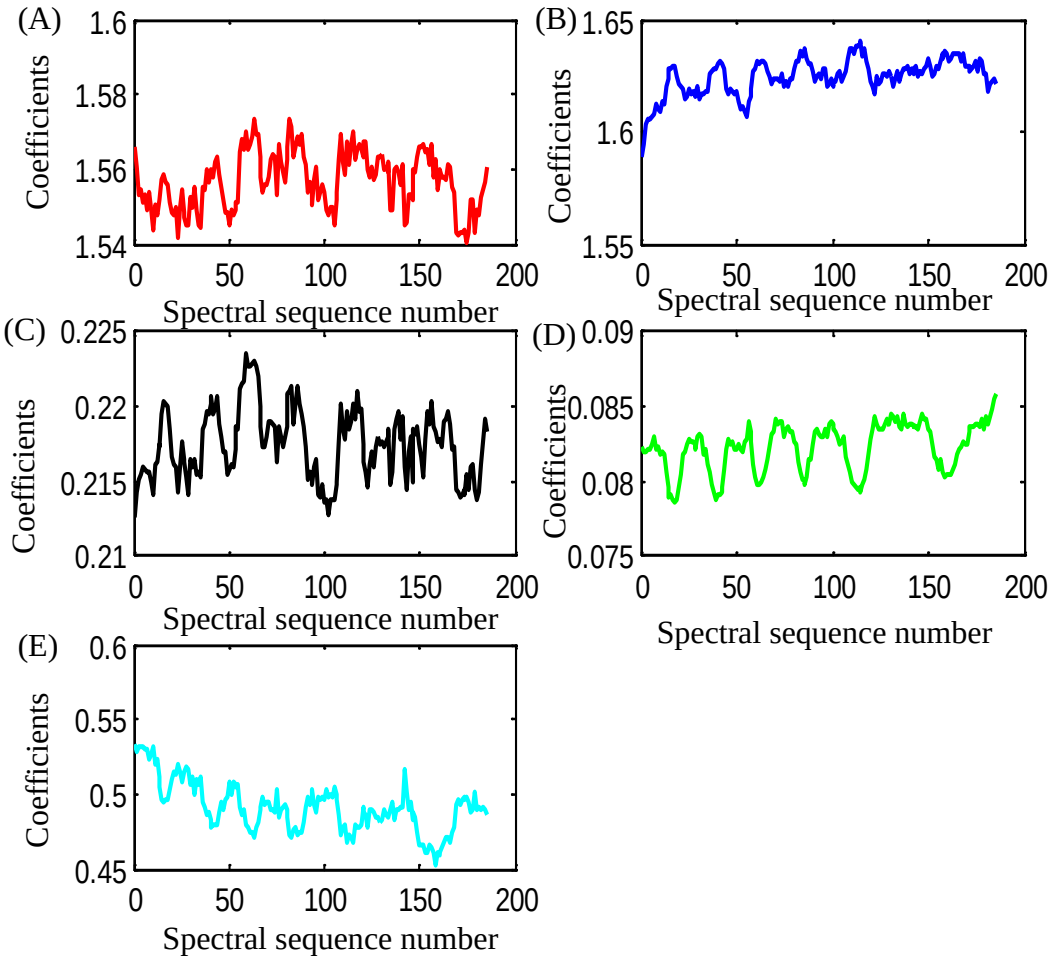


Figure 7.95. Regression coefficients computed for each of the spectra collected for the second prediction day of rat 169. Panels A-E correspond to regression coefficients for collagen, water, keratin, fat, and an intercept term. The regression coefficients represent the thickness of each component in the skin tissue matrix. The water coefficient in panel B is thicker than ideal for obtaining spectra with low noise levels.

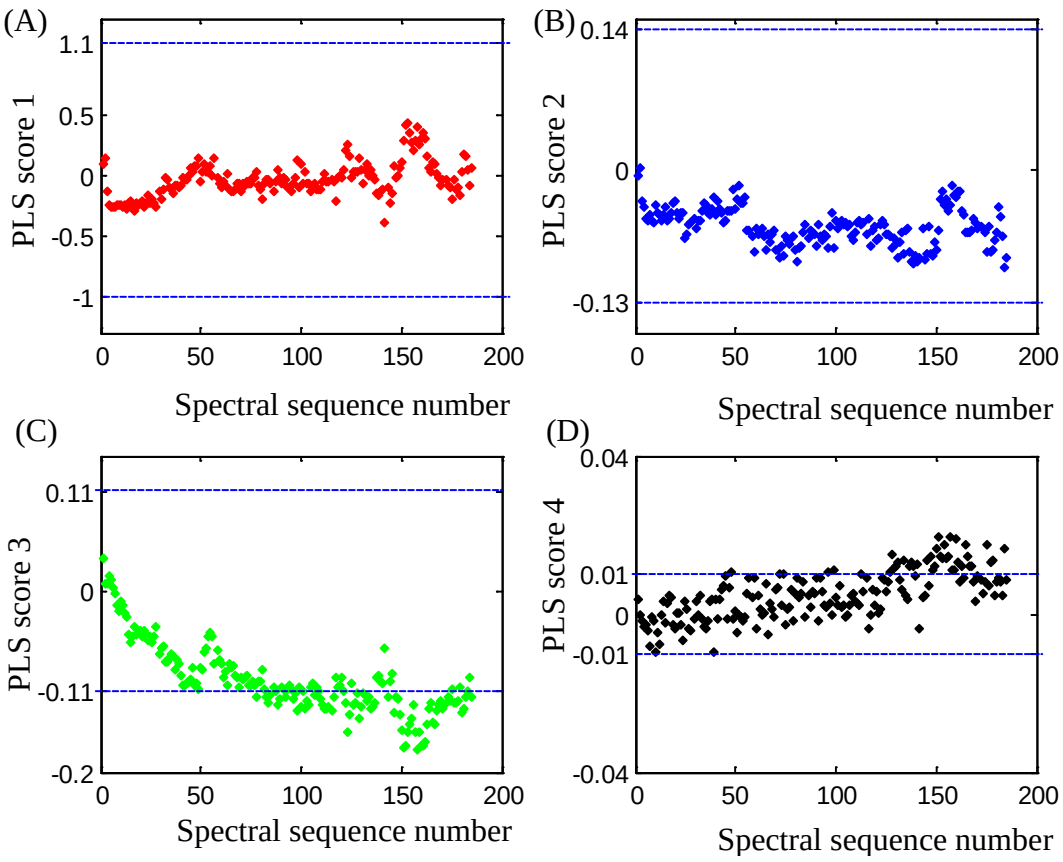
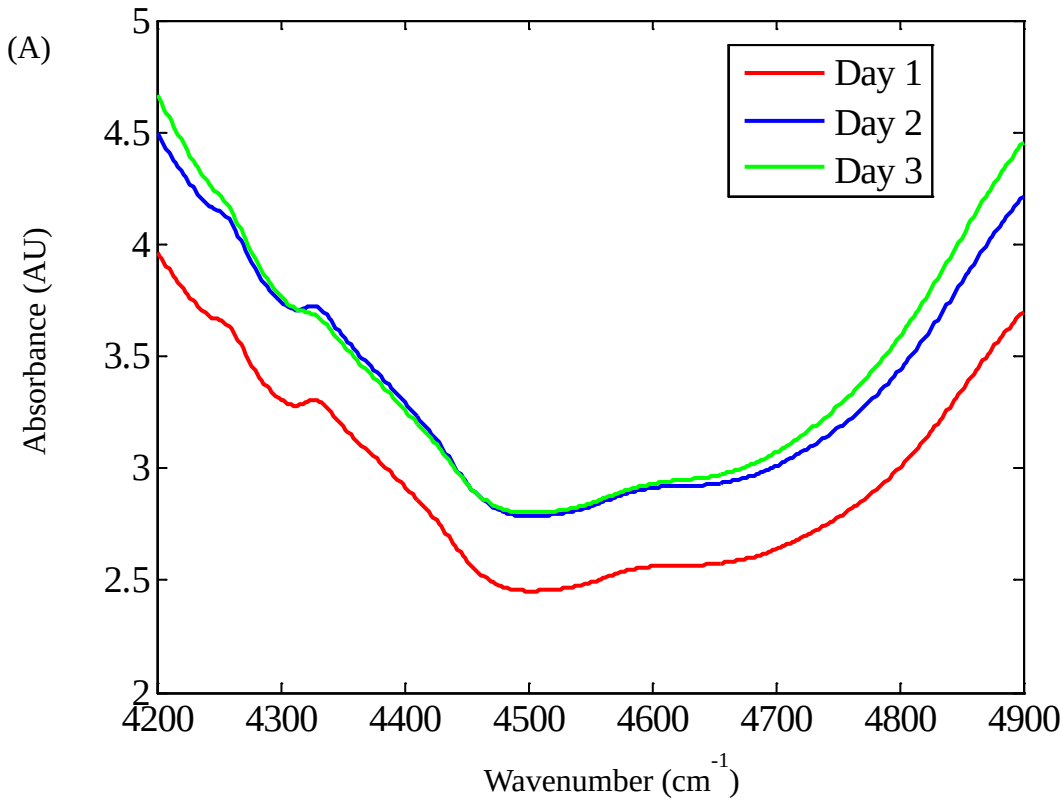


Figure 7.96. Plots of PLS scores vs. spectral sequence number for the prediction differential spectra collected on the second prediction day for rat 169. The 4850-4250 cm^{-1} wavenumber range was used for the PLS calculation. The maximum and minimum PLS scores computed with the calibration differential spectra are plotted as dashed lines. Panels A, B, C, and D correspond to latent variables 1, 2, 3, and 4, respectively.

The mean absorbance spectra collected on each day of data collection are shown in Figure 7.97 A and the corresponding mean differential spectra are plotted in Figure 7.97 B. It is clear that there are significant variations in the absorbance across the three days of data collection.

Figure 7.98 is a score plot derived from performing PCA on the combined data from rat 169. The spectra were mean-centered using the overall mean of the spectra and the region of 4900 to 4200 cm^{-1} was submitted to PCA. Clusters can clearly be observed corresponding to the day of data collection. Figure 7.99 is a corresponding score plot based on differential spectra. As observed previously with rat 134, calculation of the differential spectra helps to remove day-to-day variation, although data groupings corresponding to the day of data collection can still be seen.



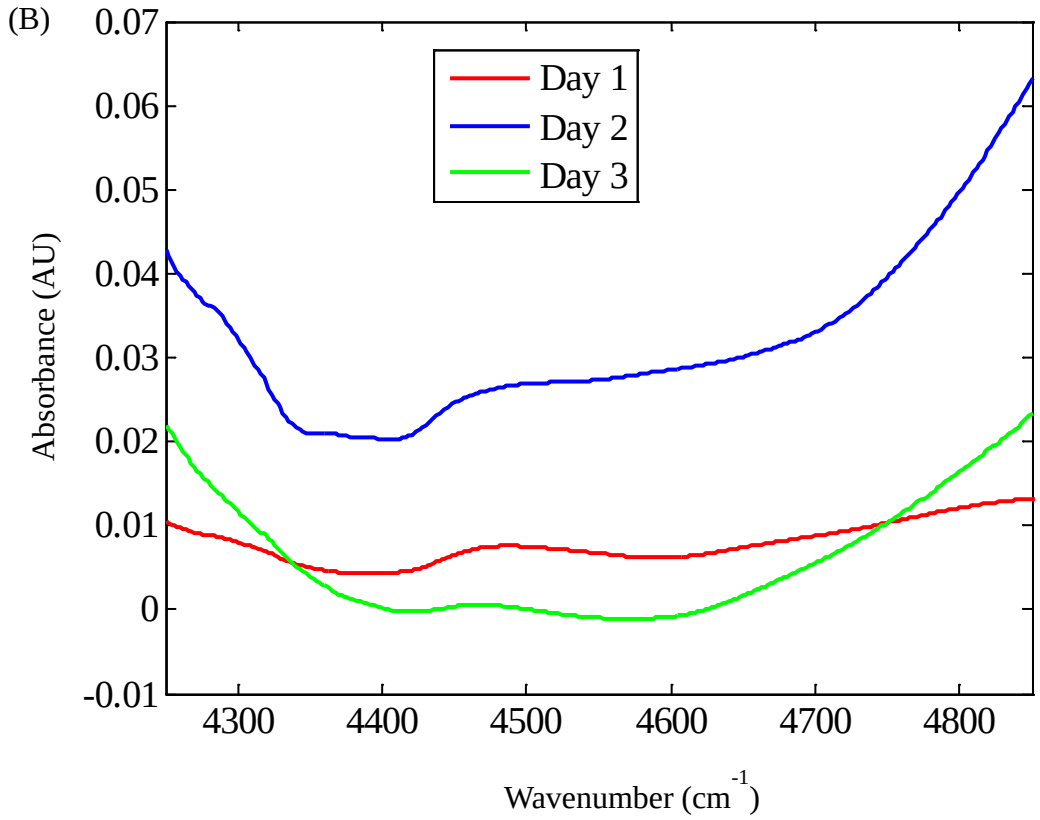


Figure 7.97. A. Mean absorbance spectra for each day of data collection for rat 169. B. Corresponding mean differential spectra.

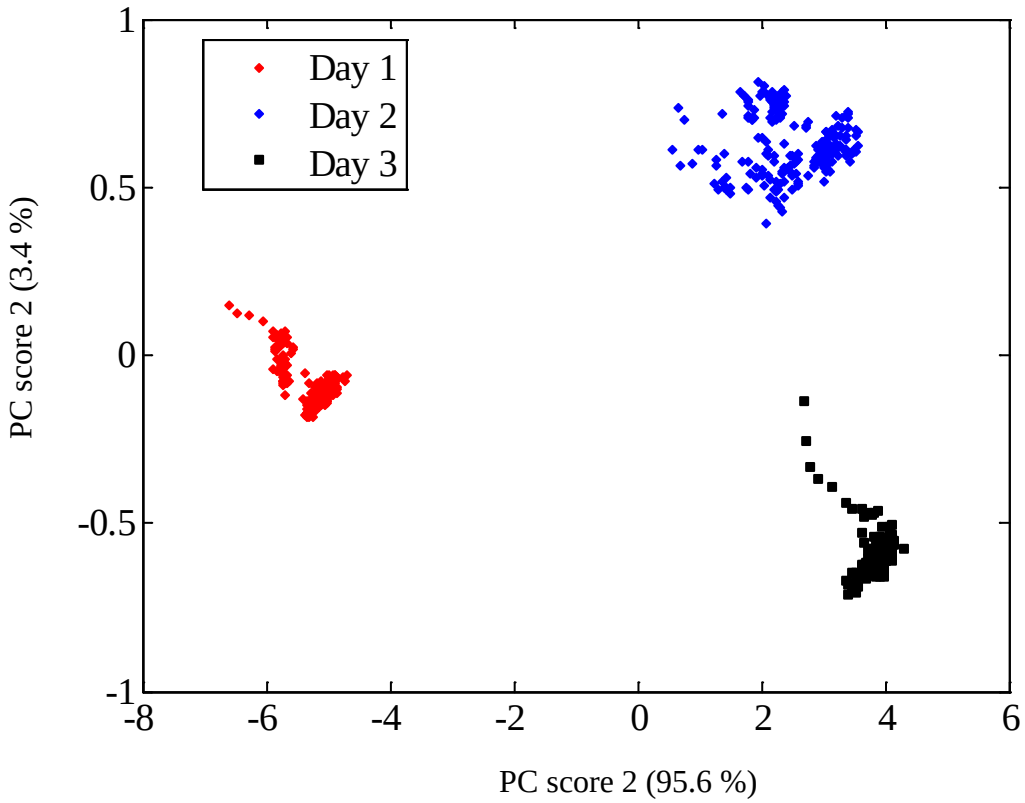


Figure 7.98. Principal component score plot for the three days of data collection for rat 169 using absorbance spectra relative to an air background. The spectra were mean-centered using the overall mean of the spectra. The wavenumber region of 4900 to 4200 cm^{-1} was employed.

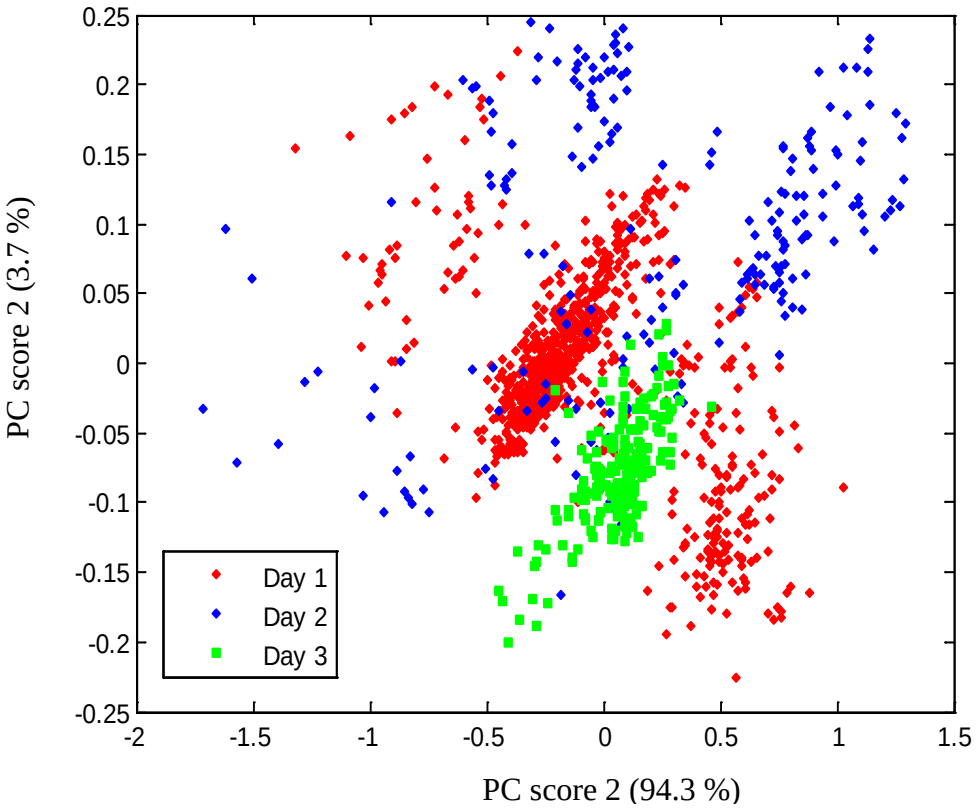


Figure 7.99. Principal component score plot for the three days of data collection for rat 169. The differential spectra for each day were used in computing the principal component scores. Spectra were mean-centered and the region of 4900 to 4200 cm^{-1} was employed. Less day-to-day variation is observed than in Figure 7.98, although data groupings are still observed corresponding to the day of data collection.

Conclusions

In this chapter, the nocturnal hypoglycemic alarm algorithm presented in Chapter 3 was tested with a series of *in vivo* measurements performed with a rat animal model. By use of glucose clamp procedures in which glucose or insulin were administered to the

animal, data were collected that simulated the glucose excursions that might occur in a human with diabetes during sleep. From these studies, the potential of the alarm algorithm was further established.

The alarm algorithm was first tested with data collected in non-survival single-day experiments in which the rat was anesthetized during the data collection. Spectra were collected continuously over several hours. The results presented here focused on measurements with two different rats (106 and 114). In each case, the data collected on a single day were used to calibrate the alarm algorithm and then to test the implementation of the alarm. While still collected on the same day, the prediction data in each case corresponded to a time period outside of the time span of the calibration.

For both rats 106 and 114, the nocturnal alarm algorithm performed well when applied to the prediction data. On average, 87.5 % of alarm events were correctly detected, and the occurrence of false alarms was at 7.2%. There was clear evidence of spectral drift during the time course of the experiment, but the alarm algorithm was able to overcome this variation without significant negative effects.

Results were also presented for two rats (134 and 169) that were used for experiments conducted over multiple days. The multiple-day data collection is much more complex than the single-day experiment. The rat is awake during the data collection which can lead to significant additional spectral variation arising from movement artifacts. Also, there are potential variations in skin composition with time. While there was an attempt to make the measurement each day at the same location on the rat skin, changes in the pressure of the interface from day to day can induce effective changes in the skin thickness present within the optical path. Also, there was an apparent change in

skin transmission over time due to a cumulative effect of the same tissue section being repeatedly placed within the interface. These effects are reflected in the regression coefficients obtained by fitting the collected spectra to pure-component spectra of the principal tissue components.

For rat 134, data were collected in five one-day sessions that spanned a total of 21 days. The calibration database was assembled from the data collected on the first day, while the remaining four days were used in testing the implementation of the alarm algorithm. For the first three prediction days, 77.1 % of alarm events were detected successfully, and the average occurrence of false alarms was 3.1 %. Performance declined on the fourth day, however. By increasing the pattern dimensionality by one PLS factor and updating the reference spectrum twice during the experiment, 86.8 % of alarm events were detected with a false alarm rate of 10.5 %. Overall, the classification performance with rat 134 was deemed promising, particularly given the many sources of variance encountered in the multiple-day experiment.

The results obtained with rat 134 also illustrated the effects of imprecision in the estimate of the lag time between arterial and tissue glucose concentrations. The lag time estimated for the calibration day was 29 minutes, a value higher than normal on the basis of previous results. For the four prediction days, the results were inconsistent with respect to whether the estimated lag time was correct. The classification results obtained for the first prediction day suggested a slight change in lag time might be appropriate, but the value of 29 minutes was not clearly wrong. Prediction days 2 and 3 had no alarm events, and thus no change in lag time would have affected the classifications. For the fourth prediction day, however, the results indicated a lag time of approximately 15 minutes was

probably more appropriate than 29 minutes. Overall, these results illustrate the importance of the lag time estimate and the inadequacies of the current methodology in identifying the appropriate lag time for a given day.

Rat 169 was employed over three consecutive days of data collection, with the first day being used for calibration and the remaining two days serving as prediction data for testing the implementation of the alarm. Noise levels for rat 169 were higher than those computed for the other rats. This increase in noise arose from a greater thickness of water in the optical path (~ 1.5 mm vs. < 1.0 mm for the other rats). Across the two prediction days, an average of 79.7 % of alarm events were detected, but the occurrence of false alarms averaged 28.2 %. These results reflect performing one reference update at the two-hour mark on the first prediction day. From the calibration data, no lag time was found to be significant, and none of the prediction results indicated a different lag time would have improved the classification performance.

Several overall conclusions can be drawn from the work presented in this chapter. First, classification performance is negatively impacted by increased spectral noise levels. Given the greater degree of spectral variance inherently encountered in the multiple-day experiments, it was especially problematic to have noise levels at 100 μAU or above for rat 169. The issue is a challenging one from an experimental standpoint, however. Increasing the pressure on the interface will result in greater compression of the tissue and a corresponding higher light throughput and lower noise, but such increased pressure will also cause greater variation in the tissue background with time. The ultimate solution would be to have greater incident source power, but that would require a higher brightness light source than is currently available. A higher brightness source might also

permit an interface based on a reflectance geometry. This might help to alleviate deformation of the tissue over the time course of the measurements.

Two other problematic issues that were made apparent by the work performed with the four rats are (1) the difficulty of establishing a workable and accurate way of handling the lag between arterial and tissue glucose concentrations and (2) the need for more dynamic components in the alarm algorithm.

The issue of lag was addressed in this work by taking a block of calibration data and fitting it to a quantitative model for concentration as the lag was adjusted. It was hypothesized that the best concentration fit would signal the most appropriate lag. Across the four rats employed, lag values of 11, 7, 29 and 0 minutes were obtained. These values are not consistent enough to be satisfying or convincing. Furthermore, they assume a constant lag both within a day and between days. The results from the fourth prediction day of rat 134 emphasize this concern, as they suggest the need for a different lag time than determined on the calibration day. A more extensive study of this issue is beyond the scope of this dissertation, but is clearly needed.

Finally, the hypoglycemic alarm algorithm worked well as long as the assumptions upon which it was based appeared to hold. Correct choice of the reference spectrum to use in the calculation of differential spectra was critical. Issues encountered in the *in vivo* work were a need to detect when the rat had equilibrated with the interface, as well as a need to determine when or if the reference required updating. Essentially ad hoc solutions to these problems were implemented here (e.g., waiting 15 minutes before taking the reference), but a more elegant dynamic approach to both problems is clearly needed.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

Summary of Thesis Contributions

The ability to measure glucose selectively in *in vivo* spectra collected in the combination region of the NIR has been demonstrated previously.[10,151] This dissertation utilized this ability to measure glucose selectively and extended it to an investigation of a possible classification method to identify occurrences of hypoglycemia. A nocturnal hypoglycemic alarm algorithm was developed and tested both with *in vitro* simulation studies and *in vivo* studies using rat animal models.

The requirement for conventional invasive reference glucose measurements limited the size of the calibration database available for use in developing the hypoglycemic alarm algorithm. In this thesis, this was addressed by using a differential spectral computation in which spectra in absorbance units were computed by taking the ratios of all combinations of the collected single-beam spectra within a defined time block. By orienting the ratios such that negative differential concentrations were obtained, a simple correspondence was established between the concentrations in the calibration database and the differential concentrations obtained when spectra collected continuously are referenced to a spectrum collected at the start of the sleep period. By obtaining a conventional invasive glucose measurement to match the reference spectrum, differential concentrations signaling hypoglycemia will necessarily correspond to a specific negative value. This allowed the calibration database to be split into two groups: differential concentrations above or below a user-specified hypoglycemic threshold. On

the basis of these groupings, classification models could be constructed to produce a yes/no response regarding the occurrence of hypoglycemia.

The use of differential spectra increased the size of the calibration database significantly. In addition, the calculation of differential spectra served to remove spectral information corresponding to a constant background between every two spectra whose ratio was taken. As an example, if the thickness of the skin tissue matrix were constant, the spectral variance associated with it would be constant and would therefore be canceled when the ratio was taken.

The glucose-related information in the sample matrix was extracted using the multivariate inverse regression method, PLS, which utilized the covariance between the spectral and concentration information to compute a set of latent variables and corresponding scores. In PLS, the dimensionality of each spectrum can be reduced from hundreds of points to an equivalent representation based on fewer than 20 factors. This data reduction/feature extraction step was very important for the nocturnal hypoglycemic alarm algorithm application discussed in this thesis. The pattern recognition method, PLDA, requires iterative optimization steps to find the optimal classification model. The optimization is both much faster and more reproducible when using PLS scores as patterns in the classification model than when using the entire spectrum.

Another important step in the design of the alarm algorithm was the calculation of replicate classification models. In implementing the alarm, three replicate classifiers were computed and the alarm decision was based on the performance of the classifiers as a group. An alarm decision rule was introduced in which any prediction pattern was termed an “alarm” if two of the three replicate classifiers predicted it as an “alarm”. This

minimized the number of missed or false alarms since the decision was not based on the precise optimization of a single classifier.

The exact number of discriminant functions used to build each piecewise linear discriminant was directly dependent on the structure of the data space formed by the patterns in the training set. For the research discussed in this thesis, the number of discriminants used for the classification was determined by the number of training patterns separated. Only the discriminants that separated a significant number of patterns were used for the classification. For the relatively simple sample matrix discussed in Chapter 4, a single discriminant function separated more than 95 % of the data and only the first discriminant was used for the classification. However, 2-3 discriminants were observed to be significant for the complex sample matrixes discussed in Chapters 6 and 7.

The choice of the optimal number of discriminants was observed to be important in helping to ensure the accuracy of subsequent data predictions. In the application of PLDA in this thesis, if any discriminant classified the unknown pattern as an alarm, the piecewise linear discriminant classified that pattern as an alarm. Therefore, if more discriminants than required are used for the classification, the probability is increased to classify unknown patterns as alarms and thus the rate of false alarms may increase. Correspondingly, if fewer than the optimal number of discriminants are used, a greater likelihood of missed alarms exists.

One of the challenges in defining an automated protocol for the implementation of the alarm algorithm was the need to optimize several important parameters. The number of PLS factors and the wavenumber range used for the classification was optimized in

two steps. In PLS, the use of too many latent variables incorporates noise or unneeded spectral features into the model. By contrast, the use of less than the optimal number of factors may prevent important spectral information from being incorporated into the model. The use of the correct wavenumber range has a similar effect: the inclusion of unneeded spectral features or the exclusion of important features.

The PLS factor-wavenumber combination was first optimized using a grid search analysis, in which an *F*-test corrected SECV value was used as the response function to find the optimum. The results presented previously in Chapters 4, 6, and 7 demonstrated that the grid search analysis optimized the wavenumber range correctly to cover the known glucose absorption features. However, the highest number of PLS factors tested tended to always be judged as optimal, and the use of the *F*-test to eliminate insignificant terms was not effective because of the high number of degrees of freedom (i.e., very small changes in *F*-values were judged significant).

For this reason, the grid search was used only to identify a set of 4-5 best-performing wavenumber ranges and a further optimization based on PLDA was used to select the final combination of wavenumber range and number of latent variables. The results discussed in the previous chapters clearly demonstrated that the optimal result obtained from the PLDA optimization was typically not the top wavenumber-latent variable combination produced by the grid search analysis. Therefore, the PLDA-based optimization was very important to find the best wavenumber-PLS factor combination for use in building the final classification model.

It was also observed that the PLDA classification results were highly sensitive to the wavenumber-latent variable combinations for the complex sample matrixes. Even

though they were judged to be well-performing by the grid search, some combinations performed poorly when applied to the monitoring data. This underscored the need to test the selected parameters with external prediction data (i.e., the monitoring set) rather than simply relying on cross-validation within the calibration set. This was especially important with the *in vivo* data presented in Chapter 7.

A series of experimental investigations was performed to validate the proposed nocturnal hypoglycemic alarm algorithm. Chapter 4 discussed a five-component dynamic system composed of glucose, urea, lactate, glyceryl triacetate and pH 7.4 phosphate buffer. The results presented in the chapter clearly demonstrated excellent data prediction, giving no missed or false alarms. The robustness of the model was further demonstrated by use of data collected over an extended time period. The PLS scores plot for the alarm and non-alarm patterns further demonstrated a clear separation between the data classes, confirming the utility of implementing the alarm decision with a classification method such as PLDA.

Chapter 5 discussed an approach to make a skin tissue phantom composed of keratin and collagen which could be used to simulate data collection in the human skin tissue matrix. The thickness of each film was evaluated using a linear regression fit to experimentally measured human skin tissue spectra, and the films were designed to match the equivalent amount of keratin or collagen observed in these spectra of human tissue. Initial studies were performed with films of uniform thickness and follow-on studies were then performed in which films of varying thickness were created. This allowed the incorporation of variation in film thickness into the *in vitro* simulation studies.

The R^2 value was used to determine how well the predicted spectra from the regression fit matched the actual spectra of skin tissue. Values of R^2 in the range of 0.97 to 0.99 demonstrated both the films of uniform and varying thickness compared well to the actual spectra. However, plots of actual vs. predicted spectra showed a clear mismatch in the protein feature area in the NIR spectrum. This might be due to differences in the protein structure of the keratin and gelatin powders used with those present in skin tissue. Another possibility is structural variation caused by the need to use formic acid as the solvent in casting the films.

In this thesis, an experimental methodology was proposed to produce variable-thickness films which provided some ability to change the effective skin tissue component content of a simulated tissue sample. This allowed the simulation of changes in the skin tissue components that might occur in *in vivo* measurements conducted at different measurement sites on the same individual or at different times.

In the work discussed in Chapter 6, the noninvasive nocturnal alarm algorithm methodology was tested and its performance was evaluated with the two tissue phantoms described above to simulate the spectral properties of human tissue. The alarm algorithm performed well for the data collected in a dynamic system in the presence of the films in the same orientation during all the days of data collection. However, more missed and false alarms were observed when the film thickness was allowed to change while the spectra were collected. These results illustrate the need for the calibration database to incorporate the range of variation to be encountered when the classification model is applied to external prediction data.

In the work discussed in Chapter 7, the noninvasive nocturnal alarm algorithm methodology was tested and its performance was evaluated with a much more complex sample matrix. The alarm algorithm was tested with the data collected from rat animal models in both single-day and multiple-day data collections. The single-day experiments were performed with anesthetized rats under a non-survival protocol. The multiple-day data collections were performed with awake rats. The results obtained clearly demonstrated that the alarm algorithm performed well with the data collected from the single-day data collection where the physical and chemical spectral variances were minimal. However, the performance of the algorithm for the multiple-day studies varied. Some days performed relatively well while some had poor predictions. The thicknesses for the skin tissue components appeared to be different for each day of data collection and even within a single day they appeared to change. Similar to the results obtained with the variable-thickness protein films described in Chapter 6, this variation may be the key factor in negatively affecting prediction accuracy.

Related to the variation in apparent skin thickness was variation in observed spectral RMS noise levels. From rat to rat and day to day, average RMS noise values varied from approximately 20 to well over 100 μAU. The largest contributor to the observed spectral noise was the intrinsic noise of the detector, made more significant as the number of signal photons decreased. A thicker section of tissue within the optical interface, such as that found for rat 169, translated into fewer transmitted photons and a correspondingly higher noise level. The task of differentiating glucose concentrations above and below the alarm threshold was clearly made more difficult in the presence of increased spectral noise.

The overall conclusion from this work is that the proposed hypoglycemic alarm algorithm can perform extremely effectively as long as high-quality, reproducible spectra are obtained that can be accurately represented by the calibration database. While the use of differential spectra as the inputs to the alarm algorithm has several practical advantages in terms of expanding the calibration database and removing common background components, these spectra are not sufficiently resistant to changes in the background matrix to allow the alarm algorithm to overcome the levels of variation observed from day to day in the *in vivo* data. Thus, further improvements in the experimental protocol are needed in the *in vivo* measurements to improve the quality and reproducibility of the spectral data.

Future Work

The performance of the alarm algorithm in the presence of some key interfering components in the NIR region was studied in the thesis. However, the sample matrix could be made more representative by incorporating additional components into the optical path of the measurement. For example, a fat phantom could be added in addition to the protein phantom described in Chapter 5. Fat present in the skin tissue matrix absorbs as well as scatters NIR light, thereby interfering with the glucose analysis. As noted previously, the fat absorption regions between 4400 and 4200 cm^{-1} interfere with the two main glucose absorption bands at 4300 and 4400 cm^{-1}. A fat tissue phantom could be prepared by melting bovine or porcine fat and applying it to the sample holder in much the same procedure as discussed previously for preparing the protein phantoms.

The aqueous matrix used with the dynamic system could also be altered to reflect the composition of human interstitial fluid more accurately.

Performing PCA with mean-centered rat skin tissue absorbance spectra demonstrated that the first principal component explained approximately 95% of the total spectral variance. The first loading for rat 106 is shown in Figure 8.1. This figure clearly exhibits broad spectral features and some baseline variation across the spectrum. This baseline variation and the broad features are indications of variation in the optical path length and the degree of light scattering encountered with the skin tissue matrix. Data preprocessing methods could be used to correct for these potential physical phenomena and interferences that result in unwanted signal variability.

In this thesis, the differential spectral computation was performed as a preprocessing method on the basis of assuming that spectral ratios would remove common background components. However, it was observed that the skin tissue matrix changes during the data collection. The light scattering from the skin tissue is manifest as a multiplicative interference across the NIR spectral region and could potentially be corrected by use of a spectral preprocessing method. For example, multiplicative scatter correction (MSC) [152] could be used to help correct this variation. For the application discussed in this thesis, the single-beam spectra could be corrected using MSC before the calculation of the differential spectra.

Variation in the thickness of the skin tissue can be observed as baseline variation in the collected spectra. Derivatives are often used to reduce this effect, and for the application discussed here, the spectra could be preprocessed using a derivative computation. This calculation also enhances small spectral features. Either before or after

the differential spectral computations, spectra could be preprocessed using a derivative calculation. Other digital filtering methods could also be explored.

One of the problems noticed in the multiple-day data collection with the rat animal models was the variation in skin thickness from day to day. The spectra collected on each day could be normalized by the sum of the regression coefficients that represent the physical thickness of each of the modeled components of the skin tissue. This could produce a normalized spectrum for each day and would help to account for the thickness variations from day to day.

The alarm decision rule could also be studied further. It was observed that in replicate classifiers, some predicted well while others predicted poorly. The number of classifiers could be increased and the alarm decision could be taken based on the performance of more classifiers rather than the use of two out of three as performed in this thesis. In addition, the time constant of the change in glucose concentration could be incorporated into the classification rule. For example, given the observed rate of change of the glucose concentration, a requirement could be implemented that several consecutive alarms must be encountered before an actual alarm is signaled. More sophisticated use of the previous discriminant scores in implementing the alarm decision for the current score could also be studied.

Another issue related to the classification models was the poor separation of alarm and non-alarm patterns obtained for the training data corresponding to several of the rats studied during the *in vivo* measurements. This indicates difficulty in establishing the classification boundary between the alarm and non-alarm data classes. In some cases, this is an artifact of the distribution of differential concentrations relative to the critical

concentration. For example, if the majority of the training data is concentrated at the alarm threshold, the separation will necessarily be challenging and the percentage of alarm patterns separated may be low. However, such a case puts an emphasis on the training methodology being able to identify an optimal placement of the discriminant boundary. The need for replicate classifiers was made apparent in the current work, as any single classification boundary may not be optimally placed.

The challenging nature of the separation of the training patterns suggests that other pattern classification methodologies should be investigated for comparison to PLDA. For example, support vector machines (SVMs) and artificial neural networks (ANNs) implement true nonlinear classification models and might provide better performance. These methods also have a fundamental design difference than PLDA in the way in which false alarms in the training data are handled. As described in Chapter 3, the requirement for single-sided discriminants biases PLDA against false alarms in the training set (i.e., they are not allowed). Both SVMs and ANNs allow false alarms during training in seeking an overall best placement of the class separation boundary. An assessment of how this design characteristic would affect the performance of the alarm algorithm would be an interesting and potentially valuable study.

The results obtained from the *in vivo* measurements also illustrate the need for an additional layer of automation and decision-making to be applied to the hypoglycemic alarm algorithm. For example, the ability to self-diagnose when the classification model is out of specification would be very valuable. In such a case, the patient could be wakened and a new reference collected. Updating the reference provided some benefit in the work with rat 169. Implementing such an approach requires a reliable performance

diagnostic, however. One potential diagnostic would be to fit each differential spectrum to the pure-component spectra of tissue components. In principle, the fitted coefficients should be zero if the current reference is valid (i.e., the calculation of the differential spectrum should have effectively subtracted the tissue components). The occurrence of regression coefficients significantly different from zero would signal the need for updating the reference. A similar approach could be used to identify spectral outliers that might arise from a sudden movement by the patient while sleeping. Automation and self-diagnosis could also be applied to issues such as the number of discriminant functions to employ and the number of replicate classifiers needed.

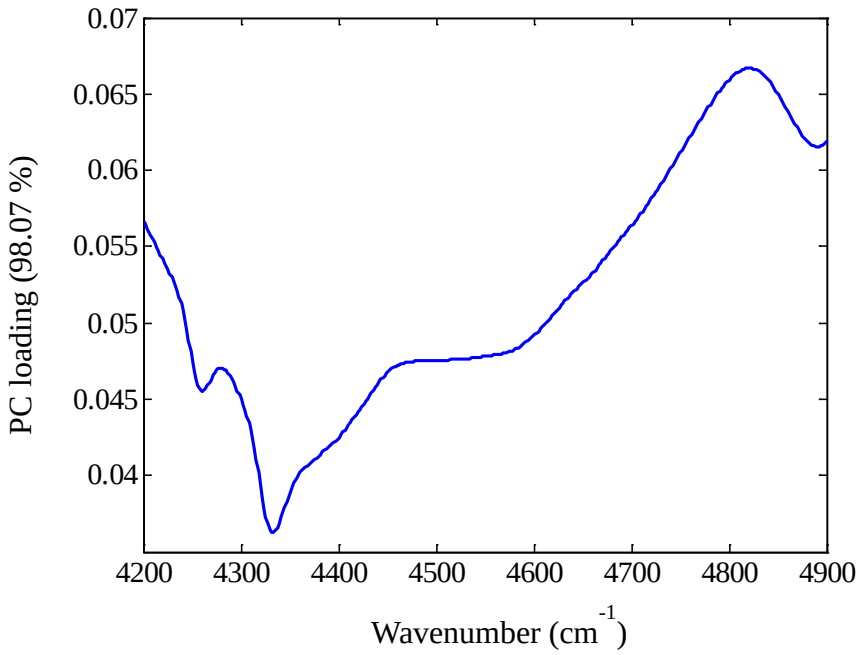


Figure 8.1. First principal component loading plot for the differential spectra collected with rat 106. The first loading explains more than 95 % of the total spectral variance. The features are broad and some baseline variation can be observed.

Overall, this thesis successfully developed an alarm algorithm that can be applied for the identification of dangerous hypoglycemic occurrences while patients are sleeping. The *in vitro* simulation work demonstrated the excellent performance of this algorithm. However, the *in vivo* studies demonstrated that some problems need to be addressed before the methodology can be applied to human subjects. As further efforts are expended to improve the experimental aspects of the *in vivo* measurements, it may well be possible to apply the alarm methodology to other applications in which continuous monitoring by NIR spectroscopy is attractive. For example, the same procedures discussed here for the nocturnal hypoglycemic alarm methodology could be used in an industrial quality control setting to identify situations in which a process goes out of specification.

REFERENCES

1. International Diabetes Federation. IDF Diabetes Atlas, 6th edition.Brussels,Belgium. http://www.idf.org/diabetesatlas (accessed 3/30 2014).

2. American Diabetes Association.Diabetes Basics.Statistics About Diabetes. http://www.diabetes.org/diabetes-basics/diabetes-statistics (accessed 03/30 2014).

3. American Diabetes Association. Diabetes Basics.Type 2. http://www.diabetes.org/diabetes-basics/type-2/ (accessed 03/30 2014).

4. American Diabetes Association.Living With Diabetes.Complications. http://www.diabetes.org/living-with-diabetes/complications (accessed 03/30 2014).

5. American Diabetes Association.Advocacy.The Cost of Diabetes. http://www.diabetes.org/for-media/2013/annual-costs-of-diabetes-2013.html (accessed 03/30 2014).

6. Virginia Mason Medical Center.Treatment.What are normal blood glucose levels. https://www.virginiamason.org/WhatareNormalBloodGlucoseLevels (accessed 03/30 2014).

7. Kerry, C.; Mitchell, S.; Sharma, S.; Scott, A.; Rayman, G. Diurnal temporal patterns of hypoglycaemia in hospitalized people with diabetes may reveal potentially correctable factors. *Diabetic Medicine* 2013, *30*, 1403-1406.

8. Kim, H. S.; Shin, J. A.; Chang, J. S.; Cho, J. H.; Son, H. Y.; Yoon, K. H. Continuous glucose monitoring: current clinical use. *Diabetes/Metabolism Research and Reviews*, *28*, 73-78.

9. Arnold, M. A.; Small, G. W. Noninvasive glucose sensing. *Analytical Chemistry* 2005, *77*, 5429-5439.

10. Olesberg, J. T.; Liu, L.; Zee, V. V.; Arnold, M. A. In vivo near-infrared spectroscopy of rat skin tissue with varying blood glucose levels. *Analytical Chemistry* 2006, *78*, 215-223.

11. Luypaert, J.; Massart, D. L.; Vander Heyden, Y. Near-infrared spectroscopy applications in pharmaceutical analysis. *Talanta* 2007, *72*, 865-883.

12. Shaffer, R. E.; Small, G. W. Improved response function for the Simplex optimization of piecewise linear discriminants. *Chemometrics and Intelligent Laboratory Systems* 1996, *32*, 95-109.

13. Kaltenbach, T. F.; Small, G. W. Development and optimization of piecewise linear discriminants for the automated detection of chemical species. *Analytical Chemistry* 1991, *63*, 936-944.

14. Hönes, J.; Müller, P.; Surridge, N. The technology behind glucose meters: test strips. *Diabetes Technology & Therapeutics* 2008, *10*, S-10-S-26.

15. Clarke, S. F.; Foster, J. R. A history of blood glucose meters and their role in self-monitoring of diabetes mellitus. *British Journal of Biomedical Science*, *69*, 83-93.

16. Cheeley, R. D.; Joyce, S. M. A clinical comparison of the performance of four blood glucose reagent strips. *The American Journal of Emergency Medicine* 1990, *8*, 11-15.

17. Wang, J. Glucose biosensors: 40 years of advances and challenges. *Electroanalysis* 2001, *13*, 983.

18. Nagy, G.; Von Storp, L. H.; Guilbault, G. G. Enzyme electrode for glucose based on an iodide membrane sensor. *Analytica Chimica Acta* 1973, *66*, 443-445.

19. Henry, C. Getting under the skin: implantable glucose sensors. *Analytical Chemistry* 1998, *70*, 594A-598A.

20. Wang, J. Electrochemical glucose biosensors. *Chemical Reviews* 2008, *108*, 814-825.

21. Oliver, N. S.; Toumazou, C.; Cass, A. E. G.; Johnston, D. G. Glucose sensors: a review of current and emerging technology. *Diabetic Medicine* 2009, *26*, 197-210.

22. Oliver, N.; Toumazou, C.; Cass, A.; Johnston, D. Glucose sensors: a review of current and emerging technology. *Diabetic Medicine* 2009, *26*, 197-210.

23. Thomé-Duret, V. r.; Reach, G. r.; Gangnerau, M. N. l.; Lemonnier, F.; Klein, J. C.; Zhang, Y.; Hu, Y.; Wilson, G. S. Use of a subcutaneous glucose sensor to detect decreases in glucose concentration prior to observation in blood. *Analytical Chemistry* 1996, *68*, 3822-3826.

24. Schmidtke, D. W.; Freeland, A. C.; Heller, A.; Bonnecaze, R. T. Measurement and modeling of the transient difference between blood and subcutaneous glucose concentrations in the rat after injection of insulin. *Proceedings of the National Academy of Sciences* 1998, *95*, 294-299.

25. Mastrototaro, J. J. The MiniMed continuous glucose monitoring system. *Diabetes Technology & Therapeutics* 2000, *2*, 13-18.

26. Gross, T. M.; Bode, B. W.; Einhorn, D.; Kayne, D. M.; Reed, J. H.; White, N. H.; Mastrototaro, J. J. Performance evaluation of the MiniMed® continuous glucose monitoring system during patient home use. *Diabetes Technology & Therapeutics* 2000, *2*, 49-56.

27. Skyler, J. S. Continuous glucose monitoring: an overview of its development. *Diabetes Technology & Therapeutics* 2009, *11*, S-5-S-10.

28. Poscia, A.; Mascini, M.; Moscone, D.; Luzzana, M.; Caramenti, G.; Cremonesi, P.; Valgimigli, F.; Bongiovanni, C.; Varalli, M. A microdialysis technique for continuous subcutaneous glucose monitoring in diabetic patients (part 1). *Biosensors and Bioelectronics* 2003, *18*, 891-898.

29. Pasic, A.; Koehler, H.; Schaupp, L.; Pieber, T. R.; Klimant, I. Fiber-optic flow-through sensor for online monitoring of glucose. *Analytical and Bioanalytical Chemistry* 2006, *386*, 1293-1302.

30. Garg, S. K.; Kelly, W. C.; Voelmle, M. K.; Ritchie, P. J.; Gottlieb, P. A.; McFann, K. K.; Ellis, S. L. Continuous Home Monitoring of Glucose Improved glycemic control with real-life use of continuous glucose sensors in adult subjects with type 1 diabetes. *Diabetes Care* 2007, *30*, 3023-3025.

31. Garg, S.; Zisser, H.; Schwartz, S.; Bailey, T.; Kaplan, R.; Ellis, S.; Jovanovic, L. Improvement in Glycemic Excursions With a Transcutaneous, Real-Time Continuous Glucose Sensor A randomized controlled trial. *Diabetes Care* 2006, *29*, 44-50.

32. Feldman, B.; Brazg, R.; Schwartz, S.; Weinstein, R. A continuous glucose sensor based on Wired Enzyme™ technology-Results from a 3-day trial in patients with type 1 diabetes. *Diabetes Technology & Therapeutics* 2003, *5*, 769-779.

33. Langemann, H.; Mendelowitsch, A.; Landolt, H.; Alessandri, B.; Gratzl, O. Experimental and clinical monitoring of glucose by microdialysis. *Clinical Neurology and Neurosurgery* 1995, *97*, 149-155.

34. Vering, T.; Adam, S.; Drewer, H.; Dumschat, C.; Steinkuhl, R.; Knoll, M.; Schulze, A.; Siegel, E. G. Wearable microdialysis system for continuous in vivo monitoring of glucose. *Analyst* 1998, *123*, 1605-1609.

35. Schoonen, A. J. M.; Wientjes, K. J. C. A model for transport of glucose in adipose tissue to a microdialysis probe. *Diabetes Technology & Therapeutics* 2003, *5*, 589-598.

36. Rhemrev-Boom, R. M.; Tiessen, R. G.; Jonker, A. A.; Venema, K.; Vadgama, P.; Korf, J. A lightweight measuring device for the continuous in vivo monitoring of

glucose by means of ultraslow microdialysis in combination with a miniaturised flow-through biosensor. *Clinica Chimica Acta* 2002, *316*, 1-10.

37. Jacobs, P. G.; Youssef, J. E.; Castle, J. R.; Engle, J. M.; Branigan, D. L.; Johnson, P.; Massoud, R.; Kamath, A.; Ward, W. K. In *Tilte*; IEEE.

38. Sieg, A.; Guy, R. H.; Delgado-Charro, M. B. a. Noninvasive and minimally invasive methods for transdermal glucose monitoring. *Diabetes Technology & Therapeutics* 2005, *7*, 174-197.

39. Rao, G.; Guy, R. H.; Glikfeld, P.; LaCourse, W. R.; Leung, L.; Tamada, J.; Potts, R. O.; Azimi, N. Reverse iontophoresis: noninvasive glucose monitoring in vivo in humans. *Pharmaceutical Research* 1995, *12*, 1869-1873.

40. Tamada, J. A.; Bohannon, N. J. V.; Potts, R. O. Measurement of glucose in diabetic subjects using noninvasive transdermal extraction. *Nature Medicine* 1995, *1*, 1198-1201.

41. Kost, J.; Mitragotri, S.; Gabbay, R. A.; Pishko, M.; Langer, R. Transdermal monitoring of glucose and other analytes using ultrasound. *Nature Medicine* 2000, *6*, 347-350.

42. Mitragotri, S.; Coleman, M.; Kost, J.; Langer, R. Analysis of ultrasonically extracted interstitial fluid as a predictor of blood glucose levels. *Journal of Applied Physiology* 2000, *89*, 961-966.

43. Burdick, J.; Chase, P.; Faupel, M.; Schultz, B.; Gebhart, S. Real-time glucose sensing using transdermal fluid under continuous vacuum pressure in children with type 1 diabetes. *Diabetes Technology & Therapeutics* 2005, *7*, 448-455.

44. Gebhart, S.; Faupel, M.; Fowler, R.; Kapsner, C.; Lincoln, D.; McGee, V.; Pasqua, J.; Steed, L.; Wangsness, M.; Xu, F. Glucose sensing in transdermal body fluid collected under continuous vacuum pressure via micropores in the stratum corneum. *Diabetes Technology & Therapeutics* 2003, *5*, 159-166.

45. Svedman, C.; Samra, J. S.; Clark, M. L.; Levy, J. C.; Frayn, K. N. Skin mini-erosion technique for monitoring metabolites in interstitial fluid: its feasibility demonstrated by OGTT results in diabetic and non-diabetic subjects. *Scandinavian Journal of Clinical & Laboratory Investigation* 1999, *59*, 115-123.

46. Volden, G.; Thorsrud, A. K.; Bjørnson, I.; Jellum, E. Biochemical composition of suction blister fluid determined by high resolution multicomponent analysis (capillary gas chromatography--mass spectrometry and two-dimensional electrophoresis). *Journal of Investigative Dermatology* 1980, *75*, 421-424.

47. Tierney, M. J.; Tamada, J. A.; Potts, R. O.; Jovanovic, L.; Garg, S. Clinical evaluation of the GlucoWatch® biographer: a continual, non-invasive glucose monitor for patients with diabetes. *Biosensors and Bioelectronics* 2001, *16*, 621-629.

48. Girardin, C. l. M.; Huot, C. l.; Gonthier, M.; Delvin, E. Continuous glucose monitoring: A review of biochemical perspectives and clinical use in type 1 diabetes. *Clinical Biochemistry* 2009, *42*, 136-142.

49. National Diabetes Data, G.; National Institute of, D.; Digestive; Kidney, D.: *Diabetes in America*; National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 1995.

50. Clarke, W. L.; Cox, D.; Gonder-Frederick, L. A.; Carter, W.; Pohl, S. L. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care* 1987, *10*, 622-628.

51. Johnson, R. N.; Baker, J. R. Error detection and measurement in glucose monitors. *Clinica Chimica Acta* 2001, *307*, 61-67.

52. Larin, K. V.; Eledrisi, M. S.; Motamedi, M.; Esenaliev, R. O. Noninvasive blood glucose monitoring with optical coherence tomography a pilot study in human subjects. *Diabetes Care* 2002, *25*, 2263-2267.

53. Larin, K. V.; Larina, I. V.; Motamedi, M.; Gelikonov, V. M.; Kuranov, R. V.; Esenaliev, R. O. In *Tilte*2001; International Society for Optics and Photonics.

54. Sapozhnikova, V. V.; Kuranov, R. V.; Cicenaite, I.; Esenaliev, R. O.; Prough, D. S. Effect on blood glucose monitoring of skin pressure exerted by an optical coherence tomography probe. *Journal of Biomedical Optics* 2008, *13*, 021112-021112-021114.

55. Sapozhnikova, V. V.; Prough, D.; Kuranov, R. V.; Cicenaite, I.; Esenaliev, R. O. Influence of osmolytes on in vivo glucose monitoring using optical coherence tomography. *Experimental Biology and Medicine* 2006, *231*, 1323-1332.

56. Cito, S.; Ahn, Y.-C.; Pallares, J.; Duarte, R. M.; Chen, Z.; Madou, M.; Katakis, I. Visualization and measurement of capillary-driven blood flow using spectral domain optical coherence tomography. *Microfluidics and Nanofluidics* 2012, *13*, 227-237.

57. Pfützner, A.; Caduff, A.; Larbig, M.; Schrepfer, T.; Forst, T. Impact of posture and fixation technique on impedance spectroscopy used for continuous and noninvasive glucose monitoring. *Diabetes Technology & Therapeutics* 2004, *6*, 435-441.

58. Kajiwara, K.; Uemura, T.; Kishikawa, H.; Nishida, K.; Hashiguchi, Y.; Uehara, M.; Sakakida, M.; Ichinose, K.; Shichiri, M. Noninvasive measurement of blood glucose concentrations by analysing Fourier transform infra-red absorbance spectra through oral mucosa. *Medical and Biological Engineering and Computing* 1993, *31*, S17-S22.

59. Burmeister, J. J.; Arnold, M. A.; Small, G. W. Noninvasive blood glucose measurements by near-infrared transmission spectroscopy across human tongues. *Diabetes Technology & Therapeutics* 2000, *2*, 5-16.

60. Hazen, K. H.; Arnold, M. A.; Small, G. W. Measurement of glucose in water with first-overtone near-infrared spectra. *Applied Spectroscopy* 1998, *52*, 1597-1605.

61. Shen, Y. C.; Davies, A. G.; Linfield, E. H.; Elsey, T. S.; Taday, P. F.; Arnone, D. D. The use of Fourier-transform infrared spectroscopy for the quantitative determination of glucose concentration in whole blood. *Physics in Medicine and Biology* 2003, *48*, 2023.

62. Hunter, M.; Enejder, A.; Scecina, T.; Feld, M.; Shih, W.-C.: Raman spectroscopy for non-invasive glucose measurements. Google Patents, 2006.

63. Lambert, J. L.; Morookian, J. M.; Sirk, S. J.; Borchert, M. S. Measurement of aqueous glucose in a model anterior chamber using Raman spectroscopy. *Journal of Raman Spectroscopy* 2002, *33*, 524-529.

64. Russell, R. J.; Pishko, M. V.; Gefrides, C. C.; McShane, M. J.; Cote, G. L. A fluorescence-based glucose biosensor using concanavalin A and dextran encapsulated in a poly (ethylene glycol) hydrogel. *Analytical Chemistry* 1999, *71*, 3126-3132.

65. Mansouri, S.; Schultz, J. S. A miniature optical glucose sensor based on affinity binding. *Nature Biotechnology* 1984, *2*, 885-890.

66. Cameron, B. D.; Baba, J. S.; CotÃ©, G. L. Measurement of the glucose transport time delay between the blood and aqueous humor of the eye for the eventual development of a noninvasive glucose sensor. *Diabetes Technology & Therapeutics* 2001, *3*, 201-207.

67. Cameron, B. D.; Anumula, H. Development of a real-time corneal birefringence compensated glucose sensing polarimeter. *Diabetes Technology & Therapeutics* 2006, *8*, 156-164.

68. Cameron, B. D.; Baba, J. S.; Coté, G. L. Measurement of the glucose transport time delay between the blood and aqueous humor of the eye for the eventual development of a noninvasive glucose sensor. *Diabetes Technology & Therapeutics* 2001, *3*, 201-207.

69. Sierra.J.F, G. J., DcMarcos.S, Castillo.J.R.;. Direct determination of glucose in serum by fluorimetry using a labelled enzyme. *Anal.Chim.Acta* 2000, *414*, 33-41.

70. Amerov, A. K.; Chen, J.; Arnold, M. A. Molar absorptivities of glucose and other biological molecules in aqueous solutions over the first overtone and combination regions of the near-infrared spectrum. *Applied Spectroscopy* 2004, *58*, 1195-1204.

71. Block, M. J.: Non-invasive IR transmission measurement of analyte in the tympanic membrane. Google Patents, 1999.

72. Heise, H. M.; Bittner, A.; Marbach, R. Near-infrared reflectance spectroscopy for noninvasive monitoring of metabolites. *Clinical chemistry and Laboratory Medicine* 2000, *38*, 137-145.

73. Maruo, K.; Tsurugi, M.; Tamura, M.; Ozaki, Y. *In Vivo* Noninvasive Measurement of Blood Glucose by Near-Infrared Diffuse-Reflectance Spectroscopy. *Applied Spectroscopy* 2003, *57*, 1236-1244.

74. Marbach, R.; Koschinsky, T. H.; Gries, F. A.; Heise, H. M. Noninvasive blood glucose assay by near-infrared diffuse reflectance spectroscopy of the human inner lip. *Applied Spectroscopy* 1993, *47*, 875-881.

75. Weiss, R.; Yegorchikov, Y.; Shusterman, A.; Raz, I. Noninvasive continuous glucose monitoring using photoacoustic technology-results from the first 62 subjects. *Diabetes Technology & Therapeutics* 2007, *9*, 68-74.

76. MacKenzie, H. A.; Ashton, H. S.; Spiers, S.; Shen, Y.; Freeborn, S. S.; Hannigan, J.; Lindberg, J.; Rae, P. Advances in photoacoustic noninvasive glucose testing. *Clinical Chemistry* 1999, *45*, 1587-1595.

77. Burmeister, J. J.; Arnold, M. A. Evaluation of measurement sites for noninvasive blood glucose sensing with near-infrared transmission spectroscopy. *Clinical Chemistry* 1999, *45*, 1621-1627.

78. Heise, H. M.; Marbach, R.; Koschinsky, T. H.; Gries, F. A. Noninvasive Blood Glucose Sensors Based on Near-Infrared Spectroscopy. *Artificial Organs* 1994, *18*, 439-447.

79. Robinson, M. R.; Eaton, R. P.; Haaland, D. M.; Koepp, G. W.; Thomas, E. V.; Stallard, B. R.; Robinson, P. L. Noninvasive glucose monitoring in diabetic patients: a preliminary evaluation. *Clinical Chemistry* 1992, *38*, 1618-1622.

80. Hazen, K. H.; Arnold, M. A.; Small, G. W. Temperature-insensitive near-infrared spectroscopic measurement of glucose in aqueous solutions. *Applied Spectroscopy* 1994, *48*, 477-483.

81. Hanlon, E. B.; Manoharan, R.; Koo, T. W.; Shafer, K. E.; Motz, J. T.; Fitzmaurice, M.; Kramer, J. R.; Itzkan, I.; Dasari, R. R.; Feld, M. S. Prospects for in vivo Raman spectroscopy. *Physics in Medicine and Biology* 2000, *45*, R1.

82. Anker, J. N.; Hall, W. P.; Lyandres, O.; Shah, N. C.; Zhao, J.; Van Duyne, R. P. Biosensing with plasmonic nanosensors. *Nature Materials* 2008, *7*, 442-453.

83. Shafer-Peltier, K. E.; Haynes, C. L.; Glucksberg, M. R.; Van Duyne, R. P. Toward a glucose biosensor based on surface-enhanced Raman scattering. *Journal of the American Chemical Society* 2003, *125*, 588-593.

84. Palmer, K. F.; Williams, D. Optical properties of water in the near infrared. *JOSA* 1974, *64*, 1107-1110.

85. Hale, G. M.; Querry, M. R.; Rusk, A. N.; Williams, D. Influence of temperature on the spectrum of water. *JOSA* 1972, *62*, 1103-1108.

86. Small, G. W. Chemometrics and near-infrared spectroscopy: avoiding the pitfalls. *TrAC Trends in Analytical Chemistry* 2006, *25*, 1057-1066.

87. Chen, J.; Arnold, M. A.; Small, G. W. Comparison of combination and first overtone spectral regions for near-infrared calibration models for glucose and other biomolecules in aqueous solutions. *Analytical Chemistry* 2004, *76*, 5405-5413.

88. Arnold, M. A.; Olesberg, J. T.; Small, G. W. Near-infrared spectroscopy for noninvasive glucose sensing. *Analytical chemistry of in vivo glucose measurements. Hoboken, NJ: John Wiley & Sons* 2009, 357-390.

89. Krämer, K.; Ebel, S. Application of NIR reflectance spectroscopy for the identification of pharmaceutical excipients. *Analytica Chimica Acta* 2000, *420*, 155-161.

90. Roggo, Y.; Chalus, P.; Maurer, L.; Lema-Martinez, C.; Edmond, A. l.; Jent, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis* 2007, *44*, 683-700.

91. Williams, P.; Norris, K.: *Near-infrared technology in the agricultural and food industries*; American Association of Cereal Chemists, Inc., 1987.

92. Larrechi, M. S.; Callao, M. P. Strategy for introducing NIR spectroscopy and multivariate calibration techniques in industry. *TrAC Trends in Analytical Chemistry* 2003, *22*, 634-640.

93. Blanco, M.; Maspoch, S.; Villarroya, I.; Peralta, X.; Gonzalez, J. M.; Torres, J. Geographical origin classification of petroleum crudes from near-infrared spectra of bitumens. *Applied Spectroscopy* 2001, *55*, 834-839.

94. Ding, Q.; Boyd, B. L.; Small, G. W. Determination of organic contaminants in aqueous samples by near-infrared spectroscopy. *Applied Spectroscopy* 2000, *54*, 1047-1054.

95. Morris, G. A.; Freeman, R. Selective excitation in Fourier transform nuclear magnetic resonance. *Journal of Magnetic Resonance (1969)* 1978, *29*, 433-462.

96. Kuroda, K.; Suzuki, Y.; Ishihara, Y.; Okamoto, K.; Suzuki, Y. Temperature mapping using water proton chemical shift obtained with 3D-MRSI: Feasibility in vivo. *Magnetic Resonance in Medicine* 1996, *35*, 20-29.

97. Amster, I. J. Fourier transform mass spectrometry. *Journal of Mass Spectrometry* 1996, *31*, 1325-1337.

98. Hoffman, F. An introduction to Fourier theory. *Web site available at http://aurora. phys. utk. edu. Accessed on Aug* 2004.

99. Harris, D. C.: *Quantitative Chemical Analysis*; Seventh Edition ed.; W.H Freeman and Company, 2007.

100. Griffiths, P.; De Haseth, J. A.: *Fourier Transform Infrared Spectrometry*; John Wiley & Sons, 2007; Vol. 171.

101. TYDEX. FTIR Beam Splitter. http://www.tydexoptics.com/products/spectroscopy/beamsplitter_substrates/ (accessed 03/30 2014).

102. Mertz, L. Transformations in Optics. *New York: Wiley, 1965* 1965.

103. Griffiths, P. R. *Analytical Chemistry* 1992, *64*, 868A-875A.

104. Chase, D. B. Phase correction in FT-IR. *Applied Spectroscopy* 1982, *36*, 240-244.

105. Bracewell, R. N.: *The Fourier Transform and Its Applications*; McGraw-Hill: Newyork, 2000.

106. Herres, W.; Gronholz, J. Understanding FT-IR data processing. Part 1: Data acquisition and Fourier transformation. *Instruments and Computers (Reprint)* 1984, *1*.

107. Zhao, X.; Zhao, B.: The Study of Apodization of Imaging Fourier Transform Spectrometer in the Spectrum Reconstruction. In *Advances in Computer Science and Information Engineering*; Springer; pp 677-682.

108. Michelson, A. A. *Philosophical Magazine* 1891, *31*, 338-346.

109. Perkins, W. D. Fourier transform infrared spectroscopy. Part II. Advantages of FT-IR. *Journal of Chemical Education* 1987, *64*, A269.

110. Brereton, R. G.: *Applied Chemometrics for Scientists*; Wiley. com, 2007.

111. Haaland, D. M.; Easterling, R. G.; Vopicka, D. A. Multivariate least-squares methods applied to the quantitative spectral analysis of multicomponent samples. *Applied Spectroscopy* 1985, *39*, 73-84.

112. Haaland, D. M.: Classical least squares multivariate spectral analysis. Google Patents, 2002.

113. Lavine, B. K. Chemometrics. *Analytical Chemistry* 2000, *72*, 91-98.

114. Haaland, D. M.; Melgaard, D. K. New classical least-squares/partial least-squares hybrid algorithm for spectral analyses. *Applied Spectroscopy* 2001, *55*, 1-8.

115. Haaland, D. M.; Melgaard, D. K. New augmented classical least squares methods for improved quantitative spectral analyses. *Vibrational Spectroscopy* 2002, *29*, 171-175.

116. Haaland, D. M.; Melgaard, D. K.: Augmented classical least squares multivariate spectral analysis. Google Patents, 2004.

117. Melgaard, D. K.; Haaland, D. M.; Wehlburg, C. M. Concentration residual augmented classical least squares (CRACLS): a multivariate calibration method with advantages over partial least squares. *Applied Spectroscopy* 2002, *56*, 615-624.

118. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 2001, *58*, 109-130.

119. Krutchkoff, R. Classical and inverse regression methods of calibration. *Technometrics* 1967, *9*, 425-439.

120. Höskuldsson, A. PLS regression methods. *Journal of Chemometrics* 1988, *2*, 211-228.

121. Testing, A. S. f.; International, M.: ASTM E1655. Standard Practices for Infrared Multivariate Quantitative Analysis. ASTM International West Conshohocken, PA, 2005.

122. Lavine, B. K.; Workman, J. Chemometrics. *Analytical Chemistry* 2002, *74*, 2763-2770.

123. Haaland, D. M.; Thomas, E. V. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry* 1988, *60*, 1193-1202.

124. Martens, H.: *Multivariate calibration*; John Wiley & Sons, 1991.

125. Wall, M.; Rechtsteiner, A.; Rocha, L. Singular value decomposition and principal component analysis. *A Practical Approach to Microarray Data Analysis* 2003, 91-109.

126. Jolliffe, I. T. Principal component analysis: A beginner's guide—II. Pitfalls, myths and extensions. *Weather* 1993, *48*, 246-253.

127. Madsen, R. E.; Hansen, L. K.; Winther, O. Singular value decomposition and principal component analysis. *Neural Networks* 2004, *1*, 1-5.

128. Liang, Y.; Lee, H.; Lim, S.; Lin, W.; Lee, K.; Wu, C. Proper orthogonal decomposition and its applications—Part I: Theory. *Journal of Sound and Vibration* 2002, *252*, 527-544.

129. Mujica, L.; Rodellar, J.; Fernandez, A.; Güemes, A. Q-statistic and T2-statistic PCA-based measures for damage assessment in structures. *Structural Health Monitoring* 2011, *10*, 539-553.

130. Bouhsain, Z.; Garrigues, S.; de la Guardia, M. PLS-UV spectrophotometric method for the simultaneous determination of paracetamol, acetylsalicylic acid and caffeine in pharmaceutical formulations. *Fresenius' Journal of Analytical Chemistry* 1997, *357*, 973-976.

131. Jonsson, P.; Johansson, A. I.; Gullberg, J.; Trygg, J.; Grung, B.; Marklund, S.; Sjöström, M.; Antti, H.; Moritz, T. High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. *Analytical Chemistry* 2005, *77*, 5635-5642.

132. Esteban, M.; Arino, C.; Díaz-Cruz, J. Chemometrics for the analysis of voltammetric data. *TrAC Trends in Analytical Chemistry* 2006, *25*, 86-92.

133. Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Analytica Chimica Acta* 1986, *185*, 1-17.

134. Frangos, C.; Stone, M. On jackknife, cross-validatory and classical methods of estimating a proportion with batches of different sizes. *Biometrika* 1984, *71*, 361-366.

135. Anderssen, E.; Dyrstad, K.; Westad, F.; Martens, H. Reducing over-optimism in variable selection by cross-model validation. *Chemometrics and Intelligent Laboratory Systems* 2006, *84*, 69-74.

136. Ding, Q.; Small, G. W.; Arnold, M. A. Genetic algorithm-based wavelength selection for the near-infrared determination of glucose in biological matrixes: initialization strategies and effects of spectral resolution. *Analytical Chemistry* 1998, *70*, 4472-4479.

137. Jiang, J.-H.; Berry, R. J.; Siesler, H. W.; Ozaki, Y. Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data. *Analytical Chemistry* 2002, *74*, 3555-3565.

138. González-Arjona, D.; López-Pérez, G.; González-Gallero, V.; González, A. G. Supervised pattern recognition procedures for discrimination of whiskeys from gas chromatography/mass spectrometry congener analysis. *Journal of Agricultural and Food Chemistry* 2006, *54*, 1982-1989.

139. Simon, L.; Nazmul Karim, M. Probabilistic neural networks using Bayesian decision strategies and a modified Gompertz model for growth phase classification in the batch culture of *Bacillus subtilis*. *Biochemical Engineering Journal* 2001, *7*, 41-48.

140. Marini, F.; Balestrieri, F.; Bucci, R.; Magrı, A.; Magrı, A.; Marini, D. Supervised pattern recognition to authenticate Italian extra virgin olive oil varieties. *Chemometrics and Intelligent Laboratory Systems* 2004, *73*, 85-93.

141. Massart, D.; Vandeginste, B.; Deming, S.; Michotte, Y.; Kaufman, L. Evaluation of precision and accuracy. Comparison of two procedures. *Chemometrics: a textbook. Elsevier Science, BV, Amsterdam, The Netherlands* 2003, 33-57.

142. Lee, T.; Richards, J. A. Piecewise linear classification using seniority logic committee methods, with application to remote sensing. *Pattern Recognition* 1984, *17*, 453-464.

143. Herman, G. T.; Yeung, K. D. On piecewise-linear classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 1992, *14*, 782-786.

144. Postaire, J.-G. An unsupervised Bayes classifier for normal patterns based on marginal densities analysis. *Pattern Recognition* 1982, *15*, 103-111.

145. Kroutil, R. T.; Combs, R. J.; Knapp, R. B.; Godfrey, J. In *Tilte*1996; International Society for Optics and Photonics.

146. Walters, F. Sequential simplex optimization-An update. 1999.

147. Bai, C. Noninvasive near infrared spectroscopy on living tissue with multivariate calibration approaches. University of Iowa, 2010.

148. Fraser, D. M.: *Biosensors in the Body: Continuous in vivo Monitoring*; John Wiley & Sons, 1997.

149. Wilson, G. S.; Gifford, R. Biosensors for real-time in vivo measurements. *Biosensors and Bioelectronics* 2005, *20*, 2388-2403.

150. Stout, P.; Pokela, K.; Mullins-Hirte, D.; Hoegh, T.; Hilgers, M.; Thorp, A.; Collison, M.; Glushko, T. Site-to-site variation of glucose in interstitial fluid samples and correlation to venous plasma glucose. *Clinical Chemistry* 1999, *45*, 1674-1675.

151. Arnold, M. A.; Liu, L.; Olesberg, J. T. Selectivity assessment of noninvasive glucose measurements based on analysis of multivariate calibration vectors. *Journal of Diabetes Science and Technology* 2007, *1*, 454-462.

152. Botros, L. L.; Jablonski, J.; Chang, C.; Bergana, M. M.; Wehling, P.; Harnly, J. M.; Downey, G.; Harrington, P.; Potts, A. R.; Moore, J. C. Exploring Authentic Skim and Nonfat Dry Milk Powder Variance for the Development of Nontargeted Adulterant Detection Methods Using Near-Infrared Spectroscopy and Chemometrics. *Journal of Agricultural and Food Chemistry* 2013, *61*, 9810-9818.