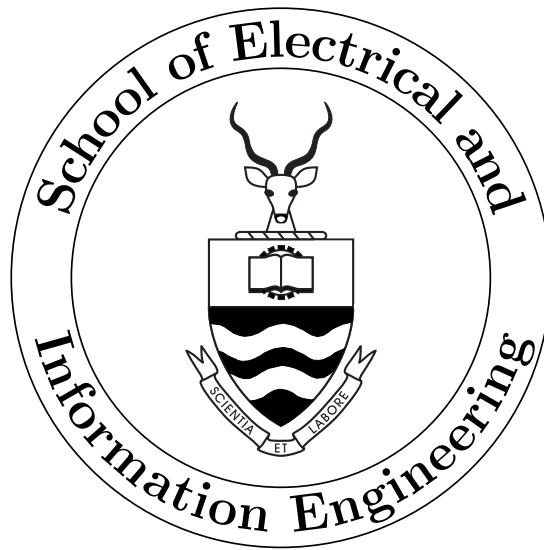


A Simulation-Based Study on the Application of Artificial Neural Networks to the NIR Spectroscopic Measurement of Blood Glucose

John David Manuell



A research report submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science in Engineering.

Johannesburg, July 2008

Declaration

I declare that this research report is my own, unaided work, except where otherwise acknowledged. It is being submitted for the degree of Master of Science in Engineering in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other university.

Signed this ___ day of _____ 20___.

John David Manuell

Abstract

Diabetes Mellitus is a major health problem which affects about 200 million people worldwide. Diabetics require their blood glucose levels to be kept within the normal range in order to prevent diabetes-related complications from occurring. Blood glucose measurement is therefore of vital importance. The current glucose measurement techniques are, however, painful, inconvenient and episodic. This document provides an investigation into the use of near-infrared spectroscopy for continuous, non-invasive measurement of blood glucose. Artificial neural networks are used for the development of multivariate calibration models which predict glucose concentrations based on the near-infrared spectral data. Simulations have been performed which make use of simulated spectral data generated from the characteristic spectra of many of the major components of human blood. The simulations show that artificial neural networks are capable of predicting the glucose concentrations of complex aqueous solutions with clinically relevant accuracy. The effect of interference, such as temperature changes, pathlength variations, measurement noise and absorption due other analytes, has been investigated and modelled. The artificial neural network calibration models are capable of providing acceptably accurate predictions in the presence of multiple forms of interference. It was found that the performance of the measurement technique can be improved through careful selection of the optical pathlength and wavelength range for the spectroscopic measurements, and by using preprocessing techniques to reduce the effect of interference. Although the simulations suggest that near-infrared spectroscopy is a promising method of blood glucose measurement, which could greatly improve the quality of life of diabetics, many further issues must be resolved before the long-term goal of developing a continuous non-invasive home glucose monitor can be achieved.

Contents

Declaration	i
Abstract	ii
Contents	iii
List of Figures	viii
List of Tables	x
List of Abbreviations	xi
1 Introduction	1
1.1 Overview of Research Report	2
2 Background	4
2.1 Diabetes Background	4
2.2 Diabetes Statistics	5
2.3 Diabetes Control Studies	7
2.4 Treatment of Diabetes	8
2.5 The Need for Continuous Monitoring	9
3 Glucose Measurement Techniques	11
3.1 Conventional Measurement Procedure	11

3.2	Fully Implanted Sensors	12
3.3	Minimally Invasive Sensors	13
3.3.1	Fluid Extraction from the Skin	13
3.3.2	Microdialysis	13
3.4	Non-invasive Sensors	14
3.4.1	Radio Wave Impedance	14
3.4.2	Polarimetry	14
3.4.3	Mid-infrared Spectroscopy	15
3.4.4	Near-infrared Spectroscopy	15
3.4.5	Raman Spectroscopy	16
4	NIR Spectroscopy and its Application to Glucose Measurement	17
4.1	Quantitative Analysis and the Laws of Absorption	18
4.2	Overview of the Measurement Technique	20
4.3	The NIR Region and Optimal Spectral Range for Glucose Measurement	21
4.4	Barriers to Continuous Blood Glucose Measurement	24
4.5	Literature Survey	25
4.5.1	General Literature Relating to Spectroscopic Glucose Measurement	26
4.5.2	Neural Networks for Spectroscopic Glucose Measurement	28
5	Data Analysis and Multivariate Calibration	30
5.1	A Brief Introduction to Neural Networks	31
5.2	Neural Networks for Multivariate Calibration of Spectral Data	31
5.3	Alternative Calibration Techniques	32
5.4	Advantages and Limitations of Artificial Neural Networks	33
5.4.1	Flexibility	33
5.4.2	Robustness	34

5.4.3	Black-box Nature of Artificial Neural Networks	34
5.5	Data Preprocessing	34
5.5.1	General Preprocessing Methods	35
5.5.2	Additional Preprocessing Techniques	35
5.6	Measurement of Performance	38
5.6.1	Standard Error of Prediction	38
5.6.2	Clarke Error Grid Analysis	39
5.6.3	Comparison of the Performance of Episodic and Continuous Meters	40
6	The Effect of Interference on Glucose Measurement in Human Blood	42
6.1	The Effect of Temperature on Glucose Measurement	43
6.1.1	The Influence of Temperature on the Water Absorption Spectrum	43
6.1.2	The Influence of Temperature on the Glucose Absorption Spectrum	46
6.2	The Effect of Interfering Analytes on Glucose Measurement	47
7	Glucose Measurement in Simulated Aqueous Solutions	50
7.1	Overview of the Simulations	51
7.2	The Simulated Aqueous Solutions	52
7.2.1	Generation of the Simulated Spectral Data	52
7.2.2	Analysis of the Spectral Data	55
7.3	Glucose Measurement in the Presence of Other Analytes	57
7.3.1	Development of the Calibration Models	57
7.3.2	Results and Discussion	57
7.4	Temperature Insensitive Glucose Measurement	59
7.4.1	Generation of the Spectral Data	60
7.4.2	Development of the Calibration Models	60
7.4.3	Results and Discussion	61

7.5	Glucose Measurement in the Presence of Random Noise	62
7.5.1	Generation of the Spectral Data	62
7.5.2	Development of the Calibration Models	63
7.5.3	Results and Discussion	64
7.5.4	Instrument Performance and RMS Noise	64
7.6	Simulations with Variable Pathlengths	65
7.6.1	Generation of the Spectral Data	66
7.6.2	Development of the Calibration Models	66
7.6.3	Results and Discussion	66
7.7	Simulations with Multiple Sources of Interference	67
7.7.1	Generation of the Spectral Data	68
7.7.2	Development of the Calibration Models	69
7.7.3	Results and Discussion	69
7.8	The Use of Data Preprocessing Techniques	72
7.8.1	Preprocessing for the Removal of Low Frequency Effects	72
7.8.2	Preprocessing for the Removal of High Frequency Noise	73
7.8.3	Preprocessing for Data Reduction	75
7.8.4	Performance of the Preprocessing Techniques	75
7.9	Findings from the Simulations	77
7.10	Limitations of Simulations using Computer-generated Spectral Data	79
8	Towards In Vivo Measurement of Blood Glucose	80
8.1	Recommendations for Future Work	80
8.2	Requirements of a Spectroscopic Glucose Monitor	82
8.3	Continuous In Vivo Glucose Measurement	83
9	Conclusion	84

References	86
A Basic Principles of Near Infrared Spectroscopy	95
A.1 Theoretical Models for IR Spectroscopy	96
A.2 Features of the Near Infrared Spectral Region	98
A.3 Measurement Modes	99
A.4 Instrumentation	100
A.4.1 Dispersive Spectrometers	100
A.4.2 Fourier Transform Spectrometers	100
B Data Handling and Processing of Spectral Data	102
B.1 Multivariate Calibration	102
B.1.1 Linear Calibration Techniques	103
B.1.2 Artificial Neural Networks	105

List of Figures

2.1	Late complications of diabetes	6
2.2	Number of people with diabetes in each continent for 2000 and 2010	6
2.3	Continuous versus episodic monitoring	9
4.1	Stages involved in the measurement of glucose concentrations	20
4.2	The infrared absorption spectrum of glucose	21
4.3	The NIR absorption spectrum of water at 37°C	22
4.4	Glucose absorptivities over combination region	23
4.5	Glucose spectrum over first overtone region	24
4.6	Molar absorptivities of glucose and water in the NIR spectral region	25
5.1	Clarke Error Grid	40
6.1	Water molar absorptivity at different temperatures	44
6.2	Difference in water molar absorptivity due to temperature changes	44
6.3	Water absorptivity changes in first overtone region due to temperature variations	45
6.4	Water absorptivity changes in combination region due to temperature variations	46
6.5	Combination region molar absorptivities of major blood components	47
6.6	First overtone region molar absorptivities of major blood components	48
7.1	Simulated spectra in combination and first overtone region	56
7.2	The effect of analyte concentrations	58

7.3	Clarke error grid for simulated data with no noise or temperature variations	59
7.4	The effect of temperature on the combination region spectrum	61
7.5	The effect of temperature on first overtone region spectrum	62
7.6	Clarke Error Grid for simulated data with temperature variations	63
7.7	Clarke Error Grid for simulated data with random noise	65
7.8	Representative 100% lines for simulated data with random noise	66
7.9	Effect of pathlength changes on the NIR spectral samples	67
7.10	Clarke Error Grid for simulated data with variable pathlengths	68
7.11	The square error vs the number of training cycles for an MLP network	69
7.12	Clarke error grid for network with multiple sources of interference	71
7.13	First overtone spectra before data preprocessing	73
7.14	Effect of various pre-processing techniques on first overtone spectral data	74
7.15	The removal of random noise with a moving average filter	75
7.16	Approximation of spectrum with six principal components	76
A.1	The electromagnetic spectrum	95
A.2	The NIR region of the electromagnetic spectrum	99
A.3	Measurement modes for NIR spectroscopy	99
B.1	PCA transformation procedure	104
B.2	A single neuron	106
B.3	Multi-layer perceptron network with one hidden layer	106
B.4	Radial basis function network	107

List of Tables

2.1	The cost of diabetes	7
5.1	Zones of the Clarke error grid	41
7.1	Concentrations of analytes used in simulated aqueous solutions	53
7.2	Glucose measurement results for simulated aqueous solutions	59
7.3	Results for simulations with temperature variations	62
7.4	Results for simulations with random noise	64
7.5	Results for simulations with random pathlength variations of $\pm 10\%$	67
7.6	Results for simulations with multiple sources of interference	70

List of Abbreviations

NIR	Near-infrared
DCCT	Diabetes Control and Complications Trial
UKPDS	United Kingdom Prospective Diabetes Study
GOD	Glucose Oxidase
FDA	Food and Drug Administration
IR	Infrared
PLS	Partial Least Squares
PCR	Principal Component Regression
ANN	Artificial Neural Network
MIR	Mid-infrared
SEP	Standard Error of Prediction
PCA	Principle Component Analysis
RBF	Radial Basis Function
MLP	Multi-layer Perceptron
PLSR	Partial Least Squares Regression
MSC	Multiplicative Scatter Correction
SEV	Standard Error of Validation

SEC	Standard Error of Calibration
EGA	Error Grid Analysis
CG-EGA	Continuous Glucose Error Grid Analysis
CGM	Continuous Glucose Monitor
R-EGA	Rate Error Grid Analysis
P-EGA	Point Error Grid Analysis
AU	Absorbance Units
RMS	Root mean square
RMSN-100%	Root mean square noise on 100% lines
NAS	Net analyte signal

Chapter 1

Introduction

Diabetes mellitus refers to several conditions that, if untreated, result in excessively high blood glucose levels. The raised blood glucose concentration can be caused by a reduction in the production of insulin or a reduced sensitivity to the action of insulin. Without adequate glucose control, diabetic patients are likely to experience severe long-term degenerative complications [1, 2].

The number of individuals affected by diabetes is growing at an alarming rate. There are currently approximately 200 million diabetics world-wide, and this figure is expected to increase to approximately 300 million by 2025 [3].

Several studies have shown that the late complications of diabetes can be delayed significantly through tight control of a diabetic patient's blood glucose levels [4]. Many diabetics make use of home glucose monitors to measure their blood glucose concentrations and inject themselves with the required doses of insulin based on the results. The current glucose monitors are episodic in nature and require the diabetic to extract a drop of blood in order for the measurements to be performed. This process can be painful and inconvenient.

A non-invasive glucose monitor would promote more frequent testing, thereby allowing tighter control of blood glucose levels and delaying the onset of the severe late complications of diabetes [5]. A continuous measurement device would encourage accurate treatment by providing information about the rate of change of blood glucose levels.

This report provides an investigation into the use of near-infrared (NIR) spectroscopy for the measurement of blood glucose levels. The measurement technique involves directing NIR radiation through a vascular region of the human body. The absorption of radiation at different wavelengths is measured. Since the absorption spectrum produced depends on the composition of the tissue that the radiation passes through, it is possible to determine the concentration of blood glucose.

The research focusses on data processing aspects and methods of extracting glucose information from spectroscopic data. Artificial neural networks (ANN's) are used for the generation

of the calibration models. Investigations are performed to determine the effects of various forms of interference on glucose measurement.

The aim of the project is determine the feasibility of using NIR spectroscopy along with artificial neural networks for the measurement of blood glucose levels. Factors which affect the measurement of glucose are analysed and methods of overcoming some of the major obstacles to development of a NIR spectroscopic glucose monitor are investigated. The research provides information which will be valuable in achieving the long term goal of developing a continuous non-invasive glucose monitor.

1.1 Overview of Research Report

Chapter 2 provides the reader with a basic understanding of diabetes by describing the causes, symptoms and complications. The alarming increase in the number of diabetics and the risk associated with uncontrolled blood glucose levels are also discussed. The method by which diabetics regulate their blood glucose levels through the use of episodic glucose monitors is mentioned, as well as the problems with these monitors. This is followed by an explanation as to why a continuous glucose monitor is required which can overcome the problems associated with these episodic measurement devices.

Chapter 3 discusses several techniques which can be used to measure blood glucose levels. It begins with a description of the enzymatic measurement process used in most episodic monitors and then describes various techniques which could be used to measure blood glucose concentrations continuously. The techniques discussed include fully-implanted sensors, minimally invasive sensors and non-invasive sensors.

Chapter 4 provides a description of near-infrared (NIR) spectroscopic glucose measurement. Background information relating to near-infrared spectroscopy and quantitative analysis is provided and an overview of the measurement procedure is given. The spectral regions which could provide the most useful spectral information are discussed and various problems which must be overcome are mentioned. A literature survey discussing previous research into spectroscopic glucose measurement is provided.

The analysis of the spectral data and the multivariate calibration process required to extract glucose information from the spectra is described in chapter 5. The chosen method of performing the calibration is with the use of neural networks. The application of neural networks to the analysis of spectral data is discussed. Alternative calibration techniques, which have been used for similar applications, are also mentioned. Data preprocessing techniques that can remove the effect of interferences, thereby improving the results achieved by the multivariate calibration models, are described.

Chapter 6 discusses how various interferences can mask the changes in the spectrum caused by variations in the glucose concentration. This complicates the process of modelling the spectra. Two of the most problematic forms of interference are those caused by changes in the concentrations of blood analytes and temperature variations. The effects of these interferences are discussed as well as the interferences due to high-frequency noise and pathlength

changes.

Chapter 7 shows the results of simulations which aim to determine the feasibility of spectroscopic glucose measurement and provide insight into the challenges and obstacles which must be overcome before a spectroscopic glucose monitor can be developed. The simulations make use of artificial neural networks to analyse simulated spectral data. The use of simulated spectral data enables initial investigations to be performed without the time-consuming and expensive task of obtaining actual spectral data using an NIR spectrometer. The computer-generated data contains many of the major components of human blood and enables the effect of interferences to be studied independently. The ability of neural networks to make accurate predictions from the spectral data of complex aqueous samples with multiple forms of interference is determined and the effects of pathlength variations and temperature changes are noted.

Chapter 8 discusses further issues which must be considered before an in vivo NIR spectroscopic glucose sensor can be developed and points out further work that is required.

Chapter 9 concludes by providing a summary of the major findings and discusses how the results obtained provide insight into the feasibility of creating a continuous glucose monitor based on the use of NIR spectroscopy and artificial neural networks.

Chapter 2

Background

2.1 Diabetes Background

Diabetes mellitus refers to a number of conditions that, in an untreated state, are characterised by excessively high blood glucose levels (hyperglycaemia). The raised blood glucose levels can be due to a reduction in insulin production, an absence of insulin production or a reduced sensitivity of the organs to the action of the insulin [1, 2].

Insulin is a hormone that is produced in the pancreas by the beta cells of the islets of Langerhans. Insulin has two important functions, to promote the entry of glucose into the liver, muscles and adipose tissue and to regulate the storage of energy in the form of glycogen, fat and protein. It therefore plays an important role in the regulation of the blood glucose levels. In order for an individual to remain healthy, it is vital that their blood glucose level remains between 3.8 and 6.7 mmol/l. Levels lower than 3 mmol/l can result in impaired brain function while glucose concentrations above 10 mmol/l exceed the renal absorption threshold leading to degenerative complications in the long-term [1].

The two most common forms of diabetes are known as type-1 and type-2 diabetes. Together these two forms of diabetes account for 99.9% of occurrences of the disease. Type-1 diabetes is responsible for approximately 10% of diabetes cases. It is the result of an autoimmune destruction of the insulin-producing beta-cells of the pancreas resulting in an absolute deficiency of insulin. The onset of Type-1 diabetes usually occurs during childhood or early adulthood. Genetic factors are thought to be the cause of the autoimmune destruction of the beta-cells, although environmental factors and viral infections may also contribute [1, 4]. The symptoms of Type-1 diabetes, which include increased thirst and urination, constant hunger, weight loss, blurred vision and extreme fatigue, usually develop over a short period of time. Untreated patients run the risk of falling into a life-threatening diabetic coma, known as diabetic ketoacidosis [4].

Type-2 diabetes is the most common form of the disease, occurring in approximately 90% of diabetic patients. It is usually diagnosed in individuals between 50 and 75 years of age. Type-2 diabetes results when the organs become less sensitive to the action of insulin (insulin resistance) or the quantity of insulin produced by the pancreatic beta-cells is insufficient [1].

Type-2 diabetes is thought to be caused by a combination of genetic factors, dietary habits and lifestyle [2]. About 80% of type-2 diabetics are overweight [4]. The symptoms of type-2 diabetes usually develop gradually. They include fatigue, nausea, frequent urination, unusual thirst, weight loss, blurred vision, frequent infections and slow healing of wounds [4].

Prolonged elevation of the blood glucose level leads to glycation of the body's proteins. This results in [2]:

- damage to small blood-vessels (micro-angiopathy),
- damage to the large blood-vessels (macro-angiopathy),
- increased rigidity of tendons, joint capsules and blood vessel walls and
- reduced elasticity of the lungs.

Long-term complications include blindness, kidney failure, coronary heart disease and impaired circulation. Figure 2.1 illustrates some of the complications which can occur if diabetes is not carefully monitored.

Diabetics that receive insulin treatment run the risk of developing hypoglycaemia (low blood glucose levels) which can at first cause confusion, and if the glucose concentration remains low for extended periods, it will result in a coma or death [1, 2].

In order to maintain their blood glucose concentration in the required range of between 3.8 and 6.7 mmol/l, type-1 diabetics need to check their blood glucose concentrations regularly and must receive insulin injections to normalise the blood glucose concentration. Type-1 diabetics are usually totally dependent on insulin for their survival. A healthy diet and physical activity are also necessary. Type-2 diabetes is initially managed through diet and oral medication, but as the disease progresses, insulin therapy may be required. Insulin injections are required by about one-third of sufferers [1, 4, 7].

2.2 Diabetes Statistics

Diabetes mellitus is a major health problem and the number of diabetics is increasing at an alarming rate. The World Health Organization (WHO) estimated that 30 million people world-wide had diabetes in 1985. By 2000, the number of diabetics had risen to 177 million and it is expected that this figure will increase to 300 million by 2025. The most dramatic increase is in the number of individuals contracting type-2 diabetes. It is estimated that 4 million people die per year due to complications related to diabetes [3]. These frightening statistics have led the International Diabetes Foundation and the World Health Organization to describe this increase in diabetes as the “Most challenging health problem of the 21st century”. The prevalence of diabetes on each continent and the dramatic increase in the number of diabetics is illustrated in figure 2.2.

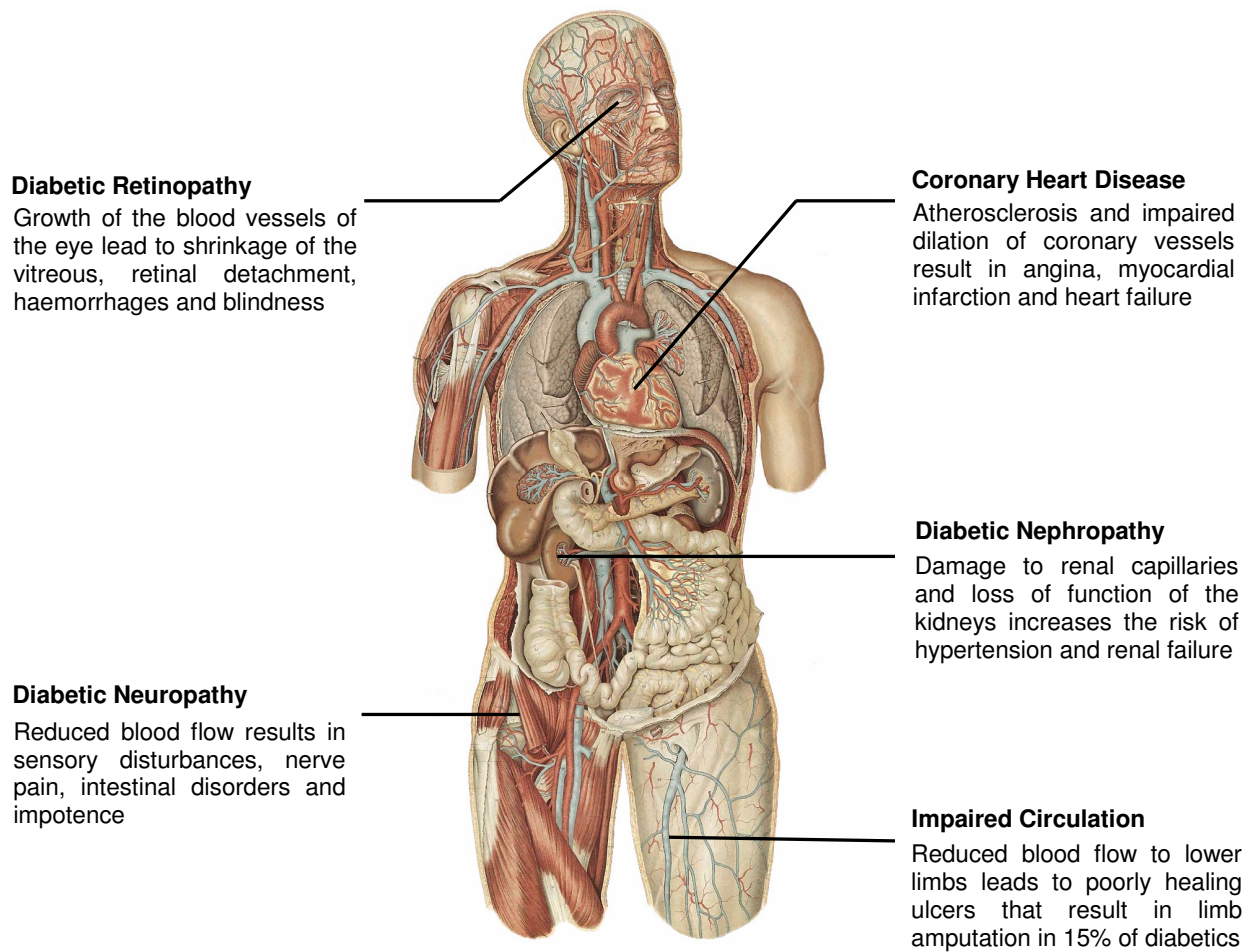


Figure 2.1: Late complications of diabetes [2]. Image obtained from [6].

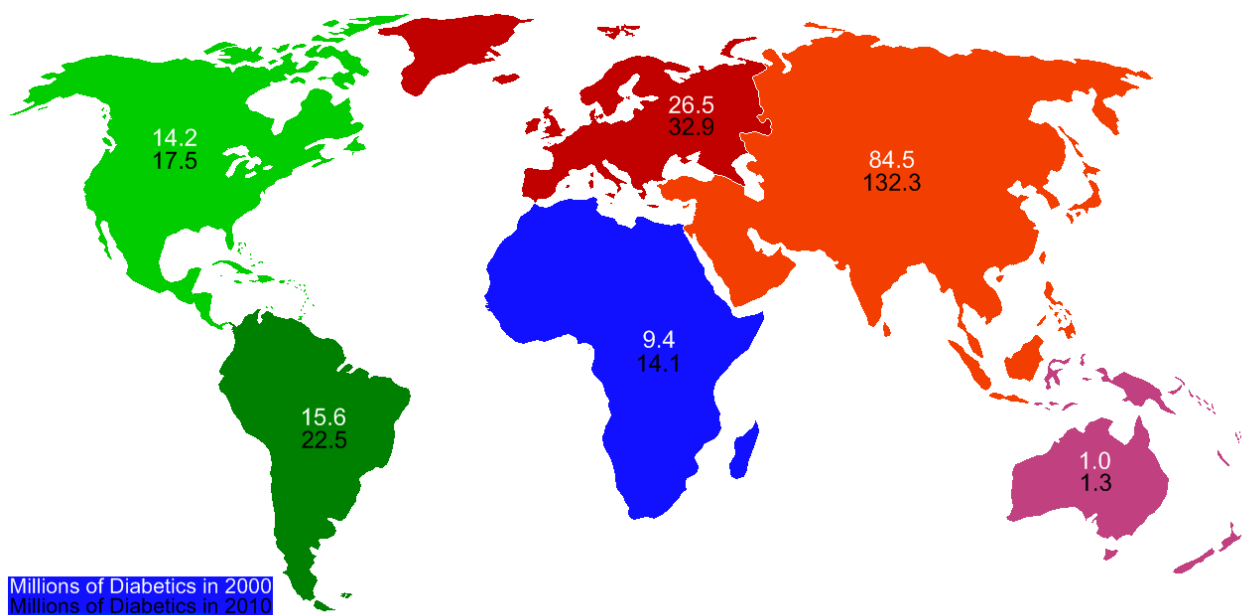


Figure 2.2: Number of people with diabetes in each continent (in millions) for 2000 and 2010 [8]. Image obtained from [9].

The reasons for this dramatic increase in diabetes are unclear. The increase in type-1 diabetes may be due to the fact that better treatment has ensured that type-1 diabetics no longer die before reaching reproductive age. This causes the genes that predispose people to diabetes to accumulate in the gene pool. The increase in the occurrence of type-2 diabetes is thought to be caused by changes in lifestyle and increased obesity due to poor dietary habits [2, 8].

Due to its chronic nature, the severity of the complications and the difficulties involved in treating these complications, diabetes is a very costly disease for both the affected individual and the healthcare authorities [3]. Some of the costs associated with diabetes are given in table 2.1.

Direct Costs	<ul style="list-style-type: none"> • Individuals and their families must pay for medical care, drugs, insulin and other supplies. • The healthcare sector is burdened by the expense of hospital services, physician services, lab tests and costs relating to the daily management of diabetes. • The direct healthcare costs of diabetes range from 2.5% to 15% of healthcare budgets depending on the prevalence of diabetes and the treatment available. • For most countries, the largest diabetes related expense is the for the treatment of patients with long-term complications.
Indirect Costs	<ul style="list-style-type: none"> • Sickness, absence from work, disability, premature retirement and premature mortality lead to a loss of productivity. • The cost relating to the loss of productivity may be as great or even greater than the direct costs.
Intangible Costs	<ul style="list-style-type: none"> • Pain, anxiety and a lower quality of life have a great impact on the lives of diabetics and their families. • Personal relationships, leisure and mobility can be negatively affected. • Self-monitoring and taking insulin injections can be time-consuming, inconvenient and painful.

Table 2.1: The cost of diabetes [3]

2.3 Diabetes Control Studies

Two Landmark clinical studies have shown the importance of tight glucose control for patients with type-1 and type-2 diabetes. The Diabetes Control and Complications Trial (DCCT) and the United Kingdom Prospective Diabetes Study (UKPDS) showed that the effective management of blood glucose levels can dramatically decrease the risk of serious complications [4].

The DCCT compared a group of individuals who had one or two insulin injections per day and monitored their blood glucose levels, with an intensive therapy group that monitored their blood glucose levels and had three or more insulin injections per day [4].

Although less than 5% of the patients in the intensive group managed to keep their glucose levels in the normal range, the intensive therapy group had significantly better control of their blood glucose levels than the conventional therapy group. The improved control delayed the onset of retinopathy, nephropathy and neuropathy by approximately 60%. The risk of macrovascular disease was also reduced [4].

The UKDPS showed intensive blood glucose control to reduce the microvascular complication rate of type-2 diabetic patients by 25% [4].

Both studies showed that tight glucose control can increase the risk of severe hypoglycaemia [4]. The two studies confirm that tight glucose control is a key aspect of diabetes management with the benefits far outweighing the risks involved.

2.4 Treatment of Diabetes

The diabetic control studies discussed in section 2.3 highly recommend frequent monitoring of blood glucose levels. In an attempt to meet the required levels of control, self-monitoring is practised by the majority of type-1 diabetics. Patients pierce their skin with a lancet to extract a drop blood, place the blood on a test strip containing chemicals sensitive to glucose and insert the test strip into a meter which displays the glucose level. The patients adjust their insulin treatment based on the results [4].

Patients can administer the insulin by injecting themselves several times a day or using insulin pumps, which supply the patient with a steady supply of glucose throughout the day [4].

Even though patients are aware of the consequences of poor glucose monitoring, studies by the American Diabetics Association have shown that only 37% of diabetics achieve the recommended level of control [4]. The average diabetic patient only tests their blood glucose level twice a day rather than the recommended 4-7 times per day [10]. Patients are reluctant to perform glucose measurement since the technique described above is invasive, painful and inconvenient [11].

A major problem with the “finger-prick” glucose measurement technique is that the glucose concentration is only known at a single point in time. Patients who use insulin therapy exclusively have fluctuating blood glucose levels. This is due to various factors including carbohydrate intake, insulin dosage, amount of activity and stress. The episodic monitors provide only a snapshot of the glucose levels at a particular moment and are not capable of determining if the glucose concentration is rising or falling. This makes it very difficult for patients to adjust their medication accurately. If a patient with a glucose level that is high but falling rapidly measures their blood glucose concentration with an episodic monitor, they will see that the reading is outside the normal range and inject themselves with insulin. This will increase the rate at which the blood glucose concentration is dropping and may lead to a dangerous hypoglycaemic episode. Since the risk of severe hypoglycaemia is two to three times higher in patients practising tight glucose control, many patients choose not to monitor their glucose levels closely due to the fear that they will trigger hypoglycaemic

events. Low glucose levels can be particularly dangerous in individuals who suffer from nocturnal hypoglycaemia and hypoglycaemia unawareness [4].

Another limitation with episodic monitors is that the readings may not identify abnormal glucose levels which occur between measurements. The illustration in figure 2.3 shows a scenario in which a patient may believe that their glucose levels are being well-controlled when the blood glucose concentrations are actually outside the recommended range at certain times.

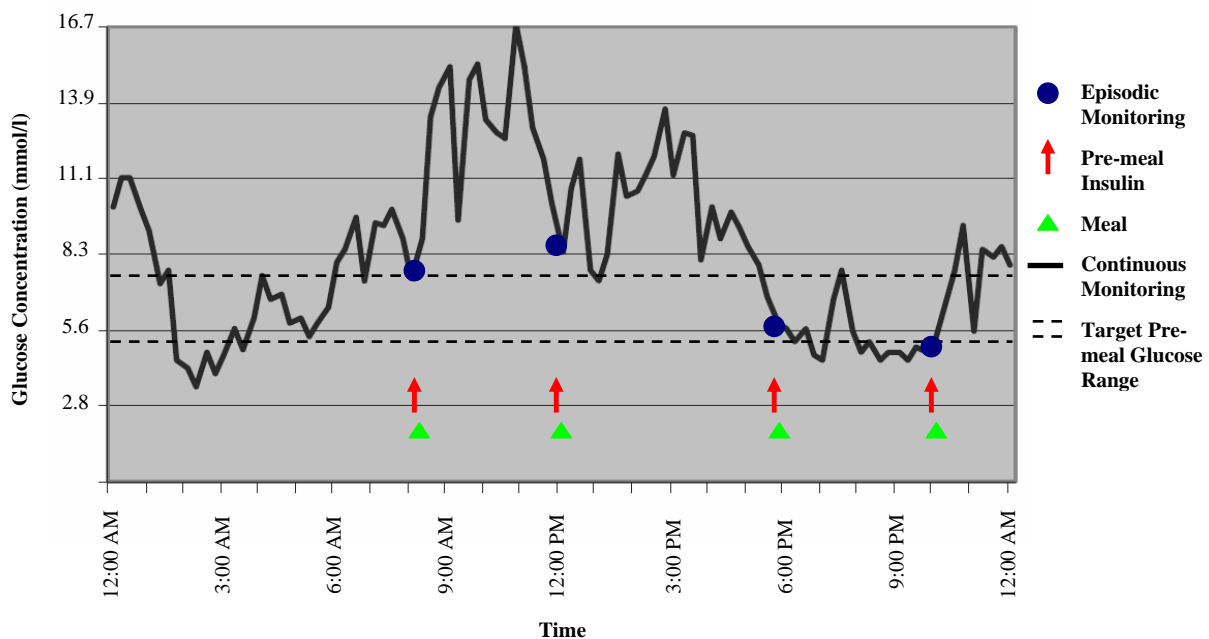


Figure 2.3: Continuous versus episodic monitoring [4]

2.5 The Need for Continuous Monitoring

There is clearly a need for a continuous glucose monitor which can operate without causing unnecessary discomfort to the patient. A non-invasive, continuous glucose monitor would improve the quality of life of diabetics by reducing stress and inconvenience, controlling glucose levels better and ultimately reducing the risk of serious diabetes-related complications.

There are several groups of individuals that would find a continuous monitor to be very useful. These groups, who would be likely to become early adopters of a continuous glucose measurement device include [4]:

- Insulin pump users that must monitor their glucose levels frequently in order to vary their insulin intake,
- People who are unable to detect hypoglycaemia due to hypoglycaemic unawareness or experience nocturnal hypoglycaemic episodes,

- Children who have difficulty understanding and responding to symptoms,
- Pregnant woman who will be highly motivated to achieve tight control in order to reduce the risk of pregnancy complications and
- Type-2 diabetics who want to make use of the device for a short period in order to make lifestyle and dietary changes in order to optimise their treatment.

The development of a continuous glucose monitor is also a vital step towards the long-term goal of developing an artificial β -cell which could autonomously measure a patient's blood glucose concentration and administer the required doses of insulin [12]. An artificial β -cell would consist of a continuous glucose monitor, which monitors the glucose levels, a control system which would calculate the required insulin doses based on the data received from the monitor and a insulin pump which would deliver the insulin to the patient. Experts predict that it will take more than a decade before a system which autonomously regulates the glucose levels is available [4].

Even though a continuous, non-invasive glucose monitor has clear benefits for diabetics, it will not become available unless it proves to be economically viable. The world-wide self-monitoring blood glucose market is estimated to be worth \$5 billion per year and this figure is expected to double within a decade. Continuous measurement devices could potentially claim a large portion of this market, which has lead many companies to invest large amounts of capital into developing replacements for the current episodic sensors. A study by the New England Health Institute, making conservative assumptions, shows that a continuous glucose monitor, that is easy-to-use and unobtrusive, would be cost-effective [4].

Chapter 3

Glucose Measurement Techniques

The sections which follow discuss several different approaches to the measurement of blood glucose. The conventional electro-enzymatic glucose measurement procedure is discussed in section 3.1. Although this approach can provide sufficiently accurate glucose readings, it has limitations in that performing the measurements are painful and inconvenient and only episodic readings can be obtained.

A large amount of research is currently being performed in order to produce a glucose measurement device which is less inconvenient to use than these conventional monitors and provides continuous glucose measurements. Many different measurement techniques have been investigated, but there are still several barriers which must be overcome before continuous glucose monitors that can replace the current episodic monitors, will be commercially available.

Continuous glucose sensors can be divided into three main categories; fully implanted sensors, minimally invasive sensors which extract interstitial fluid from the dermis or epidermis, and non-invasive techniques which generally use non-contact optical methods to measure the glucose content. Several different measurement techniques which could potentially be used for the monitoring of blood glucose are discussed in sections 3.2, 3.3 and 3.4.

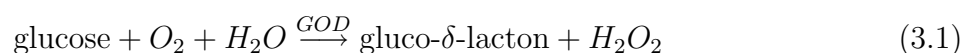
Minimally invasive techniques have shown promising results recently. Devices using this technique are likely to be the first devices to replace the episodic glucose monitors as monitors are already available which supplement the results obtained from conventional measurement devices. Non-invasive techniques, however, have advantages in terms of the frequency with which results can be obtained and they are not painful to use. These benefits are likely to make the non-invasive techniques the method of choice in the long term.

3.1 Conventional Measurement Procedure

An enzymatic electrode based on glucose oxidase was originally suggested by Clark and Lyons over 40 years ago [13]. Enzymatic glucose sensors are now the most common measurement technique for use in self-monitoring devices.

Enzymatic techniques use the natural selectivity of the enzyme glucose oxidase to achieve the required specificity within the human body. These Electrochemical measurement techniques have the advantages that they are not affected by sample colour, they require only a small quantity of blood to perform the measurements and they can easily be miniaturised [14].

The conventional procedure for the self-monitoring of blood glucose requires the patient to lance their fingertip in order to obtain a drop of blood. The blood is then placed on a test strip containing the enzyme glucose oxidase and other reagents. Glucose oxidase (GOD) acts as a catalyst in the enzymatic oxidation of glucose [1, 15].



In this reaction glucose is oxidised to gluconic acid. Glucose oxidase is an electron acceptor and is temporarily reduced to an inactive state before being reactivated by the reduction of oxygen to hydrogen peroxide [1].

Most commercial monitors measure the quantity of hydrogen peroxide produced using electrochemical or colorimetric techniques. The quantity of hydrogen peroxide produced is proportional to the glucose concentration [15, 16]. The oxygen depletion and the pH change resulting from the production of gluconic acid can also be measured in order to determine the glucose concentration [14].

3.2 Fully Implanted Sensors

Research into the development of fully implantable sensors for continuous blood glucose monitoring was first suggested in the 1960's. It is considered to be a relatively mature field of research and several devices have been partially developed [17].

Fully implanted sensors have the advantage that they are small and relatively inexpensive to produce [18]. The most common site for the implantation of glucose sensors is the subcutaneous tissue. A miniature needle is inserted directly into the tissue to monitor the glucose concentration. The subcutaneous tissue is considered to be the most appropriate location for the sensors due to its good accessibility for surgery and the relative ease with which sensors can be replaced [1]. The majority of the sensors are based on the enzymatic oxidation of glucose by glucose oxidase [1].

Several studies have been performed into the use of intravenous sensors for glucose measurement [1]. The implantation of sensors in the vascular compartment is, however, avoided by most researchers due to the risk of thrombosis, embolism and septicaemia [1].

Short-term in vivo studies have demonstrated the feasibility of using implanted needle-type glucose sensors [1]. The disadvantage of using fully implanted sensors is that biocompatibility issues, enzyme degradation and sensor drift prevent accurate readings from being attained over an extended period of time [12, 14, 18]. Biosensors using this technique for in vivo measurement seldom last for more than a few weeks.

3.3 Minimally Invasive Sensors

Minimally Invasive technologies use percutaneous sensors or needles rather than subcutaneous sampling. Fluid is extracted from the dermal layer which has many capillaries but few nerve endings. When a small needle is inserted into this dermal layer, no pain is experienced [19].

3.3.1 Fluid Extraction from the Skin

This minimally-invasive measurement technique makes use of a process known as reverse iontophoresis to extract interstitial fluid from the skin. Reverse iontophoresis involves the application of an electrical current through the skin, between an anode and a cathode, in order to extract substances from the body [12, 18]. Charged sodium ions migrate towards the cathode and uncharged molecules such as glucose are transported out of the body by electro-osmosis. The amount of glucose in the extracted interstitial fluid is proportional to the blood glucose concentration. Glucose oxidase biosensors can be used to measure the glucose concentration once the fluid has been extracted [18].

There are several problems associated with the extraction of glucose from the skin. A period of at least 20 minutes is required between the beginning of the fluid extraction process and the time at which the glucose level can be measured. This may result in fluctuations in blood glucose concentrations, which occur during this period, not being recognised. The glucose concentration in interstitial fluid is about 1000 times less than the concentration in the blood [20]. Highly accurate measurements are therefore required which increases the size and the cost of the measurement device [12]. Reverse iontophoresis can also result in skin irritation [20].

A measurement device based on this principle, known as the GlucoWatch Biographer has been approved by the United States Food and Drug Administration (FDA) for use as a supplement to a conventional glucose measurement device [15].

3.3.2 Microdialysis

Measurement of the glucose concentration of the interstitial fluid using microdialysis is one of the most promising glucose measurement techniques. Microdialysis technology attempts to simulate the action of capillaries. A catheter, containing a thin dialysis fibre, is inserted into subcutaneous fatty tissue. The fibre contains an isotonic glucose-free fluid (perfusion fluid). The catheter and the dialysis fibre have partially-permeable membranes allowing glucose to move passively from the interstitial fluid into the perfusion fluid by osmosis. The perfusion fluid is then pumped to a glucose sensor situated outside the body and the glucose concentration is measured [12].

A major problem with this technique is that there is a significant time delay between the beginning of the measurement process and the time that the glucose concentration is measured [12, 18]. The relationship between the glucose concentration in the blood and the

interstitial fluid is altered when the blood glucose levels changes rapidly [12]. This can result in inaccurate readings being obtained.

3.4 Non-invasive Sensors

The measurement techniques described above all require a probe to be inserted into the human body. Non-invasive techniques, that do not require the insertion of probes, are an active area of research. These techniques offer potential advantages for home glucose measurement as they provide painless measurements and are not affected by biocompatibility issues.

3.4.1 Radio Wave Impedance

Radio wave impedance uses the principle that when a radio wave beam is applied to an aqueous solution, non-ionic solutes attenuate the amplitude and shift the phase of the beam. This results in a change in impedance to radio wave energy proportional to the solute concentration. Since glucose is the non-ionic solute with the highest concentration, this technique can be used to determine the blood glucose levels [20].

This technique has the advantage that the majority of the components can be obtained off-the-shelf resulting in the device being relatively inexpensive. The major disadvantage is the impedance is also affected by other factors such as the concentration of electrolytes in the blood and body temperature [20].

3.4.2 Polarimetry

Polarimetry is the process of measuring the optical rotation of polarised light. The plane of polarised light rotates when it is passed through a fluid containing glucose. The rotation is proportional to the glucose concentration. This technique can be used for measuring the glucose content of the aqueous humour of the eye [20].

A beam splitter is used to divide a polarised light beam into a reference beam and a detection beam. The detection beam passes through the aqueous humour of the eye. The two beams are then compared to determine the phase shift [20].

The disadvantages of using this technique are that the signals are small and the glucose concentration of the aqueous humour may differ from that of the blood if the blood glucose levels are changing rapidly [20]. Studies performed on mammals estimate the time constant for the equilibration of blood glucose and aqueous humour concentrations to be between 20 minutes and 1 hour [21].

3.4.3 Mid-infrared Spectroscopy

Spectroscopy is an established technology used to determine composition of a body based on the electromagnetic spectrum that is reflected or absorbed by the body. Since the fundamental absorption peaks of glucose are in the mid-infrared region of the spectrum, the absorption spectrum obtained when mid-infrared radiation is passed through the human body could theoretically be used to determine blood glucose concentrations.

A major problem with the use of mid-infrared spectroscopy is the high absorbance of water which limits the pathlength which can be used to a few micrometres. Mid-infrared spectroscopy has been used with some success for glucose determination in blood and serum but has not been successfully applied to non-invasive measurement in the human body. The limited penetration depth of the radiation ensures that in vivo measurements are being performed on tissue layers near the exterior of the body which contain little or no glucose information [14].

3.4.4 Near-infrared Spectroscopy

The NIR region of the spectrum contains the first overtone and combination absorbance features as opposed to the fundamental absorbances found in the mid-infrared. The absorption of radiation in the NIR region is much weaker than that in the mid-infrared. The lower absorbances allow for pathlengths of over 1cm in aqueous samples such as the human body [14].

The aim of this glucose measurement technique is to pass NIR radiation through a vascular region of the human body and then extract information relating to the glucose concentration from the resulting spectral data [22].

The amount of near infrared radiation (NIR) absorbed by a body at each wavelength is determined by comparing a reference beam with a detection beam that has been passed through or reflected by the body. This technique has been successfully used in oximetry to determine the oxygen saturation of the blood [20].

Since the electromagnetic spectrum changes with the glucose concentration, spectroscopy can be used to measure blood glucose levels. Various measurement sites can be used including the earlobe, finger web, finger cuticle and lip mucosa [10]. Complex mathematical models are required to eliminate interferences from biological molecules, tissue structures and optical effects [18]. This optical measurement technique allows measurement of the glucose concentration of the blood directly rather than via other body fluids. Several investigators have shown that this method can potentially produce clinically useful results [11] but in vivo studies have generally produced disappointing results [20].

The major problem associated with NIR spectroscopy is the lack of selectivity for glucose [18]. Environmental factors such as body temperature, blood haemoglobin levels, skin hydration and atmospheric pressure can affect the readings obtained, resulting in the need for frequent recalibration [10, 20]. Since all individuals have unique skin and tissue properties, calibration is potentially required for each user [11].

3.4.5 Raman Spectroscopy

Raman spectroscopy is a form of vibrational spectroscopy similar to infrared (IR) spectroscopy. Raman spectroscopy often produces results which are complementary to those found using IR spectroscopy [11].

When a beam of light is focussed on a sample, photons are absorbed by the material and scattered. The majority of the scattered photons have the same wavelength as the photons in the original beam of light and are known as Rayleigh Scatter. A small portion of the scattered radiation is shifted in wavelength. These photons which undergo the shift in wavelength are known as Raman Scatter. The majority of the Raman scattered photons are shifted to longer wavelengths (Stokes Shift). The Stokes shifted Raman scattering is of interest in Raman spectroscopy [23]. The Raman spectrum is useful for the identification of molecules and could therefore be applied to the problem of measuring blood glucose levels.

A major advantage of Raman Spectroscopy over NIR Spectroscopy, is that its spectrum has distinct, pronounced peaks. This improves the specificity and lessens the effect that other metabolites have on the readings. The major problem associated with this technique is the inherent weakness of the Raman signal. The intensity of the Raman signal is about 1000 times less than the intensity of the Rayleigh scattered light used with NIR spectroscopy [23].

Chapter 4

NIR Spectroscopy and its Application to Glucose Measurement

This study investigates the use of near-infrared spectroscopy for glucose measurement. This technique is considered to be the most promising of the techniques mentioned in chapter 3, as it appears to be a feasible method of providing both continuous and non-invasive measurements with sufficient accuracy to greatly improve the quality of life of diabetic patients.

Glucose measurement using NIR spectroscopy has been an active area of research for twenty years, but researchers have not been able to produce a device with sufficient reliability and accuracy for clinical use [24]. Several major challenges must be overcome before continuous measurement devices, which can meet the approval of the healthcare organisations, can be developed.

Non-invasive near-infrared measurement devices have successfully been developed for use in the fields of pulse oximetry and bilirubinometry [25]. These devices have provided major benefits to the healthcare community due to their practicality, accuracy and speed of operation. A historical review by Severinghaus and Astrup claims that pulse oximetry is “arguably the most significant technological advance ever made in monitoring the well-being and safety of patients during anaesthesia, recovery and critical care” [26]. This demonstrates the immense potential that measurement devices based on NIR spectroscopy have for clinical monitoring. The impact of a continuous non-invasive glucose monitor could potentially be even greater than that of the pulse oximeter due to the vast number of people who are currently suffering from diabetes. NIR spectroscopic glucose measurement is extremely promising but large amounts of research are still required in order to understand the intricacies associated with the optical measurement of analytes in the human body.

The use of NIR spectroscopy for glucose measurement requires a band of infrared radiation to be passed through a vascular region of the body in order to excite vibrations in the constituent molecules. The amount of radiation absorbed at each frequency is measured, and a spectrum produced. The spectral information is analysed in order to determine the glucose concentration [27, 28].

The aim of this measurement technique is to differentiate the spectral signature of glucose

from the spectral background produced by the body tissue and other chemical components within the body. The magnitude of the glucose spectrum is dependent on the glucose concentration, thus allowing quantitative information to be obtained. A sufficiently high signal-to-noise ratio is required in order for accurate detection to occur [28, 27, 29].

It is likely that measurement devices using implanted sensors or minimally invasive techniques will be available to the public much sooner than those using the spectroscopic approach. Spectroscopic glucose measurement is, however, an extremely promising field of research and is likely to be the method of choice for glucose measurement in the long term. This form of measurement was chosen as the author considers it to be the most promising glucose monitoring technique due to several major advantages which it offers over the alternative methods of measurement.

Since spectroscopic measurement technologies are non-invasive, they offer painless measurements, rapid response times and minimal inconvenience to the user. Spectroscopic techniques do not require any surgical implantations to be performed. This enables them to overcome some major problems faced by invasive methods associated with the body's response to foreign objects or sensor deterioration over time. The use of non-invasive monitors greatly reduces the risk of infection and the sensors could potentially be used for an extended period of time without a reduction in performance.

This chapter provides an overview of the proposed glucose measurement technique. Important issues relating to the measurement of glucose are discussed and potential problems which must be overcome are mentioned. A literature survey is provided which describes the findings of other researchers who have attempted to measure glucose spectroscopically. A brief introduction to spectroscopy and a discussion of some basic spectroscopic principles are given in Appendix A. Background information relating to the data handling and signal processing aspects is provided in Appendix B. The author suggests that individuals who are not familiar with spectroscopic techniques should read these sections before continuing with the discussion of this proposed glucose measurement technique.

4.1 Quantitative Analysis and the Laws of Absorption

In order for useful analytical information to be obtained from the spectral data of a solution, it is necessary to relate the information from the absorption spectra to the concentration of the analyte. The relationship which enables quantitative information to be attained from absorption spectra is known as the Bouguer-Beer-Lambert Law, commonly called Beer's Law.

According to Beer's Law, absorption of a single compound within a homogeneous medium, is given by [30, 31]:

$$\Delta A(f) = \Delta C \epsilon(f) l \quad (4.1)$$

where: $\Delta A(f)$ is the change in the absorbance at frequency f
 ΔC is the change in analyte concentration
 $\epsilon(f)$ is the absorptivity of the analyte at frequency f
 l is the path length of light through the medium

According to equation 4.1, if the path length is constant, the change in the absorption at a certain frequency will be directly proportional to the change in the glucose concentration. The product of the concentration and the absorptivity of an analyte is known as the extinction coefficient (K) [32].

$$K = \epsilon C \quad (4.2)$$

In a solution containing a mixture of n absorbing compounds, the total absorbance at a specific frequency is the sum of the extinction coefficients of the each of the absorbing compounds multiplied by the pathlength, l .

$$\begin{aligned} A &= (K_1 + K_2 + \dots + K_n)l \\ &= (\epsilon C_1 + \epsilon C_2 + \dots + \epsilon C_n)l \end{aligned} \quad (4.3)$$

Direct measurement of the absorbance (A) of a medium is not possible using a spectrometer. The absorbance is calculated from the change in intensity of radiation as it passes through the medium. When spectroscopic measurements are performed, the intensity of the incident radiation (I_0) and the intensity of the radiation after passing through the sample (I) are measured. The ratio of these quantities is known as transmittance (T). Absorbance (A) is a dimensionless quantity, often stated in Absorption Units (AU), which can be related to the transmittance and the intensity of the radiation by equation 4.4 [32].

$$A = \log\left(\frac{I_0}{I}\right) = \log\left(\frac{1}{T}\right) \quad (4.4)$$

Beer's Law describes an ideal straight-line relationship between analyte concentrations which cannot be achieved in practice as it relies on the use of a monochromatic light source. Non-linearities in the relationship between concentration and absorbance occur due to the finite bandwidth of the light source, effects of stray radiation and scattering and unusual characteristics of certain absorbing compounds [32, 31]. It does, however, provide insight into the effect of analyte concentrations on the NIR spectrum and can provide acceptable results under controlled conditions. Under conditions where deviations from Beer's Law occur, but the law of additivity (equation 4.3) is obeyed, acceptable calibration models can be produced with the use of sophisticated calibration techniques such as Partial Least Squares (PLS), Principal Component Regression (PCR) and Artificial Neural Networks (ANN's) [30].

4.2 Overview of the Measurement Technique

The NIR glucose measurement procedure involves passing NIR radiation through a vascular region of the human body and using the transmitted spectral information to determine the blood glucose content.

The major stages and components of the measurement procedure are shown in figure 4.1. An NIR spectrometer is required to obtain the transmission spectrum of the human tissue. The operation of NIR spectrometers is briefly described in section A.4. The spectrometer's source generates infrared radiation and an interferometer or monochromator ensures that radiation of the required frequency is focussed on the sample. The NIR radiation which is transmitted through the sample is received by the detector. An NIR transmission spectrum is generated from the transmission values detected at many different frequencies. Several regions of the body are potential sites for NIR glucose measurement including the tongue, finger-webbing, cheek, lip, ear-lobe and nasal septum. The choice of measurement site is a vital consideration for in vivo glucose measurements as the spectral quality is strongly dependent on sample thickness [33]. Due to the absorption of water, RMS noise increases exponentially as a function of sample thickness. Shorter pathlengths, however, limit the measurement sensitivity. The choice of pathlength therefore requires a compromise between low RMS noise and high sensitivity [33].

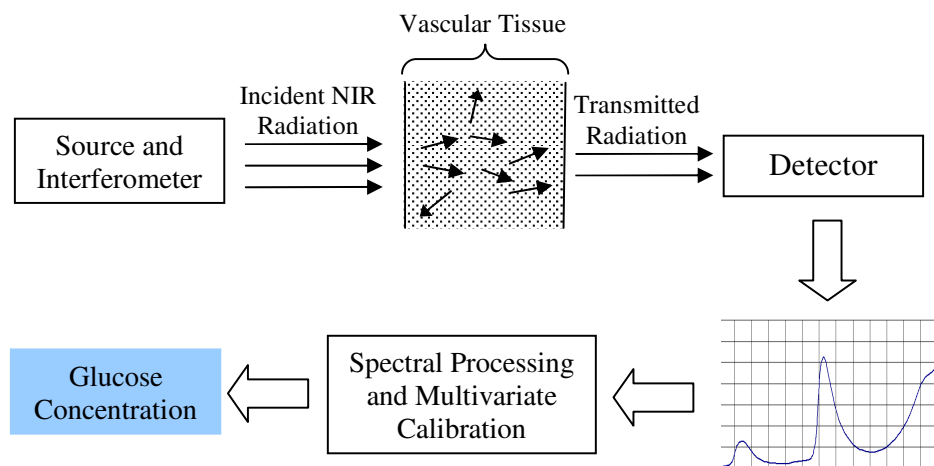


Figure 4.1: Stages involved in the measurement of glucose concentrations

The near-infrared spectrum received by the detector contains information about all the components in the optical path. The spectral information is therefore dominated by the absorption of the major components of human tissue. These include water, protein and fats. The glucose concentrations in human blood are relatively low compared to those of many other components. Sophisticated multivariate calibration techniques are therefore required to differentiate the effect that changes in glucose concentration have on the spectral data from changes caused by variations in the concentrations of other analytes. Processing of the spectral data to remove the effects of scattering, changes in pathlength and temperature fluctuations, is also required.

4.3 The NIR Region and the Optimal Spectral Range for Glucose Measurement

In order to obtain useful quantitative information, it is necessary to understand the characteristics of the NIR spectral region, the glucose spectrum and the properties of interfering compounds. This enables regions of the spectrum to be chosen in which sufficient glucose information is present and the effect of interference can be minimised.

The spectral signature of glucose shows strong absorption of radiation in the mid-infrared region due to the fundamental bending and stretching modes of C-H, N-H and O-H bonds [34]. The mid-infrared region has sharp absorption bands and offers highly specific information about the molecular make-up of solutions [16]. The mid-infrared spectrum of glucose is shown below.

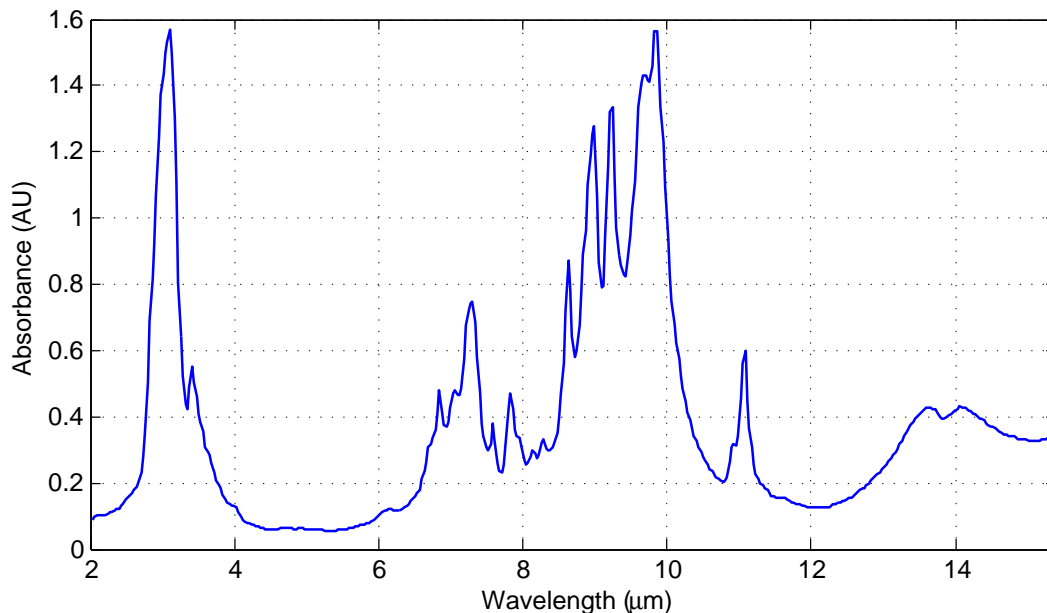


Figure 4.2: The infrared absorption spectrum of glucose. Data obtained from [35]

Despite the specificity of the mid-infrared region, it is not applicable for in vivo glucose measurement [16, 27, 29]. The absorption of water and other blood and tissue components is very high which severely limits the penetration depth of radiation. A pathlength of 100 μm or less is therefore required which is difficult to attain in the human body. The dynamic range required to measure the sharp peaks is also difficult to achieve [16].

In contrast to the mid-infrared (MIR) spectrum, NIR radiation passes relatively easily through water and body tissues allowing moderate pathlengths, in the millimetre to centimetre range, to be used [16]. NIR analysis can be performed with no sample preparation and no reagents are required [36].

The NIR region is considered to best balance the need for absorption band strength and light penetration depth required to measure the changes in glucose concentration [31]. Due

to the high concentrations of water in biological tissues, the relatively strong absorptivity of O-H groups and the broad nature of the absorption features, the Near-infrared spectrum of biological tissues is dominated by O-H stretching vibrations caused by the first, second and third overtones of the vibrational transitions of water [29, 14]. Water is therefore the primary interference in NIR spectroscopy.

The water absorption spectrum contains several intense peaks, in which low light throughput occurs resulting in a very low signal-to-noise ratio. Between these intense peaks, areas of lower absorption are found [14]. Each of these spectral windows is a potential site for spectroscopic glucose measurement, as sufficient light is transmitted through the sample. The absorption spectrum of water in the NIR region is shown in figure 4.3.

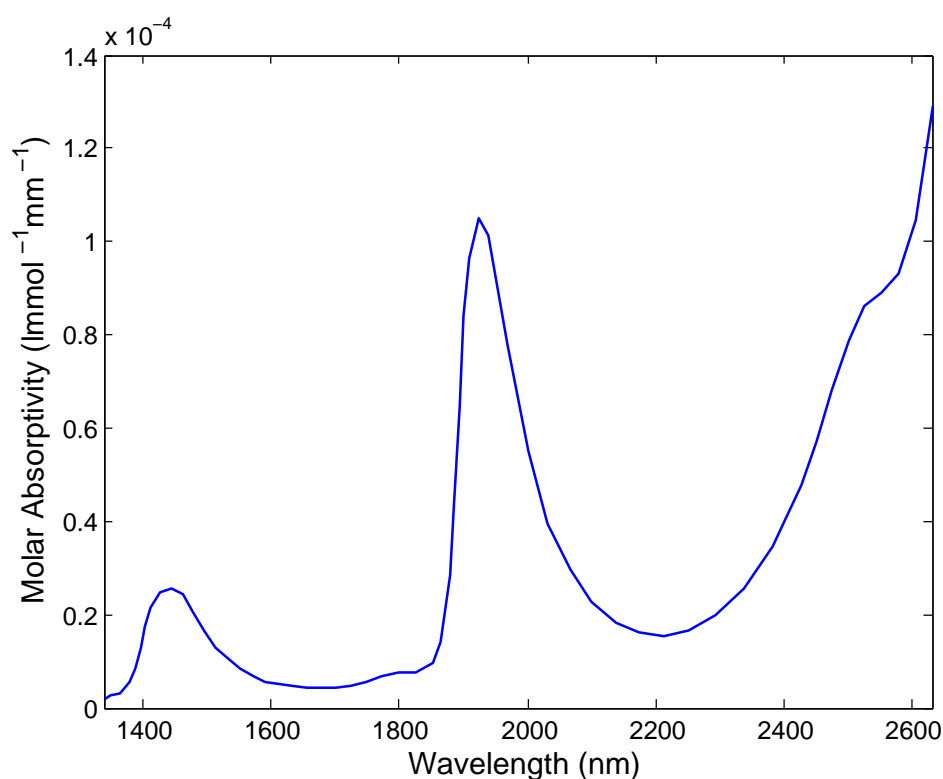


Figure 4.3: The NIR absorption spectrum of water at 37°C. Data obtained from [37]

NIR spectroscopic glucose measurement requires a spectral region in which glucose characteristic peaks are present and the water absorption is not excessively high. The Near-infrared (NIR) region contains three areas which meet these requirements [29]:

- The combination region: 2.0 - 2.5 μm (5000 - 4000 cm^{-1})
- The first overtone region: 1.54 - 1.82 μm (6500 - 5500 cm^{-1})
- The short-wavelength near infrared region: 0.7 - 1.33 μm (14286 - 7500 cm^{-1})

The glucose absorption bands in the short-wavelength NIR region are centred at 0.76, 0.92, and 1.00 μm . Measurement of these characteristic peaks is difficult due to their extremely

low absorptivities [29]. This region is, therefore, seldom used for quantitative glucose measurement.

The other two spectral regions are more promising for glucose measurement. The region between 2.0 and 2.5 μm known as the combination region is formed by the combination of stretching and bending modes. The characteristic peaks are centred at 2.10, 2.27, and 2.32 μm [29]. This region has relatively strong absorption peaks and the light penetration is sufficient to allow for pathlengths of several millimetres [14].

The region between 1.5 and 1.8 μm results from the first overtone of C-H stretching modes and is known as the first overtone region. This spectral region has absorption bands centred at 1.560, 1.695, and 1.770 μm [38]. The glucose absorption peaks are weaker than those in the combination region but due to the lower water absorption, longer pathlengths, of up to a few centimetres, can be used. The spectral features in the first overtone region are broader than those in the combination region [14]. The glucose spectra of these regions is shown in figure 4.4 and 4.5.

Using data from more than one spectral region could potentially provide improved spectral information but is not practical due to the different pathlengths required for the spectroscopic measurements in each region.

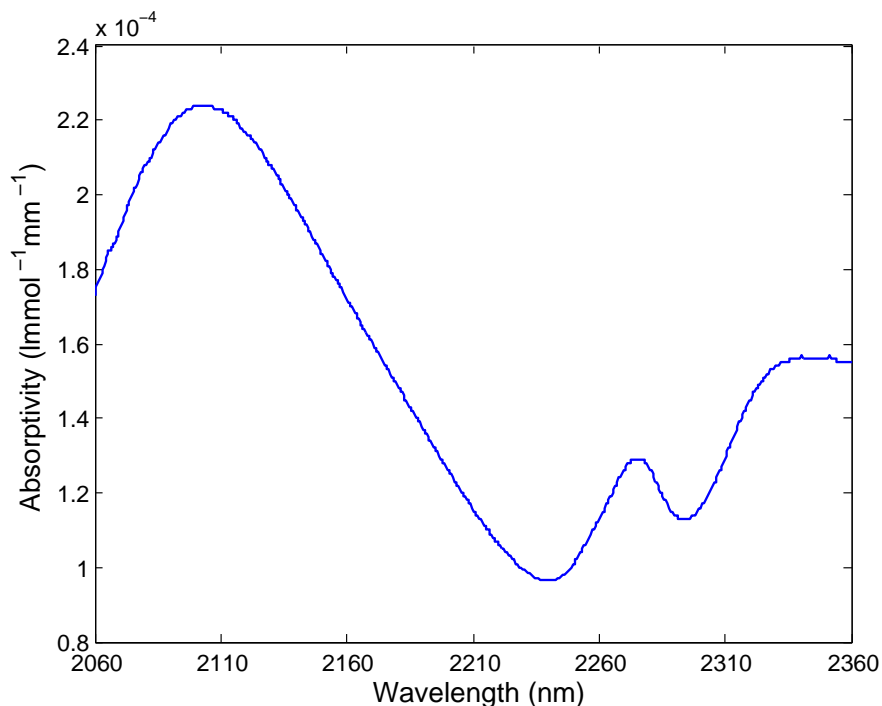


Figure 4.4: Glucose absorptivities over combination region. Data obtained from [38]

Figure 4.6 provides a comparison of the glucose absorption spectrum and the NIR spectrum of water. The figure shows that the glucose spectral features in the combination and first overtone regions fall within “water transmission windows” where the absorption of radiation due to water molecules is fairly weak. The investigations performed into the measurement of blood glucose will make use of one or both of these regions in order to determine glucose concentrations. It is not clear which of these regions will provide the most reliable glucose

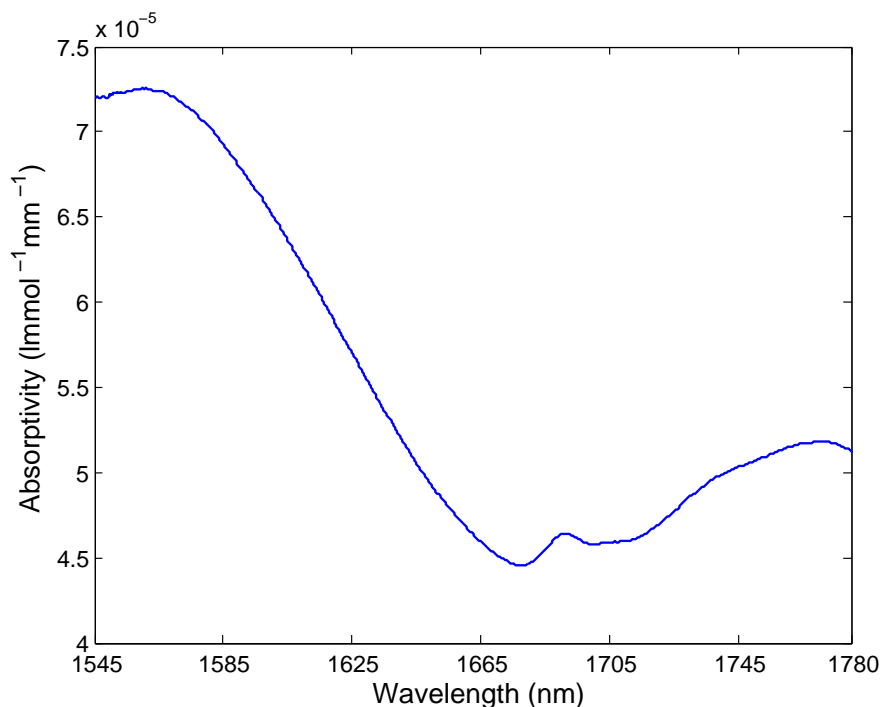


Figure 4.5: Glucose spectrum over first overtone region. Data obtained from [38]

information as the quality of the information attained will depend on many factors including the pathlength, the measurement site and interferences from water, protein, fat and other analytes.

Within the combination and first overtone regions, absorption bands for fat, proteins and various other substances will provide interfering spectral information, thus complicating the task of determining the glucose concentration [29, 34, 27, 39]. The glucose spectral peaks result from vibrations of C-H and O-H bonds. Since these bonds are present in the majority of biological molecules, spectral peaks from other blood constituents overlap with the glucose absorption features [14].

4.4 Barriers to Continuous Blood Glucose Measurement

Several challenges must be overcome before NIR spectroscopy can be effectively used for the detection of analytes. Some of the major challenges relating to the use of spectroscopy for glucose measurement are given below.

- Spectral variations due to changes in glucose concentration are extremely small compared to those from other biological components [24].
- The skin is chemically and morphologically complex and scatters light strongly, resulting in difficulties in determining the path of light travelling through the tissue [31].

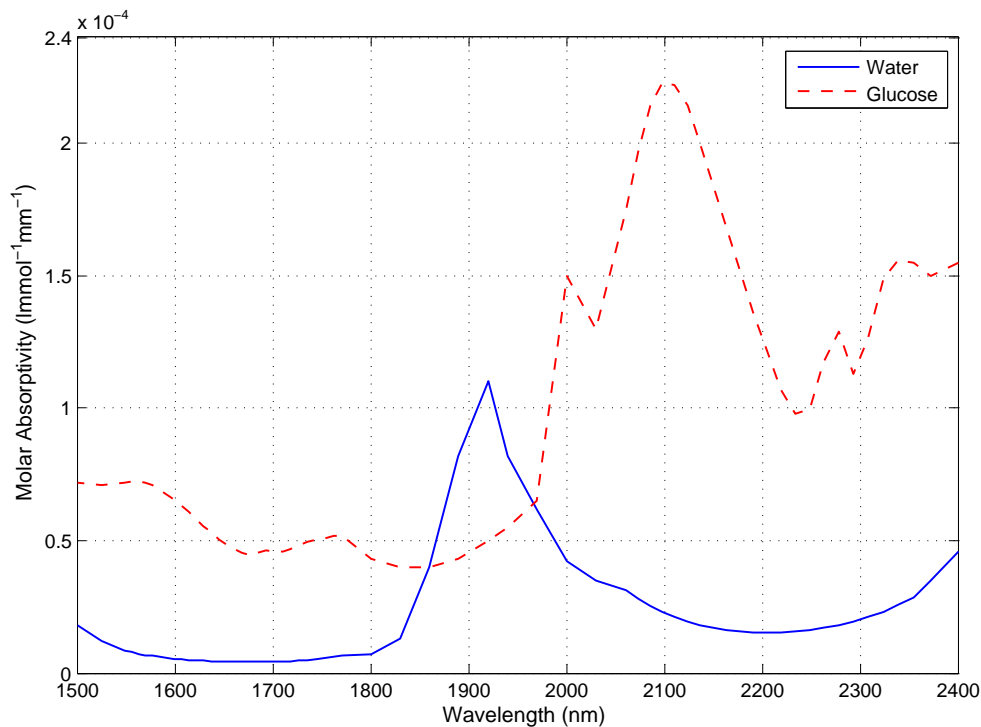


Figure 4.6: The molar absorptivities of glucose and water in the near infrared spectral region. Data obtained from [38, 37].

- The distribution of analytes in human tissue is not uniform. For example, the glucose, water and collagen concentrations differ in blood vessels, interstitial fluid and skin layers [31].
- The primary NIR absorbers in tissue are water, proteins and fat. These substances are present in far greater quantities than glucose and have spectral signatures which overlap with the glucose absorption bands [27]. The spectra of many analytes with low concentrations in blood also overlap with the glucose spectral features.
- The NIR spectrum for water is sensitive to temperature variations. The glucose absorption bands therefore lie on top of a temperature sensitive background [27].
- Pathlength variations and sensor drift result in variations in the NIR spectrum of blood which could be incorrectly interpreted as changes in analyte concentrations.

4.5 Literature Survey

A large amount of previous work has been performed that relates to the use of near infrared spectroscopy for the measurement of blood glucose levels. None of the studies have successfully managed to produce a glucose monitoring device which can overcome all the problems associated with performing in-vivo spectroscopic measurements and attain sufficient accuracy to be used by diabetic patients. Many of the studies have however successfully tackled

several major aspects relating to the measurement of glucose through tissue, and have shown the use of NIR spectroscopic measurements to be a viable method of non-invasive glucose measurement, which could have massive benefits for diabetic patients. A brief description of the techniques employed by previous researchers and the results which were attained, are given below. General research relating to spectroscopic glucose measurement is provided followed by a discussion of research involving the use of artificial neural networks. The majority of the literature relates to the use of partial least squares regression or principal component regression to perform the multivariate calibration. Only a limited amount of research has been performed which makes use of ANN's to analyse NIR spectroscopic data for the prediction of glucose concentrations.

4.5.1 General Literature Relating to Spectroscopic Glucose Measurement

One of the earlier attempts to measure glucose using near-infrared spectroscopy was performed by Dull and Giangiacomo in 1984 [40]. They showed that the concentration of glucose in water can be quantified by NIR spectroscopy but encountered difficulties at low concentrations due to the strong absorption of water. A 50 fold improvement would be required for use in blood glucose measurement.

A large amount of research has been performed by Arnold and co-workers at the University of Iowa. Arnold has questioned the validity of studies claiming to have performed highly accurate in vivo measurements [28, 41]. He suggests that fundamental research using matrices of known composition is required to verify the validity of these claims and argues that issues such as the most suitable spectral range and the influence of interferences can be studied most effectively in an environment in which variables can be examined in a controlled and systematic manner [28]. A large portion of their research therefore focuses on the measurement of glucose in simplified aqueous solutions [42, 43, 44, 45].

Arnold and Small performed measurements in a simplified aqueous solution in the combination region of the spectrum ($2\mu\text{m} - 2.5\mu\text{m}$) [42]. They used dynamic area calculations and baseline correction to provide an integrated area that is linearly related to glucose concentration. Digital filtering was used to remove high frequency noise and low frequency baseline variations. A predicted error of 7.8% was achieved [42].

Several other papers have been published by this research group which attempt to measure glucose in a protein matrix [43], an aqueous solution of protein and triglycerides [44], plasma [45] and human serum [46]. The studies analyse the combination region of the spectrum using Gaussian-shaped digital filters and partial least squares regression. The standard error of prediction (SEP) attained in the protein matrix was 0.24 mmol/l whereas the SEP in the human serum samples was 1.29 mmol/l. Marquardt et al. showed that multivariate techniques are far superior to univariate calibration techniques [43].

A Ph.D. thesis by Hazen shows the results from research performed into glucose measurement in the combination and first overtone regions of the NIR spectrum using Fourier filtering and PLS regression [14]. Simulations were performed in water, serum, whole blood and the

human body. Preprocessing of spectral data using Fourier filtering was found to improve the performance of PLS calibration models in the presence of large baseline deviations and high-frequency noise and over narrow spectral ranges. The use of Fourier filtering had little effect when spectral regions containing multiple glucose absorption bands were analysed.

Ham et al. used Partial Least Squares (PLS) and Principle Component Analysis (PCA) to analyse spectral data from measurements performed in blood serum in the wavelength range 870 nm-1098 nm [47]. The standard error of prediction was 1.64 mmol/l and 1.60 mmol/l using PLS and PCA respectively. Ham also performed measurements in the wavelength range 1.48 μm to 2.5 μm [48]. An accuracy of 0.73 mmol/l was achieved when a frequency-warping procedure was used to reduce the number of spectral components. Time-domain digital butterworth filtering was used for the pre-preprocessing of the data and PLS regression was used to produce the calibration model. Ham achieved a standard error of prediction of 0.585 mmol/l in aqueous solution using time-domain filtering and PLS regression [49].

Tarumi performed in vitro measurements in the 1200-1800nm range and investigated the effects of temperature changes and scattering [50]. His research shows that the effect of temperature changes and scattering on the spectral data is more than 25 times lower when multivariate calibration techniques are used than when univariate techniques are used. The error induced by scattering and temperature changes was found to be less than 1.1mmol/l when multivariate techniques were used.

Youcef-Toumi and Saptari designed and built a custom modular Fourier transform NIR spectrometer for the purpose of glucose measurement [51, 52]. They found the combination region of the spectrum to provide better results than the first overtone region. They achieved an accuracy of 0.5 mmol/l in an aqueous solution using PLS regression. Fourier filtering was used to compensate for temperature variations.

Liu et al. investigated the effect of chance correlations on in vivo and in vitro NIR glucose measurements [24]. They state that although several researchers have obtained satisfactory prediction results using multivariate calibration techniques, none of them have been able to show that the results are based on glucose specific spectral information rather than accidental time-related conditions with the experiments. Reasons for these chance correlations, such as signal drift, interference and physiological factors were investigated and methods to avoid these correlations were investigated. These methods included random sampling and background spectrum calibration. The researchers found that chance correlations did not contribute to the multivariate model of glucose concentration.

Burmeister et al. studied the suitability of various measurement sites for first overtone NIR transmission spectroscopy [53]. The cheek, lower lip, upper lip, nasal septum, tongue and webbing tissue between the thumb and forefinger were examined. The physical and chemical properties of the sites were evaluated. The tongue was found to be the most suitable measurement site since it had a suitable pathlength and the least fat of the measurement sites. They found that the highest signal-to-noise ratios could be obtained when the percentage of fat tissue is low. Burmeister went on to perform in vivo measurements across the tongues of five type-1 diabetic patients [54]. The standard error of prediction was 3.4 mmol/l for the best calibration model. Burmeister determined the glucose-specific information is available in the first overtone region but significant improvements are needed before clinically useful

measurements are possible.

Brown et al. used NIR spectroscopy as a diabetes screening technique [55]. In vivo reflectance spectroscopy in the wavelength range $1.25\ \mu\text{m}$ to $2.5\ \mu\text{m}$ was performed on the forearm of patients. PLS regression was used to develop the calibration model. Brown found the results obtained to be comparable to those attained using fasting plasma glucose tests. The study showed that the largest absorption in the overtone region was due to water, collagen and lipids.

Du and co-workers studied the first overtone NIR spectra of the skin [56]. They showed that region orthogonal signal correction can be used for the removal of the interference signals caused by water. PLS regression was used and a SEP of $0.883\ \text{mmol/l}$ was obtained.

Hopkins and Mauze measured the diffuse reflectance from the forearm in the $1\ \mu\text{m}$ - $2.5\ \mu\text{m}$ range [57]. 19 non-diabetic patients were involved in the study and PCA was used. The authors stated that the calibration model had no predictive value due to the presence of large amounts of noise.

Malin et al. produced statistically valid calibration models for 3 out of the 7 diabetic patients involved in a glucose measurement study using PLS regression [58]. Reflectance measurements were performed on the forearm in the wavelength range $1050\ \text{nm}$ - $2450\ \text{nm}$. A SEP of $1.41\ \text{mmol/l}$ was obtained.

Maruo et al. performed in vivo measurements in dermis tissue using NIR reflectance spectroscopy. The light path in tissue was modelled using a Monte Carlo method. PLS regression was used and the SEP was $1.79\ \text{mmol/l}$.

Robinson et al. experimented with transmission measurements through the finger in the $870\text{-}1300\text{nm}$ range [59]. The spectral data was analysed using PLS regression and principal component regression. The authors claimed that an absolute error of $1.1\ \text{mmol/l}$ was attained.

Heise et al. derived a pulsatile blood spectrum from diffuse reflectance spectra of oral mucosa by Fourier analysis [60, 61]. Pulsatile spectroscopy involves the analysis of the change in light transmission during an arterial pulse. This allows the highly complex non-pulsatile characteristics of tissue to be eliminated. Heise et al. found that special spectral variable selection based on choices made from the optimal PLS calibration models achieves better results than using the full spectrum to obtain the calibration models. The wavelength range of $1.1\ \mu\text{m}$ - $1.8\ \mu\text{m}$ was used and an SEP of $2.1\ \text{mmol/l}$ was obtained.

4.5.2 Neural Networks for Spectroscopic Glucose Measurement

The quantity of literature available, which relates to the use of artificial neural networks for the measurement of blood glucose, is limited. This is surprising since those researchers who have made use of ANN's for multivariate analysis have found their performance to be comparable and in some instances better than that of partial least squares regression and principal component analysis.

One of the most frequently cited studies using ANN's for glucose measurement was performed by Jagemann et al. [62]. Diffuse reflectance spectroscopy was used to obtain spectral readings in the 900 nm to 1200 nm wavelength range from the middle finger of 14 patients. Multivariate calibration was performed using PLS and radial basis function (RBF) neural networks. The RBF calibration models outperformed PLS models for 11 out of 14 test patients.

Bhandare et al. performed glucose measurements in simulated serum solutions using Mid-IR spectroscopy to investigate the effect of spectral interferences [63]. The performance of three multivariate techniques, PLS, PCR and ANN's, were compared to a univariate method. The PLS technique was found to have the lowest standard error of prediction (0.939 mmol/l). The Standard errors of prediction for ANN and PCR were both 1.04 mmol/l whereas the error for the univariate technique was 2.228 mmol/l. Another paper by Bhandare and co-workers found that a method combining PLS and neural networks could produce a smaller SEP than PLS or PCR when measuring the glucose concentration of whole blood samples using spectral data between 6.67 and 13.33 μm [64].

Ham and co-workers have shown that ANN's are capable of detecting glucose concentrations in aqueous solutions from Mid-IR spectral information. One of their studies made use of a two-layer feedforward architecture while another used a hybrid ANN containing a multi-layer perceptron (MLP) and a counter-propagation network [65, 66].

Lin et al. applied PLSR, MLP networks and RBF networks to the analysis of the spectral data of glucose solutions [67]. The PLSR and MLP models performed adequately with a small number of samples but the performance degraded when the sample size became too large. The RBF method was the only method with which the performance improved as the sample size increased.

Fischbacher et al. performed diffuse reflection measurements in the 850 nm - 1350 nm range on the finger of diabetic patients [68]. They showed that the prediction error of RBF networks can be decreased significantly through the use of outlier detection techniques.

A paper by Bhandare and Mendelson makes use of theoretical near-infrared spectra of an aqueous solution containing glucose, urea, albumin and histidine [69]. Twenty-four equally spaced wavelengths were selected in the 1015 nm - 2487 nm range. The researchers found that an MLP neural network could accurately predict the glucose concentration when the concentrations of the other analytes were varied. The introduction of random spectral noise was found to decrease the predictive ability of the network. The prediction error of the network was considered to be acceptable with random spectral noise corresponding to a 1.33 mmol/l glucose change.

Chapter 5

Data Analysis and Multivariate Calibration

A modelling process is required to determine the relationship between the spectral information received from the NIR spectrometer and the corresponding glucose concentrations. In order to be clinically useful, the calibration model must be insensitive to baseline variations that can be orders of magnitude larger than the glucose absorption bands, and sufficiently robust to reject interferences caused by changes in temperature, pathlength and concentrations of other analytes [70].

The simplest method of performing the calibration is to consider the spectral information at a single frequency and monitor how it is affected by glucose concentration changes. Several researchers have shown that this univariate approach is not sufficient for complex problems [43, 71]. This can be explained by the fact that several factors other than changes in glucose concentration, including temperature, pathlength and changes in the concentration of other analytes, can cause changes in the absorbance at a single frequency. The selectivity is therefore unacceptably low. Multivariate calibration techniques, which consider the spectral information available at a range of frequencies have better selectivity and are thus better suited for tasks, such as the measurement of blood glucose levels, which require measurements in the presence of many interferences. Partial least squares regression and principal component analysis are commonly used by researchers for the analysis of spectral data. Neural networks are favoured in this paper as they are a promising calibration technique, which have provided good results in a number of chemometric applications, but have not frequently been used for the purpose of determining blood glucose levels. Further research into the use of neural networks is therefore considered to be necessary. Neural networks have several advantages over linear modelling techniques such as PLSR and PCR when modelling non-linear data [72].

This chapter provides a short introduction to neural networks and discusses their application to the analysis of spectral data. Alternative multivariate modelling techniques are briefly mentioned. The benefits of using data preprocessing techniques to reduce the effects of various forms of interference are described and methods of measuring the performance of glucose calibration models are discussed. More detailed background information relating to

multivariate calibration techniques is available in Appendix B.

5.1 A Brief Introduction to Neural Networks

Artificial neural networks are a form of artificial intelligence roughly based on the operation of neurons in the human brain. Neural networks are capable of generating non-linear relationships between the inputs and outputs to a system. They are able to operate without a priori knowledge of the form of the model. Neural networks make use of a non-centralised form of information storage in which information is distributed among network nodes and the strength of the connections between the nodes [72].

Network parameters are determined using an iterative training procedure. The parameters are initially given random values which are adjusted during a training process which makes use of a training data set containing input values and the corresponding outputs. The training process is an optimisation problem in which the error between the actual network output and the target network output is minimised by adjusting the network parameters. The network is unlikely to find an absolute minimum in the multi-dimensional problem space but will locate a local minimum with an acceptably low error for the problem considered [72].

Neural networks are universal approximators as they are able to model any continuous function in a domain defined by bounded inputs to an arbitrary degree of accuracy [72].

The most commonly used forms of neural networks are multi-layer perceptron (MLP) and radial basis function (RBF) networks. A discussion on the structure and operation of these networks is provided in Appendix B.

5.2 Neural Networks for Multivariate Calibration of Spectral Data

The applicability of Artificial Neural Networks as a modelling tool for multivariate calibration is well established, leading to ANN's being widely used in many applications especially in the field of chemometrics. The wide variety of applications in which ANN's are used for multivariate calibration are illustrated in papers by Brown et al. [73] and Despagne and Massart [72].

ANN's can create an empirical multivariate calibration model which relates instrument measurements to a property of a target analyte [72]. In the case of the spectroscopic measurement of glucose, the model will relate the absorptivities at various frequencies to the glucose concentration.

The ANN builds a calibration model of the form $Y = F(X) + \epsilon$ known as an inverse calibration model, where X is the matrix of the spectral measurements, Y is the target glucose concentration and ϵ is the residual [72].

Even though the potential benefits of using ANN's for the purpose of multivariate calibration are well known, only a limited amount of research has been performed into the possible use of neural networks for the determination of glucose concentrations of aqueous solutions. The majority of researchers have favoured the use of linear modelling techniques such as PLS and PCR. The work that has been performed into the use of neural networks for the analysis of spectral data has produced promising results.

Since neural networks perform non-linear modelling, they are particularly well suited to applications in which non-linearities are present in the data set [27, 72]. Non-linearities between the spectral data and the quantitative information of interest are not easily accommodated by PLS and PCR [71].

Non-linearities are expected when spectroscopic measurements are performed on blood or tissue samples. Deviations from the Beer-Lambert Law, which relates absorbance of species in a solution to analyte concentration, are likely to occur due to the high absorbance of certain constituents, non-homogeneity of the samples and interferences due to overlapping spectra of different analytes [72]. A non-linear detector response, scattering and the presence of stray light can also introduce non-linear system behaviour. Chemical factors, such as changes in temperature or solvent composition, can also induce non-linearities resulting in shifting and broadening of the absorption bands [72]. Non-linear effects can be removed in certain cases through the use of data preprocessing techniques. Preprocessing techniques do however have limitations and can result in detrimental effects such as a reduction in the signal-to-noise ratio or the introduction of non-linearity in the wavelength space [72].

Before quantitative results can be gained from NIR spectral data, it is necessary to undergo a training process in order to generate the multivariate calibration model. The first step involved in the calibration process is to select a representative calibration sample set for which the spectral data and the corresponding glucose concentrations are known. The neural network training process is then used to determine the relationship between the spectral variations and the glucose levels. Once this has been performed, a validation process is used ensure that the model is capable of making acceptable predictions when unseen data is applied to the calibration model.

5.3 Alternative Calibration Techniques

In order to gain a better understanding of the advantages and disadvantages associated with the use of neural networks, it is necessary to consider the alternative methods which can be applied to multivariate calibration.

Linear methods, such as Partial Least Squares, Principal Component Regression and Multiple Linear Regression (MLR), are the most commonly used calibration techniques. These techniques generate the components that are used for modelling from linear combinations of the original variables. Like ANN's, they make use of a least squares criterion for minimisation of the error. If a linear data set is modelled using an ANN with linear transfer functions, it will converge to a MLR solution. The major difference between ANN's and MLR is in method by which parameters are estimated. ANN's use an iterative optimisation

process while MLR uses a matrix inversion [72]. Neural networks can also represent a PLS or PCR model if linear transfer functions are used. PLS and PCR, however, take into account constraints such as scores (PLS and PCR), orthogonality (PLS and PCR), maximisation of X-data variance (PCR) or X-Y covariance (PLS) during the parameter optimisation process. These restrictions aren't present in the optimisation process used by ANN's [72]. The linear methods can be used to model non-linear data, if suitable transformations can be taken to linearise the data or if higher order terms are used in the regression equation. Transformations, however, require a priori information which is not always available and higher order terms introduce irrelevant information to the model [72].

Several non-linear modelling techniques have been developed to overcome some of the problems associated with use of linear methods to model non-linear data. Non-linear forms of PCR and PLS such as polynomial PCR and quadratic PLS assume a simple relationship between the response modelled and the components which is not always a valid assumption. Local Weighted Regression (LWR) creates a global non-linear model based on local linear PLS or PCR models. LWR has been found to perform well in multivariate calibration but the local model parameters are less stable since they are generated from a reduced data set [72]. Other less commonly used non-linear techniques include alternating conditional expectations, smooth multiple additive regression technique, classification and regression trees, multivariate adaptive regression splines and spline PLS. These techniques perform well on non-linear data but are more computationally expensive than linear techniques and, like ANN's, they can be prone to overfitting [72].

5.4 Advantages and Limitations of Artificial Neural Networks

5.4.1 Flexibility

The flexibility of neural networks is a major advantage over many of the other non-linear modelling techniques. Neural networks do not require the assumption of a hard model before the commencement of the modelling process. This is particularly useful when using NIR data as the hard models are particularly difficult to generate due to the significant overlap in the spectra of different analytes. A priori information about non-linear effects, which may occur in practical experiments, are very difficult to model. ANN's avoid the time-consuming and difficult task of hard model identification [72].

A drawback to the flexibility of neural networks is the tendency to overfit the calibration data resulting in an inability to generalise when new data is applied to the system [72]. Neural networks can perform poorly in situations where extrapolation is required. Their ability to extrapolate is often worse than that of PLS and other linear modelling techniques [72].

5.4.2 Robustness

The distribution of information among several weights and nodes ensures that neural networks are robust with respect to random noise in the input data. Neural networks are capable of obtaining acceptable results in the presence of significant amounts of noise and their performance degrades gradually when the noise levels in the training data are increased [72, 74]. The performance of neural networks is generally significantly better than most other modelling techniques in noisy environments. This property should make neural networks particularly suitable for vivo measurements as the presence of noise and perturbations such as temperature effects are expected [72].

The gradual degradation in performance of neural networks can be attributed to the signal-averaging effect that occurs due to the summations at the network nodes and the non-localised storage of information resulting from the high interconnectivity between nodes [72, 75]. Neural networks automatically incorporate redundant nodes which improves their fault tolerance [75].

PLS and PCR accommodate non-linearities by including higher order terms, causing them to perform badly in noisy environments [72].

5.4.3 Black-box Nature of Artificial Neural Networks

The performance of neural networks is comparable and in many cases better than that of other modelling techniques. A major criticism of ANN's is that model interpretation is significantly more difficult than with PCA or PLS [72]. This is due to the summations and application of the transfer function that occur at each network node. These operations prevent the derivation of simple mathematical expressions which can relate the input and output variables [72].

5.5 Data Preprocessing

Data preprocessing is required in order to transform the spectral data received from the spectrometer into a form which will maximise the performance of the neural network calibration models.

Several general preprocessing steps, that ensure that the data is in a form which will enable the ANN's to perform adequately, are discussed in section 5.5.1. Additional preprocessing techniques, which aim to remove the effect of interferences from the spectroscopic data, are also discussed. Reducing the effect of interferences before an attempt is made to generate the calibration model simplifies the modelling process and can lead to improved predictive ability.

5.5.1 General Preprocessing Methods

Several general data preprocessing techniques are required in most applications which make use of neural networks. These techniques ensure that the data is in a form which will allow the neural network training algorithm to obtain acceptable results.

The detection of outliers is an important prerequisite to the neural network training procedure. Outliers may be introduced due to errors during the process of obtaining the spectral data or transcription errors. Removal of this erroneous data is necessary as a few data samples with major differences to the other samples can have a significant influence on the parameter estimation during the modelling process, and can lead to the generation of a poor calibration model [72].

The simplest method of outlier detection involves a visual inspection of the spectral data to determine if any data samples vary vastly from the expected values. Several more advanced outlier detection techniques are discussed in [72]. Data that has been identified as an outlier should not be used during the training process.

Data partitioning divides the data into training, validation and testing subsets. The training subset is used for parameter estimation, the validation subset is used to detect the generalisation ability of the model to prevent overtraining and the testing subset is used once the training process is complete to measure the performance of the network when unseen data is used. Each subset should be independent of the other subsets and must be representative of the the entire data set.

Data scaling is required to ensure that the training begins in the active range of the non-linear activation functions and to prevent large variations in the weights during initial training [72].

5.5.2 Additional Preprocessing Techniques

Due to inherent problems associated with multivariate calibration using spectroscopic data, additional preprocessing techniques are often required in order for acceptable results to be obtained. These problems include baseline and low frequency variations, high frequency noise and multiplicative effects.

Through the use of effective data preprocessing techniques, it is possible to create neural networks which are more rugged in terms of their ability to make accurate predictions in environments where variations in the spectral data could occur due to minor inaccuracies in the calibration of the measurement equipment, baseline variations due to temperature changes, changes in pathlength and the scattering of light. Preprocessing techniques are also used to correct for non-linear effects due to a non-ideal detector response and the presence of stray light [72]. Several techniques which can be used to improve the predictive ability of calibration models for spectral data are discussed below.

Several researchers have attempted to reduce the effect of baseline variations by calculating the first or second derivative of the spectral data and then using the derivative data as the input to the multivariate calibration algorithm [76, 36, 77, 78]. Shen et al. state that

the use of derivative spectra aids in the elimination of baseline and slope fluctuations and can narrow and enhance spectral features [77]. Small et al. suggest that the performance improvements are due to the fact that the process of obtaining the derivative is essentially high-pass filtering as high frequencies are amplified while low frequencies are suppressed [45]. The disadvantage of generating calibration models using derivative spectra is that a high signal-to-noise ratio is essential as obtaining derivatives enhances spectral noise [76]. Wang et al. claim that using derivative spectra only results in minor improvements in performance of the calibration model in most cases [77].

Many filtering techniques have been used in order to improve the calibration models that can be generated from spectral information. Filtering the data before multivariate calibration is performed is useful for suppressing low-frequency variations, such as temperature effects and pathlength changes, and for attenuating high frequency noise. Both time and frequency domain filtering techniques have been successfully applied. Heise et al. have used of time-domain butterworth filtering [61] while several other researchers have made use of digital Fourier filters [45, 79, 80, 43, 52]. The filtering of the spectral data relies on the fact that undesired spectral features, such as those caused by temperature variations, occur at much lower frequencies than the spectral features of glucose, while measurement noise occurs at higher frequencies. The effect of these interferences, which are likely to adversely affect the predictive ability of the calibration model, can therefore be removed through the use of band-pass filters.

The use of Fourier filtering involves an orthogonal transformation of the spectral information into a sum of sine and cosine shaped spectral contributions of different frequencies [81]. The transformation of data into the Fourier domain enables the components of the spectrum at different frequencies to be analysed. Windowing functions are used to attenuate undesired frequency components.

Spectroscopic measurements include both multiplicative and additive interferences that must be compensated for. The additive interferences include the spectral signals caused by analytes other than glucose, whereas multiplicative interferences are caused by changes in pathlength and light scattering. Linear modelling techniques are effective at handling the additive effects but are not able to model the multiplicative effects accurately [81]. Since ANN's are non-linear modelling techniques, they are better suited to handling multiplicative interferences but the preprocessing of the spectral data in order to remove these multiplicative effects can still improve the performance of the calibration models. Two techniques which are commonly used to remove multiplicative effects are normalisation by closure and multiplicative scatter correction (MSC).

Normalisation by closure is a simple normalisation technique that can be used for the removal of undesired scale variations. For a given spectrum, i , the absorbance value ($x_{ik_{inp}}$) at frequency k is divided by the sum of the absorbance values at all the measured frequencies to give a relative output x_{ik} . This can be expressed mathematically as [81]:

$$x_{ik} = x_{ik_{inp}} / \sum_{m=1}^K x_{im_{inp}} \quad (5.1)$$

where K is the number of frequencies at which measurements are taken and $x_{im_{inp}}$ is the absorbance value at an individual measured frequency.

Multiplicative scatter correction was originally developed in order to correct for light scattering variations in reflectance spectroscopy, a variation which has a strong multiplicative component. MSC has developed into a general technique which is widely used for separating multiplicative variations from the additive (chemical) information [81]. MSC relies on the property that the regression of spectral values against the mean spectral values for a group of samples is usually approximately linear [82]. Adjusting the slope and offset of the sample to that of the average spectrum enables the chemical of information to be preserved while other differences between spectra are reduced [83].

MSC fits each spectrum to the average of the group of spectra using least squares [83].

$$x_i = a_i + b_i m_j + e_i \quad (5.2)$$

where,

x_i is a data point from individual spectrum, i ,

m_j is the mean spectrum of the group,

e_i is the residual which should contain the chemical information,

a_i and b_i are coefficients determined during the data fitting process.

The corrected spectrum, $x_{i_{msc}}$, is calculated as follows [83]:

$$x_{i_{msc}} = \frac{x_i - a_i}{b_i} \quad (5.3)$$

Spectroscopic measurements generate large amounts of data. Reducing the size of the input data set can be beneficial in that it can reduce training time and eliminate redundant or irrelevant information. It can also lead to better generalisation and better performance in the presence of noise [72].

The most popular method of data reduction in chemometrics is Principal Component Analysis (PCA). PCA can summarise the majority of the variance of a large data set in a few orthogonal principal components. It is discussed in greater detail in section B.1.1. A potential disadvantage is that since PCA is a linear projection method, it can fail to preserve the structure of non-linear data sets [72].

Smoothing is used to remove high-frequency ripple noise, rather than systematic variations. Various filtering techniques can be used to remove these high frequency variations. One of the simplest smoothing techniques is the moving average filter. The reading x_{ik} at each value of k is replaced by a weighted average of x_{ik} and its neighbours from $k - D$ to $k + D$ [81]:

$$x_{ik} = \sum_{d=-D}^{+D} u_d \cdot x_{i,k+d} \quad (5.4)$$

The convolution weights, u_d , determine the amount of smoothing. For example, $u_d' = (0, 0, 0, 1, 0, 0, 0)$ would offer no smoothing while $u_d' = (0, 1, 2, 4, 2, 1, 0)$ would offer a weighted sum of x_{ik} and the surrounding values [81].

5.6 Measurement of Performance

An adequate method of measuring performance is essential in order to generate acceptable calibration models, to compare a specific glucose monitor to other similar devices and to determine if the calibration technique can provide sufficient accuracy to be clinically useful. The two methods used in this report for the purpose of monitoring the performance of models are the standard error of prediction and Clarke error grid analysis.

5.6.1 Standard Error of Prediction

Analysis of the standard error of prediction, validation and calibration is frequently used for the monitoring of the predictive ability of calibration models in a large variety of applications. The majority of papers discussing the use of multivariate calibration techniques for glucose measurement quote the accuracy of calibration models in terms of the standard error of prediction. This measure of prediction is also extensively used in this report to calculate the accuracy of the glucose measurements.

The standard error of prediction (SEP), the standard error of validation (SEV) and the standard error of calibration (SEC) are defined as follows [41, 84]:

$$SEP = \sqrt{\frac{1}{N_T} \sum_{i=1}^{N_T} (C_{a_{test}} - C_{p_{test}})^2} \quad (5.5)$$

$$SEV = \sqrt{\frac{1}{N_V} \sum_{i=1}^{N_V} (C_{a_{val}} - C_{p_{val}})^2} \quad (5.6)$$

$$SEC = \sqrt{\frac{1}{N_C} \sum_{i=1}^{N_C} (C_{a_{train}} - C_{p_{train}})^2} \quad (5.7)$$

where N_T, N_V and N_C correspond to the number of samples in the testing set, validation set and training set respectively. $C_{a_{test}}$ and $C_{p_{test}}$ are the actual and predicted glucose concentrations for the testing data set, $C_{a_{val}}$ and $C_{p_{val}}$ are the actual and predicted glucose concentrations for the validation data set and $C_{a_{train}}$ and $C_{p_{train}}$ are the actual and predicted glucose concentrations for the training data set.

During the optimisation process, the ANN training algorithm aims to minimise the SEC. The SEV is required to prevent overtraining. During the initial stages of training, the SEC and SEV will both decrease. With time, the network will begin to overfit the training data resulting in the SEV increasing. This will lead to poor generalisation of the network and therefore the training process must be terminated before this occurs. The SEP is the error obtained when unseen input data is supplied to the network and gives an accurate indication of the actual predictive ability of the network.

5.6.2 Clarke Error Grid Analysis

General measures of performance, such as SEP, determine accuracy in ways which are not necessarily clinically useful for the determination of glucose in blood samples. For example, a measurement device with a small SEP over the whole measurement range may produce inaccurate results in certain ranges which could lead to incorrect and potentially dangerous treatment [62, 85]. Clarke et al. claim that the accuracy measurements used by most researchers make it difficult to evaluate the clinical significance of using a particular measurement technique and to compare the performance of different measurement devices [85].

Techniques such as SEP are unable to relate the measured accuracy to the clinical interpretation of the information. The most crucial aspect of the monitoring of blood glucose is that the patient receives the correct treatment, rather than the accuracy of the measurement. For example, a 100% deviation from an actual blood glucose level of 1.5 mmol/l would still result in an identification that the patient is hypoglycaemic and the correct treatment would be provided. A 30% deviation from an actual blood glucose reading of 5 mmol/l may however result in inappropriate treatment which could potentially have dangerous consequences.

Clarke et al. recognised these problems with the accuracy measurement techniques used for the determination of blood glucose levels and developed a system which considers both the difference between the measured and target glucose levels and the clinical significance of this difference [85]. They developed a system of performance measurement known as Clarke error grid analysis (EGA), which is able to provide more appropriate accuracy measurements and is not dependent on the design of the monitor.

The Clarke error grid, shown in figure 5.1, defines a set of co-ordinates with the x-axis as the reference blood glucose level and the y-axis as the value obtained from the glucose monitor. The diagonal represents the ideal case in which the actual and predicted glucose values are equal [85].

Clarke error grid analysis is based on four assumptions used in clinical centres [85]:

- The target blood glucose range is between 3.9 mmol/l and 10 mmol/l.
- Patients will attempt to correct glucose levels above and below this target range, but not those within the target range.
- Corrective treatment which results in glucose levels outside the target range is inappropriate.

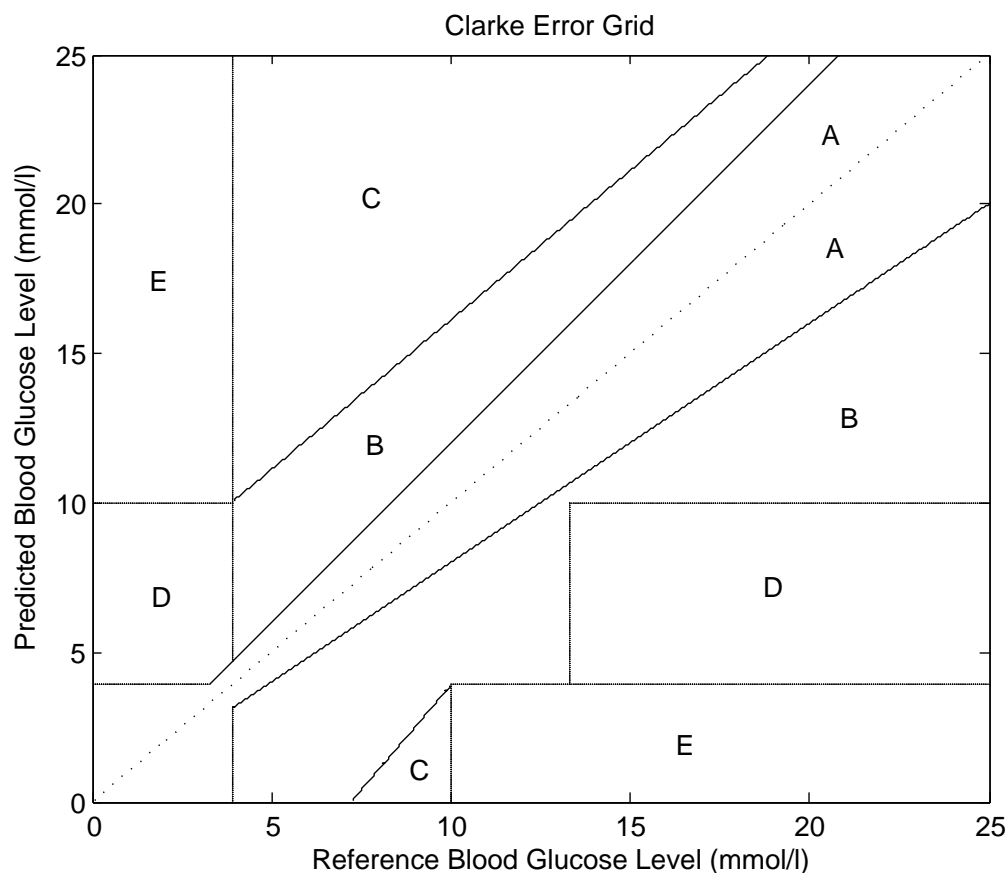


Figure 5.1: Clarke Error Grid. Adapted from [85], [86]

- Failure to treat blood glucose levels below 3.9 mmol/l or above 13.3 mmol/l is inappropriate.

In order to apply these four assumptions, the grid is divided into five regions. Each region represents a different degree of accuracy in the estimation of blood glucose levels. Values in zone A and zone B are considered to be clinically acceptable while those in zones C, D and E are clinically significant errors as they are potentially dangerous. The significance of each zone is given in table 5.6.2 [85].

5.6.3 Comparison of the Performance of Episodic and Continuous Meters

When Clarke EGA is used for the measurement of performance of traditional blood glucose monitors, 95% to 99% of the measurements fall within clinically accurate A region of the grid. The accuracy of continuous monitors is considerably lower with 58% to 70% of the readings falling within the clinically accurate range [86].

Lower accuracy than that attained by conventional glucose monitors is acceptable for continuous measurement as continuous monitors provide information about the direction and rate

Zone	Degree of Accuracy
A	Represents values that differ from the reference by less than 20% or fall in the hypoglycaemic range (<3.9mmol/l) when the reference value is <3.9mmol/l. Readings which fall in this range are clinically accurate as they would lead to the correct treatment.
B	Represents values that are more than 20% from the reference value but would lead to benign treatment or no treatment.
C	Represents readings that would result in overcorrection, which could lead to the blood glucose levels falling outside the acceptable range.
D	Represents measurements for which the actual glucose levels are outside the target range but the measured values are within the target range. This could be dangerous as unacceptable blood glucose levels will not be detected.
E	Represents cases in which the treatment is opposite to that which is desired. This erroneous measurement could result in severe consequences.

Table 5.1: Zones of the Clarke error grid [85], [86].

of changes of glucose levels which is not provided by episodic devices. A direct comparison of the performance of glucose is therefore not a meaningful method of measuring performance [86].

Knowledge of the rate at which the glucose levels are rising or falling will enable more accurate treatment decisions to be made and will allow corrective action to be taken more rapidly if incorrect decisions are made. This ensures that the potential for incorrect measurement results harming the patient is reduced. For example, if a patients glucose level are high, but falling sharply, a patient using an episodic glucose monitor would see that the blood glucose concentration was above the required range and inject themselves with insulin. This could lead to dangerous hypoglycaemic episodes. A continuous monitor, which may not be as accurate as an episodic monitor, would detect that the glucose level was falling and would therefore suggest a more appropriate course of action.

Clarke has developed a technique for measuring the performance of continuous glucose monitors called continuous glucose error grid analysis (CG-EGA) which focuses on the clinical implications of inaccuracies in readings from a continuous glucose monitor (CGM) [86]. This model looks at two important aspects of CGM, rate accuracy and point accuracy. Rate error grid analysis (R-EGA) plots the sensor blood glucose rate against the reference blood glucose rate on a grid containing five zones which represent the clinical meaning of a certain result. Point error grid analysis (P-EGA) is based on a similar principle as EGA. It provides a plot of the reference blood glucose level against the sensor blood glucose level at a particular moment in time. A detailed discussion of CG-EGA is provided in [86].

CG-EGA is not used for the determination of the predictive ability of measurement systems in this study since even though the aim of the project is to investigate continuous glucose measurement, the experiments performed in this study are episodic in nature. The fact that the interpretation of the results for continuous glucose measurement systems differs from that for episodic devices will however be considered when analysis of the experimental results is performed.

Chapter 6

The Effect of Interference on Glucose Measurement in Human Blood

Before an attempt can be made to detect glucose concentrations in human blood samples, it is necessary to have a thorough understanding of the components present in blood and how they affect the near infrared spectrum. An understanding of the effects that variations in temperature is also of vital importance.

Scattering and molecular absorption are responsible for the attenuation of NIR radiation as it passes through human tissue. Absorption dominates in the combination region where water, fat and protein are the major absorbers of radiation. The effect of scattering is more significant in the first overtone spectral range than the combination region with the absorbance of water and fat also being important [53]. Many other analytes in blood also provide interfering spectral peaks which overlap with the glucose spectral information.

Temperature changes have a significant effect on the NIR spectrum of blood. During in vivo measurements it is not possible to maintain a constant temperature. A spectroscopic glucose monitor therefore requires methods of compensating for temperature variations so that changes in temperature are not incorrectly detected as changes in glucose concentrations.

Light scattering occurs due to differences in refractive indices of elements such as erythrocytes and leukocytes and the surrounding plasma. The exact nature of the scattering depends on several factors including size, shape, orientation and refractive index of the scattering bodies and is dependent on the wavelength of the NIR radiation [87]. The scattering is dependent on the glucose concentration. This is thought to be due to an increased refractive index mismatch as the glucose concentration is increased or a reduction in the effective size of red blood cells leading to greater packing density [87]. Scattering introduces multiplicative effects which the calibration model must compensate for in order to obtain meaningful results.

The effect of high frequency noise resulting from inaccuracies in the measurement process will also interfere with the glucose measurement. A low signal-to-noise ratio will prevent accurate predictions from being made. Pathlength variations will introduce multiplicative interferences which must be removed by the calibration model.

The sections which follow discuss the effect of temperature variations and changes in analyte concentrations in more detail.

6.1 The Effect of Temperature on Glucose Measurement

As mentioned in chapter 4, the effect of variations in temperature on the NIR spectrum of water is significant and could lead to major inaccuracies in the measured glucose concentration. Jensen et al. state that a 0.1°C change in temperature causes a modification to the NIR spectrum which is comparable to a 5.55mmol/l change in glucose concentration [37]. Effective compensation for temperature variations is therefore crucial if sufficiently accurate measurements are to be attained.

Section 6.1.1 and section 6.1.2 discuss the effect that changes in temperature have on the NIR spectrum of water and glucose respectively.

6.1.1 The Influence of Temperature on the Water Absorption Spectrum

Biological systems contain a high water content and in the case of *in vivo* measurements, precise control of temperature is not feasible. The strong dependence of the absorption of water on temperature is therefore a vital consideration [37]. Even though the core temperature of the body remains fairly constant, the temperature of the extremities can drop several degrees below this core temperature [37].

The majority of information below is attained from a detailed study performed by Jensen et al. [37] who examined the influence of temperature on water and glucose absorption spectra at physiologically relevant temperatures. Jensen et al. focused on the practical consequences of temperature variations for quantitative measurement of trace elements in aqueous solutions.

The variations which occur in the NIR spectrum of water due to changes in temperature are shown in figure 6.1 and 6.2. Figure 6.1 shows the molar absorptivity at three physiologically relevant temperatures. Figure 6.2 shows the difference spectra attained by subtracting the absorption spectrum at 37°C from the spectrum attained at temperatures ranging from 30°C to 42°C . It can be clearly seen that the variations are highly non-linear and that the magnitude of the variation is strongly dependent on wavelength. The symmetry around the reference value of 37°C is also visible.

The water absorption peaks shift to higher frequencies as the temperature is increased due to changes in the extent of hydrogen bonding [14, 70]. This can be observed in figure 6.1.

The changes in the spectrum due to temperature are of a broad-band nature with the largest changes occurring in the vicinity of the water absorption bands. In regions of strong absorption, small changes in temperature have a major impact on the water spectrum which

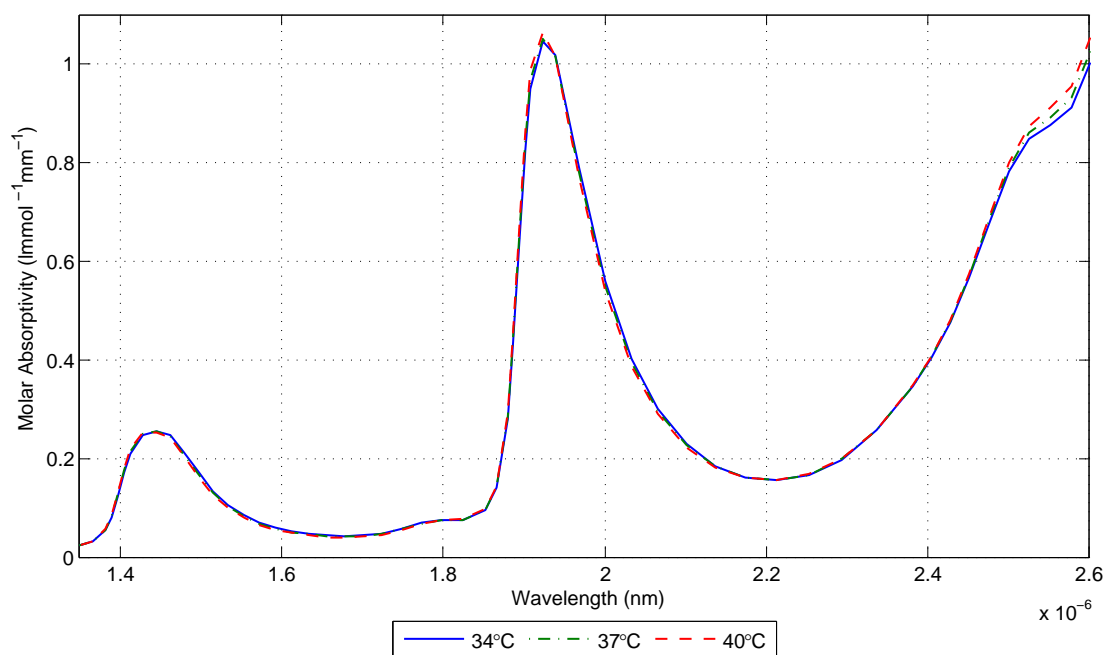


Figure 6.1: Water molar absorptivity in the NIR region at different temperatures [37]

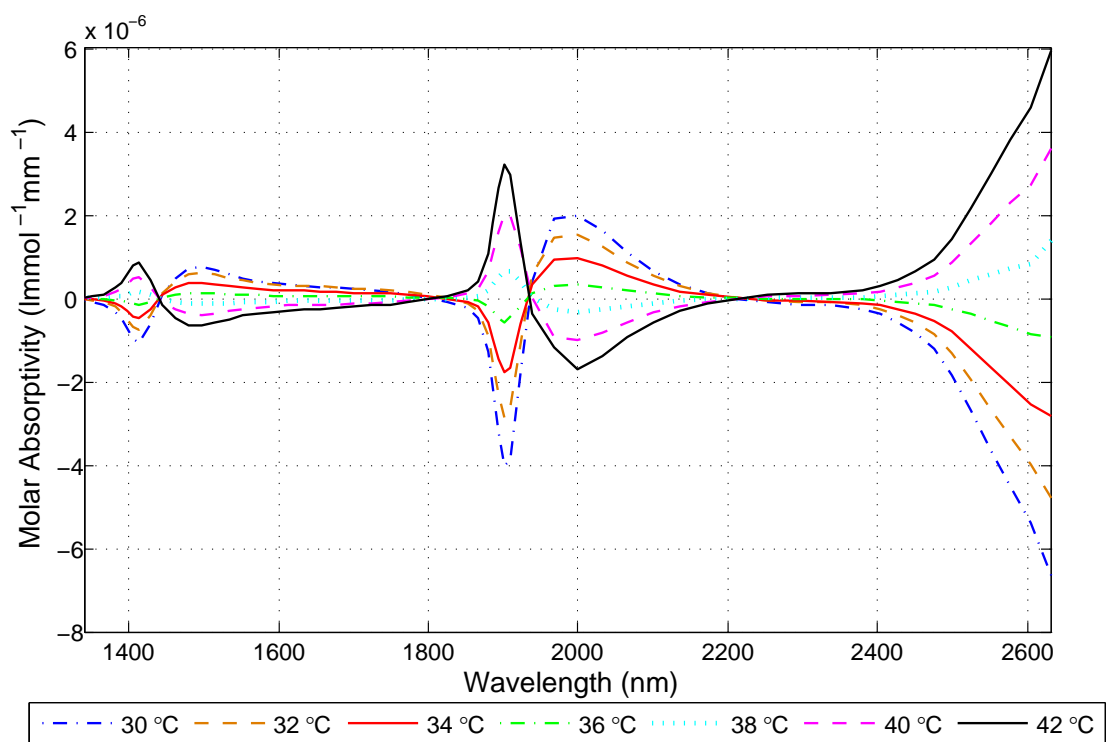


Figure 6.2: Difference in water molar absorptivity due to temperature changes (reference temperature: 37°C) [37]

would make in vivo measurement very difficult. There are, however, a few spectral windows in which quantitative analysis is possible [37].

The regions of the spectrum that are of particular interest are those between $2.2 \mu\text{m}$ and $2.38 \mu\text{m}$ and between $1.54 \mu\text{m}$ and $1.80 \mu\text{m}$ which show an almost flat dependency on temperature [37]. These regions of the spectrum are shown in figure 6.3 and 6.4.

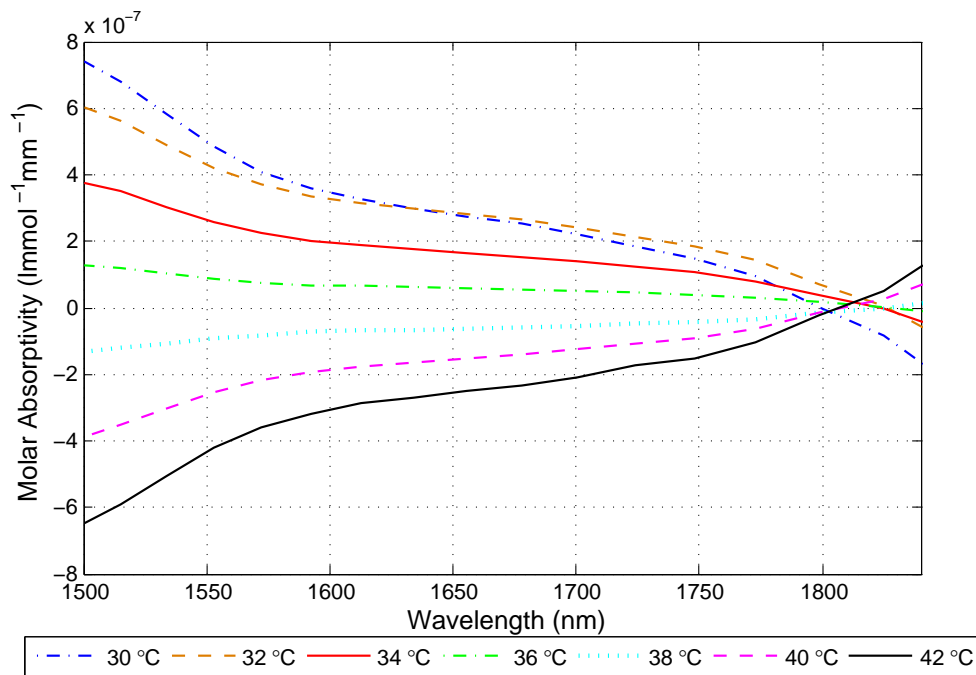


Figure 6.3: Difference in water molar absorptivity in first overtone region due to temperature changes (reference temperature: 37°C) [37]

Glucose has spectral features in both of these regions. When glucose is present in a solution, the glucose spectral information is superimposed on an almost flat baseline [37].

The glucose spectral bands in these regions are much narrower than the broad-band variations of water. The effect of the water spectrum can therefore be removed using a high-pass filter. Hazen, Arnold and Small have successfully used filtering techniques to remove these low frequency variations due to temperature changes [70, 46].

The addition of 5.55mmol/l of glucose to an aqueous solution has an influence comparable to a temperature change of 0.1°C [37]. This suggests that even in situations where tight control of temperature is possible, significant baseline variations will be introduced and compensating baseline correction will still be required in order for glucose quantification to occur.

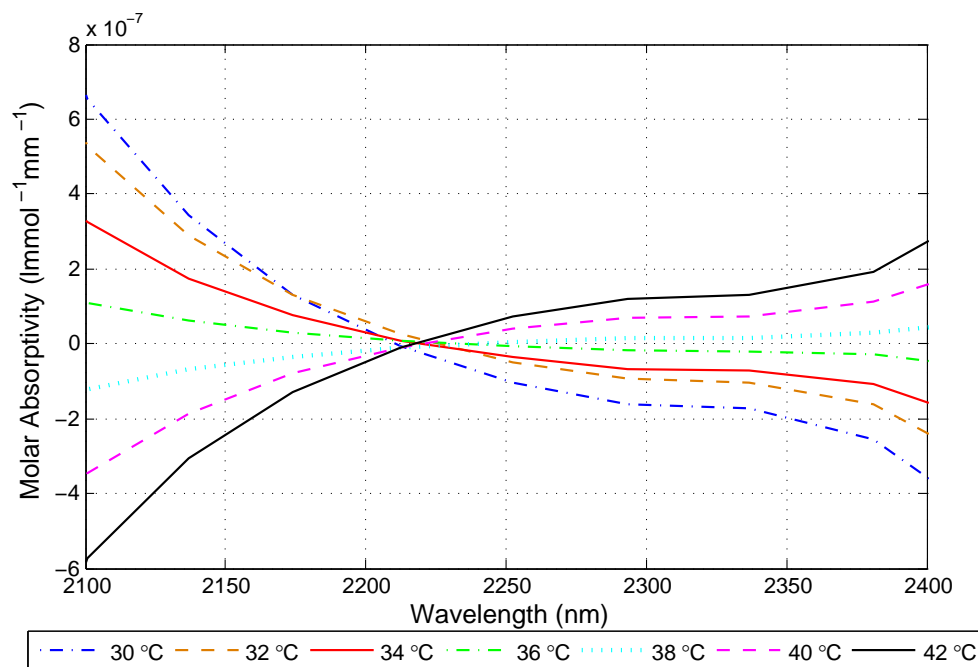


Figure 6.4: Difference in water molar absorptivity in combination region due to temperature changes (reference temperature: 37°C) [37]

6.1.2 The Influence of Temperature on the Glucose Absorption Spectrum

Studies that have attempted to measure the effect of temperature on aqueous glucose concentrations have found the spectral changes to be dominated by variations in the spectrum of water [37, 70]. Temperature variations therefore complicate the process of measuring changes in the glucose absorption spectrum significantly. A study performed by Hazen et al. shows that the positions of the three major glucose absorption bands in the combination region of the NIR spectrum remain constant over the temperature range 33 °C to 41 °C. The position, size and shapes of the bands situate near 2.27 μm and 2.326 μm were found to remain constant in this temperature range. The integrity of the 2.1 μm band was, however, compromised due to the decreased optical throughput caused by the shift in the absorption spectrum in water [70].

The insensitivity of the 2.27 μm and 2.326 μm spectral features to temperature is due to these bands being caused by combinations involving C-H stretching vibrational transitions [39, 70]. Vibrational characteristics of O-H groups are affected by hydrogen bonding with water leading to O-H vibrational transitions in the combination and first overtone region being temperature sensitive [70].

The three major glucose bands in the first overtone region of the NIR spectrum are positioned around 1408nm, 1536nm and 1688nm [39]. The 1688nm band is unlikely to be sensitive to temperature changes as it results from the first overtone of the C-H bond [39]. The features at 1408nm and 1536nm bands are a first overtone O-H band and an C-H + O-H combination band, respectively [39]. The presence of the O-H bonds increases the temperature sensitivity

of these spectral features but there is no evidence to suggest that temperature variations in the physiologically relevant range will have a significant effect on the position of these spectral bands.

6.2 The Effect of Interfering Analytes on Glucose Measurement

A major problem with performing blood glucose measurement is that many interfering analytes are present in blood that distort the spectral information thereby complicating the task of detecting changes in the glucose concentration. The concentrations of these analytes may be significantly higher than the blood glucose concentrations. Many of the analytes in blood have similar chemical compositions to glucose and similar chemical bonds resulting in the spectral peaks of these analytes overlapping with the glucose absorption bands. The concentrations of many of these components of blood vary with time. In order to be clinically useful, a glucose monitor must have sufficient selectivity to differentiate changes in glucose levels from variations in the concentrations of other analytes.

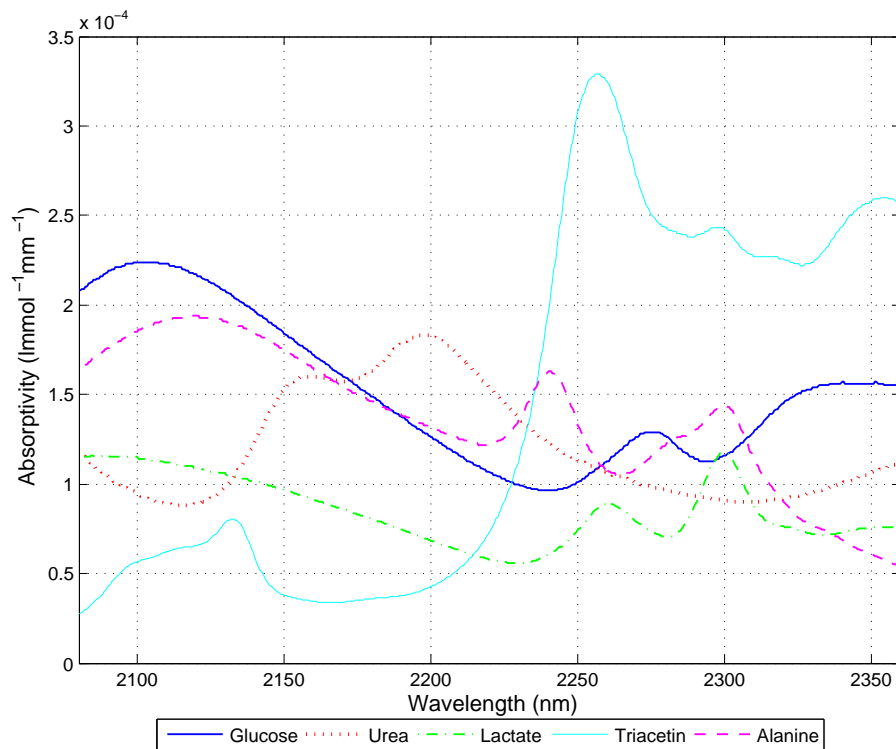


Figure 6.5: Combination region molar absorptivities of major blood components. Data obtained from [38].

Figure 6.5 and figure 6.6 show the combination region and first overtone spectra of glucose and various other analytes which are likely to interfere with glucose measurement. Alanine is an amino acid, which represents the effect that proteins will have on the NIR spectrum while triacetin is a triglyceride representing the effect of fats.

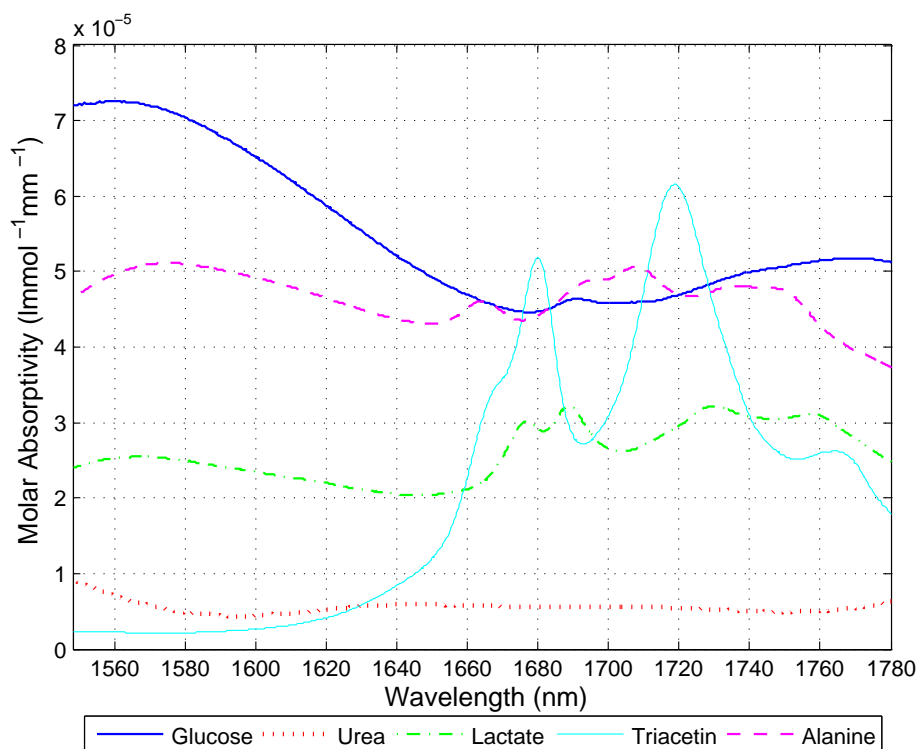


Figure 6.6: First overtone region molar absorptivities of major blood components. Data obtained from [38].

As discussed previously, water is the primary absorber of NIR radiation. The NIR spectrum of blood is therefore dominated by water with the absorbance due to other analytes being several orders of magnitude lower than that of water. The absorption due to protein and lipids is also significant due to their high concentrations and well-defined absorption peaks in the NIR spectral region. The concentration of blood proteins is approximately 6-8g/dl and that of lipids is approximately 500-900mg/dl. The concentrations of urea and lactate are slightly lower than that of glucose [88].

There are several other solutes in blood with relatively low concentrations that have spectral features within the NIR spectral region which may also interfere with the measurement of glucose. These include cholesterol, uric acid and glycerol [63].

The absorbance of NIR radiation by fat, skin and muscle is also an important consideration for in vivo glucose measurement. A paper published by Burmeister et al. discusses the combination and first overtone spectra of animal tissue [22]. Fatty tissue attenuates strongly at the lower frequencies of the first overtone with little light being transmitted at wavelengths less than 2273nm. Strong absorption bands are centred at 2299nm and 2342nm. The spectral features of skin are similar to those of fat, but of smaller magnitude. The absorption features of muscle are similar to those of water with additional features around 2174nm and 2288nm. The resemblance between the spectra of water and muscle is due to the high water content of muscle. The additional features are due to the presence of protein. The absorbance due to muscle and water is lowest between 2170nm and 2270nm [22].

In the first overtone region, similar results are found, with fat absorbance dominating at lower frequencies and the absorbance of water dominating at higher frequencies. Two strong, overlapping fat absorbance bands are centred at 1729nm and 1757nm. The absorbance of protein is much lower in the first overtone region than in the combination region [22, 53].

The components of human blood and tissue will mask changes in glucose concentration by providing interfering spectral features in the spectral ranges of interest. The calibration models are required to differentiate between the spectral characteristics of glucose and the spectral features of these other constituents which will be present in the optical path of a spectroscopic measurement device.

Chapter 7

Glucose Measurement in Simulated Aqueous Solutions

Development of an in vivo glucose measurement system is a complex problem. It is therefore not possible to perform detailed research into all aspects of the problem. The focus of the research is on several major aspects involved in the development of a continuous monitor, mainly relating to the data processing of spectral data the extraction of glucose information from spectroscopic data. The development of a specialised spectrometer for use with a glucose measurement device, miniaturisation and cost considerations are not considered in detail and no in vivo measurements are performed.

This chapter discusses the use of computer-based models to simulate the spectroscopic results which would be obtained from the analysis of aqueous solutions that approximate the spectra of human blood samples. Artificial neural networks are used to analyse the simulated spectroscopic data.

The use of computer-based simulations enables factors affecting glucose measurement to be determined and thorough theoretical understanding to be gained. The feasibility of non-invasive spectroscopic glucose monitor can be determined and factors affecting glucose measurement can be analysed in a systematic manner under controlled conditions.

The use of neural networks for extracting the glucose information from the spectral data is investigated. Only a limited amount of research has been performed into the use of ANN's for NIR spectroscopic glucose measurement with the majority of researchers favouring the use of linear calibration techniques such as PLS and PCA [62, 63, 64, 65, 66, 67, 68, 69]. ANN's appear to be a promising alternative as neural networks have been successfully applied in many spectroscopic applications [89]. Neural networks potentially have advantages over linear calibration techniques when non-linear data is being modelled. Two neural network architectures are used during the modelling process, MLP and RBF networks. The performance of these network architectures is compared.

The use of simulated spectral data enables various questions relating to the use of NIR spectroscopy for glucose measurement to be answered rapidly without the cost implications and time delays introduced when laboratory experiments are performed. The knowledge gained

through the use of simulations will enable informed design decisions to be made during research involving *in vitro* and *in vivo* spectroscopic measurements. It will also allow various network architectures to be experimented with and demonstrate the advantages and disadvantages associated with use of the different data preprocessing techniques. The position of glucose absorption bands, tolerable noise levels, the required spectrometer resolution and the number of samples required for effective network training, are also considered. The optimal frequency range for NIR spectroscopic glucose measurements is investigated. Simulations are performed using data from both the combination and first overtone region of the spectrum.

The primary aim of this stage of the project is to determine whether NIR spectroscopy and artificial neural networks can be used to perform glucose measurements in complex solutions with sufficient accuracy to be clinically useful. In order to be useful for *in vivo* measurements, the monitoring system must be sufficiently robust to operate in environments which include scattering effects, temperature variations, pathlength changes, instrumentation drift and variations in the concentration of other analytes. Simulations are performed using data which incorporates the effect of various forms of interference to determine how they are likely to affect *in vivo* glucose measurements.

Secondary considerations include the effects of using different network architectures, the determination of the optimal network parameters, the calculation of the minimum amount of input data required to attain satisfactory results and the effect of the various data processing techniques.

The analysis of computer-generated data, enables the effect of various interferences to be analysed independently under controlled conditions. Measuring the effects of each interference independently can enable a deeper understanding of the NIR spectral information to be gained. A major benefit of performing simulations using computer-generated spectral data is that it enables experiments to be performed which would be difficult to perform in a systematic manner under laboratory conditions. Performing controlled pathlength variations and temperature changes experimentally is time-consuming and expensive. The effect of these interferences will be simulated based on the findings of previous researchers. Methods of overcoming these interferences are to be determined.

MATLAB release 14 was used for the data analysis and the development of the multivariate calibration models. The open-source Netlab toolbox, created by Ian Nabney, was used for the implementation of the neural networks [90].

7.1 Overview of the Simulations

The chosen approach to the problem is to begin by determining the glucose concentrations of aqueous solutions under simple conditions and then to gradually increase the complexity until a point is reached where the majority of interferences present in *in vivo* measurements are considered. By initially modelling simple systems, and then gradually increasing the complexity, it is possible to determine the accuracy of the model at each step. Interfering factors can be compensated for as they are introduced and an understanding of the factors

involved in spectroscopic glucose measurement can be gained [39]. The spectral data is generated based on the findings of previous researchers [38, 37].

Researchers using an empirical approach for the *in vivo* measurement of glucose are unable to validate that measured changes are due to variations in glucose concentrations rather than indirect changes, chance correlations or time-dependent factors [5, 14, 24, 28, 41]. This has led several researchers to state that a systematic approach, which allows a better theoretical understanding to be gained, should be favoured [28, 39].

Several different sets of simulated spectral data were developed to test the ability of the calibration models to overcome the major forms of interference which will be present during *in vivo* glucose measurement. The various simulations that were performed are stated below:

- The first simulation investigates the ability of neural networks to predict glucose levels in aqueous solutions containing several interfering analytes of varying concentrations.
- The second simulation investigates the effect of temperature variations on glucose measurements and determines the ability of the calibration models to perform temperature insensitive measurements.
- The third simulation allows the effect of random high frequency spectral noise to be analysed.
- The fourth simulation investigates the ability of the networks to overcome the multiplicative variations caused by pathlength changes.

For each of these simulations, the advantages and disadvantages of using several different network architectures were considered.

Once the effect of each of the interferences had been analysed individually, a simulation was performed using computer-generated data which simulated the spectral data of a complex solution containing multiple interferences. This simulated data incorporated the effects of all the forms of interference mentioned above.

The final phase of the simulations investigated how the predictive ability of ANN calibration models could be improved through the effective use of data preprocessing techniques.

7.2 The Simulated Aqueous Solutions

7.2.1 Generation of the Simulated Spectral Data

Simulated spectral data was generated for aqueous solutions that contain the major components of human blood. The samples each contain different glucose levels and different concentrations of several interfering analytes. Spectral data was generated for both regions of interest, the first overtone region and the combination region. The spectral data described

in this section forms the basis for the data used in the simulations discussed later in this chapter.

The simulated aqueous solutions were generated using the molar absorptivities of water, glucose, alanine, triacetin, lactate and urea in the NIR region published in a paper by Amerov et al. [38]. Amerov et al. obtained the spectral information using a Nexus 670 Fourier transform spectrometer with a 20-watt tungsten filament lamp, calcium fluoride beam splitter and a cryogenically cooled indium fluoride detector. Optical filters were used to isolate the first overtone and combination region independently. The temperature was maintained at $37.0 \pm 0.1^\circ\text{C}$ [38].

Alanine, triacetin, lactate and urea were selected for inclusion in the simulated solutions as they represent the spectral influence of major components of human blood. Due to the locations of their spectral features, they are likely to interfere with the glucose measurements. Triacetin is a simple triglyceride which is used to simulate the effects of fat. Triglycerides are the main constituent of animal fats. Alanine is an amino acid found in many proteins. It is used to simulate the effect that proteins will have on the NIR spectrum.

The aqueous solutions were created by varying the concentrations of each of the analytes within physiologically relevant ranges, as shown in table 7.1.

Analyte	Concentration Range
Glucose	2.0 - 25.0 mmol/l
Lactate	0.5 - 3.3 mmol/l
Urea	2.3 - 8.3 mmol/l
Alanine	68.0 - 90.0 mmol/l
Triacetin	6.0 - 24.0 mmol/l

Table 7.1: Concentrations of analytes used in simulated aqueous solutions [88, 91, 92]

Models used to simulate spectral data are often based on the assumption that the molecular interaction between the molecules is negligible. The total absorbance of the solution is therefore a summation of the individual absorbance values of the analytes in the solution. This can be expressed using the Beer-Lambert Law [38]:

$$A = \sum_{i=1}^n \epsilon_i c_i b \quad (7.1)$$

where A is the absorbance of the solution containing n analytes, b is the pathlength of light through the solution, ϵ_i is the molar absorptivity for analyte i and c_i is the concentration of analyte i .

This approach is not sufficient for the modelling of NIR spectra of aqueous solutions as solvent absorption is significant. Dissolution of solutes results in solute molecules displacing a specific molar volume of water leading to a reduction in the number of the water molecules in the optical path [38].

The model used to generate the spectral data described in this section incorporates the effect that water displacement has on the NIR spectral data. The water displacement effect can be modelled by summing the absorption of each analyte and subtracting the loss of absorbance due to the water displacement caused by the presence of the solutes. This can be represented mathematically as [38]:

$$\begin{aligned} A &= \sum A_{solute} - \sum A_{water_displacement} \\ &= \sum_{i=1}^n \epsilon_i c_i b - \sum_{i=1}^n \epsilon_w f_w^i c_i b \end{aligned} \quad (7.2)$$

where ϵ_w is the molar absorptivity of the water and f_w^i is the water displacement coefficient for the analyte i .

The generation of the computer-generated spectral data is based on equation 7.2. The water displacement coefficients of the various analytes are obtained from [38].

Equation 7.2 only considers the displacement of water by solute molecules. Displacement of solutes from the optical path by co-solutes will also occur. For the measurement of milli-molar concentrations, this co-solute displacement effect is negligible and is therefore neglected in the model [38].

The regions of the NIR spectrum chosen for the simulations were 1544.0 - 1780.1 nm and 2059.7 - 2360.2 nm. These ranges were selected since they contain glucose spectral features but do not include regions in which temperature changes have a major effect on the underlying water spectrum. This is discussed in more detail in section 6.1.1. The absorptivity of water is sufficiently low in these regions for an acceptable throughput to be attained.

The pathlength was set at 10mm for the measurements in the first overtone region of the spectrum. Pathlengths between 5 and 10mm are recommended for measurements in this spectral region as the pathlength is long enough to obtain acceptable sensitivity and the absorption is sufficiently low to obtain an adequate signal-noise-ratio. Various potential measurement sites in the human body have pathlengths within this range [53]. The pathlength for the measurements in the combination region was set to 2mm as the strong absorptivity of water in this wavelength range reduces the transmission of radiation. A shorter pathlength is acceptable in this region of the spectrum as the glucose spectral features have larger magnitudes.

The spectral data from [38] has a resolution of approximately 0.6 nm in the combination region and 0.3nm in the first overtone region. Since the spectral features in the NIR region are broad and the long term aim is to make an affordable glucose monitor which is suitable for home use, only selected frequencies were considered to simulate a resolution of approximately 5nm in the combination region and 3nm in the first overtone region.

The simulated spectral samples were randomised and then divided into 3 data sets for the purpose of training neural networks; a testing set, a training set and a validation set. The purpose of randomising the data is to ensure that each of the data sets is representative of the entire input space. This ensures that each set contains many sources of variance, that

the data sets are independent and that the network is not required to perform extrapolation [72]. Each of three data sets contains 200 randomly selected samples.

7.2.2 Analysis of the Spectral Data

Graphs showing the spectral data for five randomly selected samples, from the data generated using the process described above, are given in figure 7.1. These samples contain analyte concentrations within the physiological ranges shown in table 7.1.

As discussed in section 6.2, the spectral data from both the combination and the first overtone region is dominated by the NIR spectrum of water. This is due to the relatively high absorptivity of water and its high concentration in the aqueous solutions. The combination region (figure 7.1a) and first overtone region (figure 7.1b) spectra generated from the simulated aqueous solutions are almost indistinguishable from the spectra of water.

Figure 7.1c and 7.1d show the spectra from the same samples as figure 7.1a and 7.1b but with the spectrum of water removed. The spectral features of alanine, which represent the role of proteins, dominate. This is due to the concentration of protein being significantly larger than the concentrations of the other analytes. The spectral peaks around 1580nm, 1665nm and 1710nm in the first overtone region and those around 2120nm, 2240nm and 2300nm all due to the strong absorption of alanine at these wavelengths.

Figure 7.1e and 7.1f show the NIR spectrum with the absorption due to both water and alanine removed. This allows the spectral features of the other analytes to be observed. Several characteristic peaks of the remaining analytes are clearly visible. The combination region spectrum is dominated by a pronounced absorption peak at 2260nm and smaller peaks at 2140nm and 2350nm which result from the strong absorbance of triacetin. Triacetin has sharper absorption features than the other analytes and relatively large concentrations in the aqueous solutions. The spectral peak at 2300nm is due to the close proximity of spectral peaks of lactate and triacetin. Minor ripples in the spectra can be observed at 2150nm and 2200nm in some of the samples due to the presence of urea. Even once the absorbance of water and alanine have been removed, the spectral peaks of glucose are still difficult to observe. The broad nature of the glucose spectral features make it difficult to detect the glucose absorbance bands, visually. The peak at 2110nm can be seen in the figure but those at 2275nm and 2325nm are difficult to differentiate from absorbance features of the other analytes.

Once the absorbance due to water and alanine has been removed, the five spectral samples from the first overtone region also clearly show the sharp absorption peaks of triacetin (1680nm, 1720nm and 1765nm). A broad peak, mainly due to the absorption of glucose, is visible around 1565nm. The glucose features at 1690nm and 1770nm are difficult to observe.

Figure 7.1 clearly illustrates how the glucose spectral features are masked by those of interfering analytes and provides insight into why the spectroscopic determination of glucose will require the use of sophisticated multivariate calibration techniques. A univariate technique would not be able to differentiate changes in glucose concentrations from variations in the concentrations of other analytes.

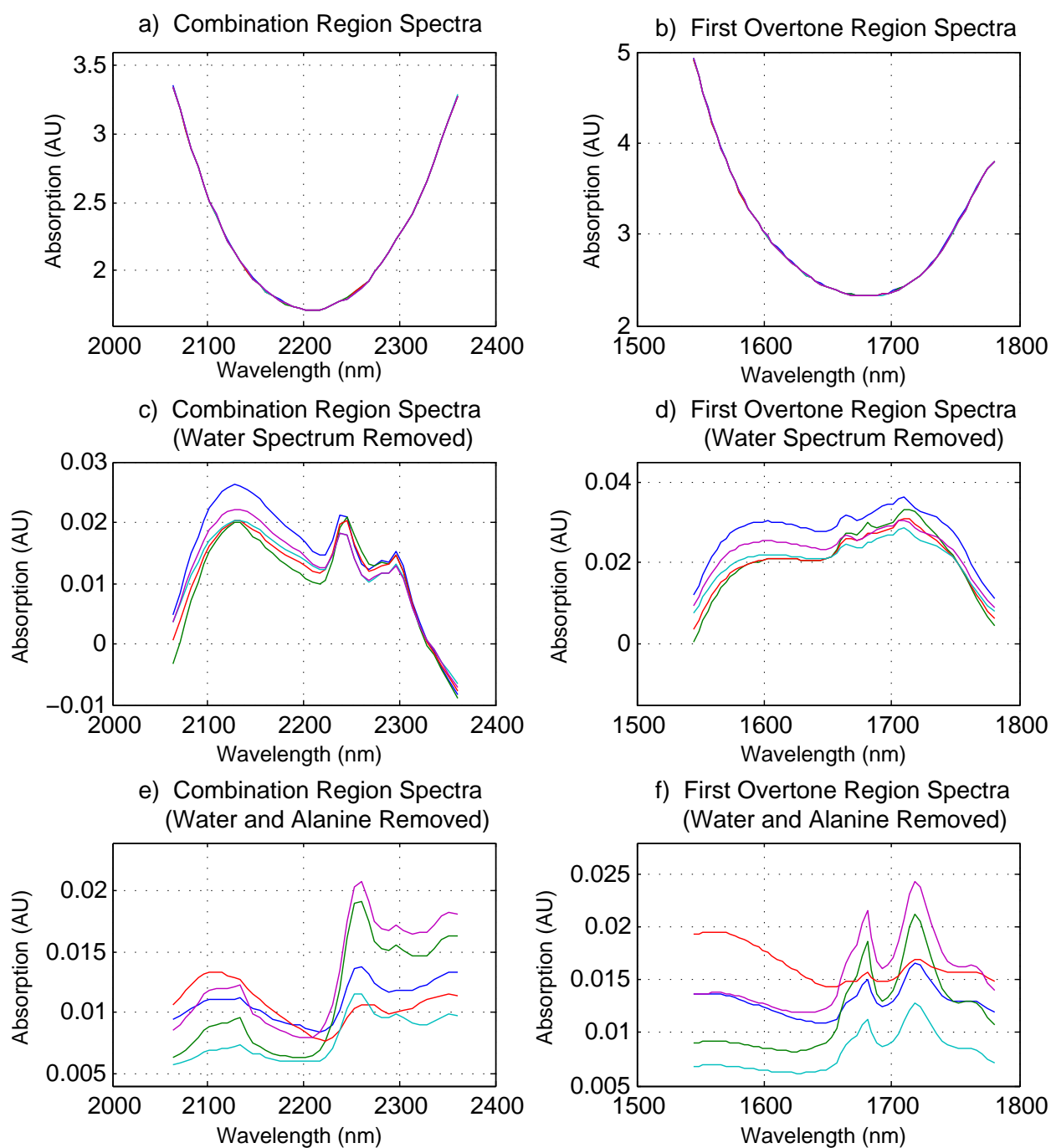


Figure 7.1: Simulated spectra of five randomly selected spectral samples in the combination and first overtone region

- Combination region spectra of simulated aqueous solutions
- First overtone region spectra of simulated aqueous solutions
- Combination region spectra with the spectrum of water removed
- First overtone region spectra with the spectrum of water removed
- Combination region spectra without the absorbance of water and alanine
- First overtone region spectra without the absorbance of water and alanine

7.3 Glucose Measurement in the Presence of Other Analytes

The first simulation to be performed was the detection of glucose in an aqueous solution containing several analytes of varying concentrations. This is a vital step towards in vivo glucose measurement as varying concentrations of analytes in the blood will distort the NIR absorption spectrum and mask the effect of changes in glucose concentration. The simulation will also determine the feasibility of measuring glucose in first overtone and combination region of the spectrum. As discussed in section 6, the spectra of many of the analytes found in blood have spectral features overlapping the spectral features of glucose. The changing concentration of these interfering analytes in the blood is one of the major challenges which must be overcome in the process of designing on continuous NIR spectroscopic glucose monitor.

7.3.1 Development of the Calibration Models

Neural networks were used to detect the presence of glucose in the simulated aqueous solutions. Two different network architectures were used, Multilayer perceptrons and radial basis functions. The networks have hyperbolic tangent hidden layer activation functions and linear output activation functions. The hyperbolic tangent activation functions in the hidden layer provide an approximately linear output for small inputs but are bounded which prevents excessively large weights from occurring. A hidden layer of hyperbolic tangent functions provides a distributed representation of the input in which information is stored across many hidden layer units [72, 93]. Various network parameters, including the number of hidden layer units, the number of training cycles, the choice of optimisation algorithm and the weight decay rate were adjusted to optimise the performance. Training was performed using data from the first overtone and the combination region of the NIR spectrum and the results were compared. The training, validation and testing data sets each contained 200 spectral samples. The training data was used by the neural networks during the supervised learning process to manipulate the network weights and thereby generate calibration models with good predictive ability. The training process attempts to minimise the network error which can lead to overfitting of the training data. To prevent this from occurring, the validation data is applied to the network periodically and the error is determined. When the validation error begins to increase, the network training is stopped. The testing data is only used once the training process is complete. The testing data is unseen by the network during the training process and therefore provides a good indication of the predictive ability of the neural network. Clarke EGA is used to ensure that the calibration models provide clinically accurate results.

7.3.2 Results and Discussion

The effect that changing the concentration of analytes has on the NIR spectrum, and consequently on the detection of glucose levels, can be clearly seen in figure 7.2. This figure

shows the simulated spectra for two samples with the same glucose concentration. The two spectra have significant differences due to the presence of different concentrations of alanine, triacetin, lactate and urea. The ANN's are required to generate a model which is capable of differentiating the glucose spectral information from the spectral interferences caused by the other components of the aqueous solutions.

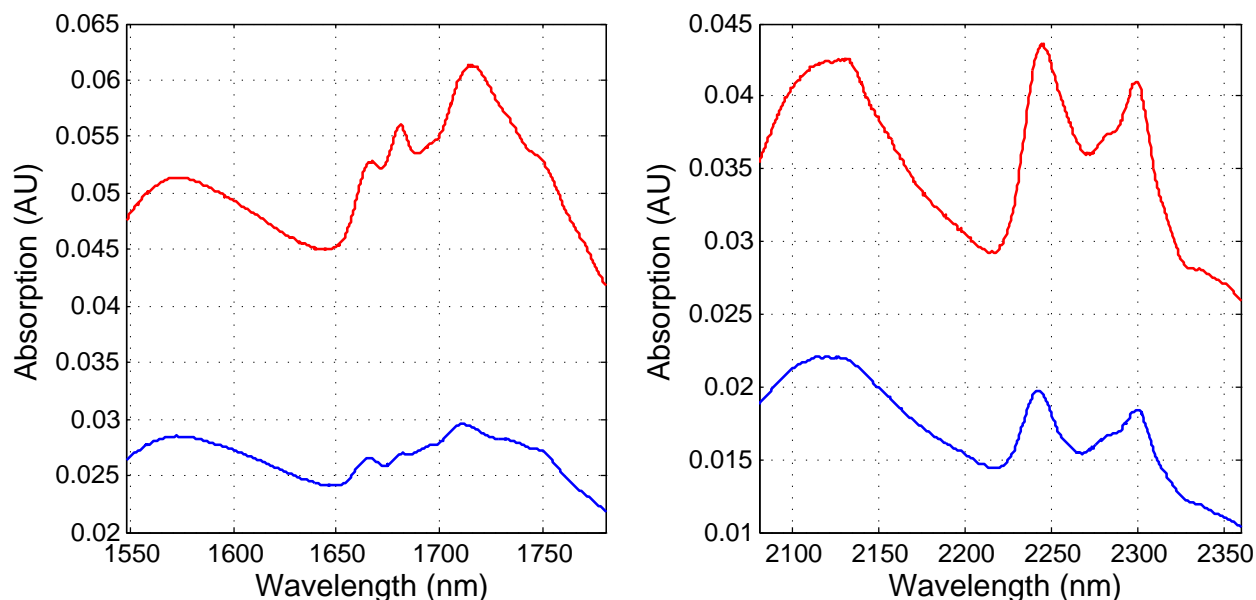


Figure 7.2: The effect of analyte concentrations in the first overtone region (left) and the combination region (right). The curves represent the spectra of samples with the same glucose concentrations and randomly chosen concentrations of lactate, urea, alanine and triacetin within physiologically relevant ranges. Absorbance due to water is removed for clarity.

The neural networks were able to detect the glucose concentrations with very high accuracy using data from either the combination region or the first overtone region. The results from this simple simulation are promising as they indicate that use of neural networks to monitor blood glucose could potentially provide clinically relevant results under controlled conditions. The most accurate results obtained from the simulations are shown in table 7.2. Use of the Quasi-Newton optimisation algorithm, which uses the BFGS (Broyden-Fletcher-Goldfarb-Shanno) Update Formula to calculate the Hessian matrix, during the training process was found to provide the best results.

Training the networks with the combination region data and the first overtone region data produces comparable results. Both ranges produce highly accurate results suggesting that they could both be successfully used for a spectroscopic glucose monitor. The performance of the MLP and RBF network architectures was very similar. The RBF networks proved to be marginally more accurate but were more susceptible to becoming trapped in poor local minima. The RBF networks were less inclined to becoming overtrained than the MLP's.

The Clarke error grid in figure 7.3 shows the results obtained when an MLP neural network was trained using combination region data. The predicted values are almost identical to the target values with all data points falling within region A of the error grid. The RBF networks and the MLP network using the first overtone data produced similar error grids.

Network Type	Spectral Region	Training Cycles	Hidden Layer Nodes	Time to Train (s)	Time to Run (s)	SEC (mmol/l)	SEP (mmol/l)
MLP	combination	400	8	11.87s	<0.01s	7.35E-4	7.73E-4
MLP	first overtone	300	8	14.03s	<0.01s	8.08E-4	8.11E-4
RBF	combination	1500	15	7.64s	<0.01s	2.98E-4	2.96E-4
RBF	first overtone	1000	10	4.11s	<0.01s	4.19E-4	4.21E-4

Table 7.2: Glucose measurement results for simulated aqueous solutions containing interfering analytes

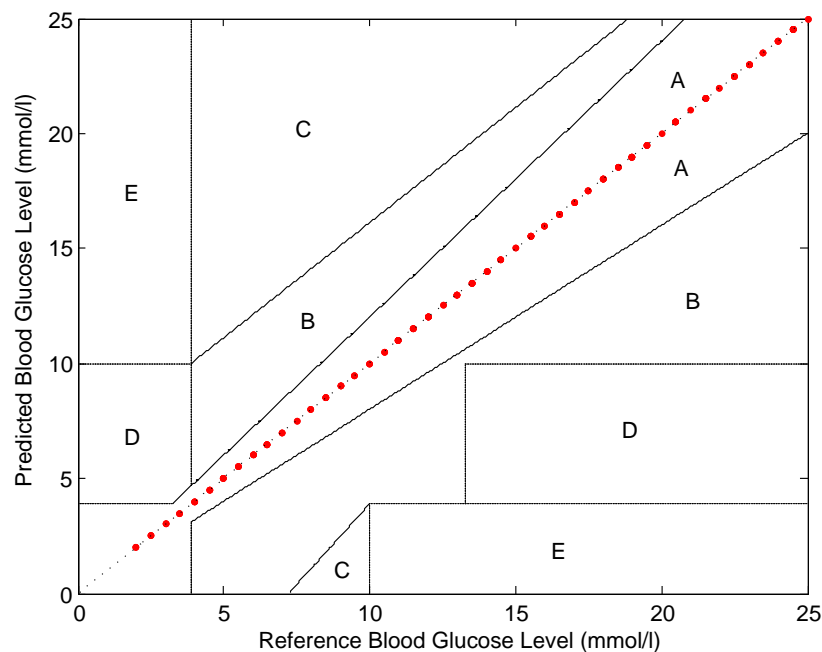


Figure 7.3: Clarke error grid for simulated aqueous solutions containing several analytes with no noise or temperature variations

7.4 Temperature Insensitive Glucose Measurement

Once it was determined that the glucose concentration could be successfully determined in a solution containing varying concentrations of other analytes, the next phase of the simulations was to ensure that accurate measurements were still attainable when there are variations in the temperature of the samples. As discussed in section 6.1.1, the underlying water absorption spectrum is strongly affected by temperature changes. Temperature changes lead to baseline and low frequency variations which mask the changes in glucose concentration. Since it is not practical to control the temperature precisely during in vivo measurements, a glucose monitor must be capable of differentiating the effects of temperature variations from changes in analyte concentrations.

7.4.1 Generation of the Spectral Data

The spectral data used for this simulation was based on the spectral data containing several analytes of varying concentrations described in section 7.2, but was modified to incorporate the effect that performing measurements at temperatures other than 37 °C has on the NIR spectra.

Each data sample from the simulated spectral data described in section 7.2, was randomly assigned a temperature between 34 °C and 38 °C. The difference between the spectrum of water at 37 °C and the randomly selected temperature, was added to the spectrum of each sample.

$$X_{i_{temp_var}} = X_{i_{37}} + (W_{37} - W_{temp_i}) \quad (7.3)$$

where $X_{i_{temp_var}}$ is the i th data sample at a randomly selected temperature, X_i is the i th data sample at 37 °C, W_{37} is the water spectrum at 37 °C and W_{temp_i} is the water spectrum at the randomly selected temperature.

The temperature range of 34 °C to 38 °C was chosen since the body temperature for in vivo measurements is likely to fall within this range.

The data describing the effect of temperature variations on the absorption spectrum of water was obtained from a paper by Jensen et al [37]. Since the spectral resolution of this data was different to that of the spectral data created previously, Matlab curve-fitting algorithms were used to approximate the difference spectra caused by the temperature variations with a continuous curve. The data was fit to a polynomial of degree four by minimising the least square error. These curves were then sampled to obtain the required number of discrete spectral values. The error introduced by the curve fitting is negligible.

As discussed in section 6.1.2, the shifting in frequency of the glucose spectral peaks, due to temperature changes within the physiologically relevant range, is minimal and can therefore be neglected.

7.4.2 Development of the Calibration Models

Both RBF and MLP neural networks were constructed to model the effect of temperature changes. Spectral data from the first overtone and combination regions was used independently to train the networks. The network parameters were adjusted in order to determine the parameters which resulted in optimal performance. The training, validation and testing data sets each contained 200 spectral samples. The training data was used to train the networks and the SEV when the validation data was applied to the network was checked periodically so that the training process could be terminated at the appropriate time. The unseen testing data was applied to the networks once the training process had been completed and the SEP was analysed in order to determine the performance of the networks. Clarke EGA was used to ensure that all the predictions made fell within clinically acceptable regions of the error grid.

7.4.3 Results and Discussion

Figure 7.4 and figure 7.5 show how the near infrared spectrum of one of the simulated aqueous solutions is modified by changes in temperature. The graphs show the NIR spectrum of the same aqueous solution at different temperatures. The graphs illustrate that temperature changes lead to low frequency fluctuations in the spectral data of aqueous solutions. The magnitude of the spectral variations caused by temperature changes are significantly larger in magnitude than the glucose spectral features. Even though the effect of temperature changes is significant, with careful observation, the characteristic peaks of analytes in the solution can still be observed in spectral data from both the combination region and the first overtone region.

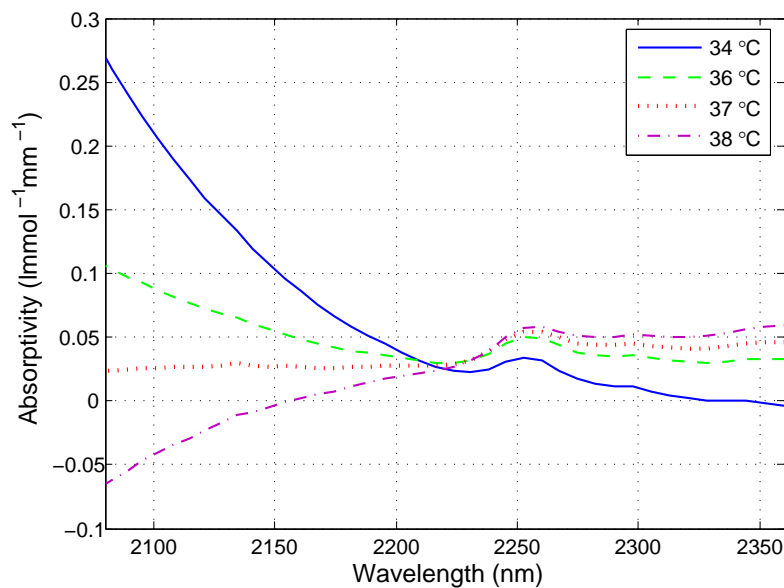


Figure 7.4: The combination region spectrum of a single sample at various different temperatures (water spectrum removed for clarity)

The best results obtained from the simulations for each network architecture and spectral region are shown in table 7.3. The ANN's were once again able to provide highly accurate predictions. The accuracy obtained using the first overtone and combination region data is comparable and the difference in performance between MLP and RBF network architectures is minimal. Clarke EGA (figure 7.6) shows the predicted glucose levels to be almost identical to the actual glucose levels. This suggests that the ANN calibration models are able to differentiate changes in analyte concentrations from the low frequency temperature effects and can definitely model temperature variations with sufficiently accuracy to provide clinically relevant information.

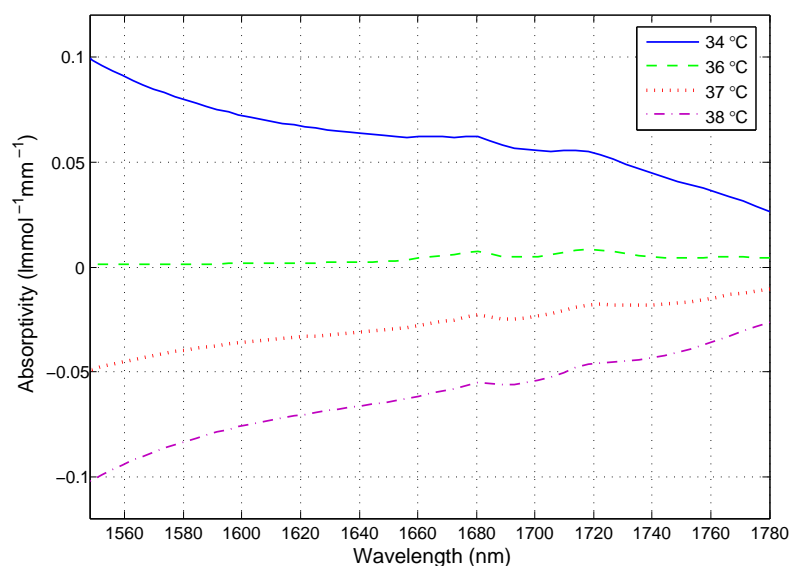


Figure 7.5: The first overtone region spectrum of a single sample at various different temperatures (water spectrum removed for clarity)

Network Type	Spectral Region	Training Cycles	Hidden Layer Nodes	Time to Train (s)	Time to Run (s)	SEC (mmol/l)	SEP (mmol/l)
MLP	combination	400	8	10.516s	<0.01s	1.46E-3	1.56E-3
MLP	first overtone	300	8	15.11s	<0.01s	2.02E-3	2.24E-3
RBF	combination	1500	15	8.63s	<0.01s	2.19E-3	3.00E-3
RBF	first overtone	1500	13	8.42s	<0.01s	3.30E-3	3.38E-3

Table 7.3: Glucose measurement results for simulated aqueous solutions containing interfering analytes and temperature variations in the physiologically relevant range

7.5 Glucose Measurement in the Presence of Random Noise

The spectral data obtained during in vivo spectroscopic measurements is likely to contain unmodelled high frequency noise caused by errors in the measurement process. The calibration models must be sufficiently robust to provide clinically relevant results when noisy spectral data is supplied. The simulations below determine the ability of the ANN's to make accurate predictions when random noise is added to the spectral data.

7.5.1 Generation of the Spectral Data

In order to simulate the effect of high frequency interferences, random noise was added to the spectral data samples described in section 7.2. The following equation was used to modify each spectral value by a random value:

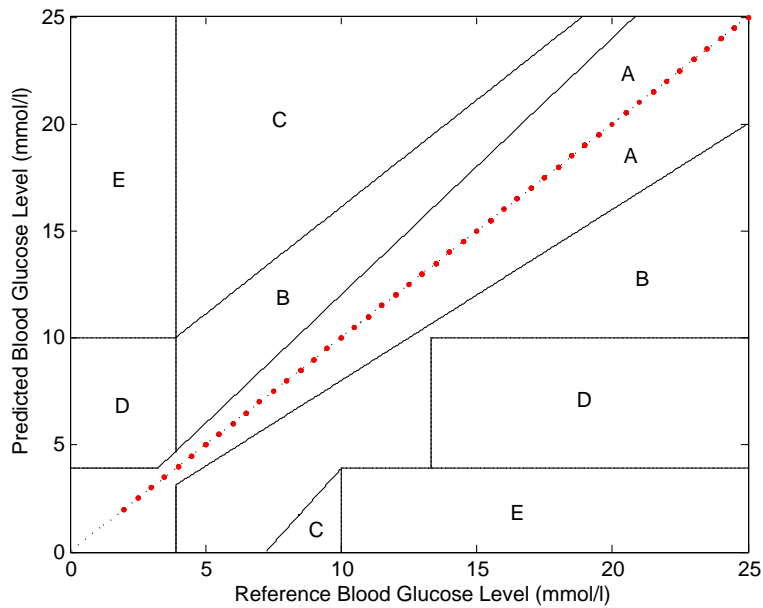


Figure 7.6: Clarke Error Grid for spectral data from simulated aqueous solutions with temperature variations within the physiologically relevant range

$$xn_{ij} = x_{ij} + rand * noise_level - noise_level/2; \quad (7.4)$$

where,

x_{ij} is the original spectral value at frequency i of spectrum j ,

xn_{ij} is the spectral value with random noise,

$rand$ is a random number between zero and one (with 15 significant digits),

$noise_level$ is the difference between the maximum and minimum amount of noise which could be added to the spectral value.

7.5.2 Development of the Calibration Models

RBF and MLP networks are trained using a similar process to that described previously. Noisy combination region data and first overtone region data was used independently to train neural networks. Network parameters were adjusted to maximise the predictive ability of the networks. The results obtained using the different network architectures and spectral regions for training were compared.

An attempt was made to determine the maximum amount of noise which could be added to the spectral data without the network making inaccurate predictions which lead to data points falling outside the A-region of the Clarke error grid.

7.5.3 Results and Discussion

The results obtained by MLP and RBF networks trained with spectral data containing random noise are shown in table 7.4.

Network Type	Spectral Region	Training Cycles	Hidden Layer Nodes	Time to Train (s)	Noise Level (AU)	SEC (mmol/l)	SEP (mmol/l)
MLP	combination	250	8	6.94s	4.0e-4	0.278	0.359
MLP	combination	150	8	4.37s	5.8e-4	0.359	0.471
RBF	combination	1500	20	11.00s	4.0e-4	0.300	0.370
RBF	combination	2000	16	12.13s	5.6e-4	0.466	0.498
MLP	first overtone	100	8	4.95s	6.1e-4	0.289	0.525
RBF	first overtone	1500	20	14.42s	6.0e-4	0.448	0.536

Table 7.4: Glucose measurement results for simulated aqueous solutions containing random noise

In the combination region of the NIR spectrum, glucose has a larger absorptivity than in the first overtone region. This would suggest that using combination region data to train the neural networks should result in calibration models which are less susceptible to the presence of random noise. The first overtone region, however, allows for greater transmission of radiation which means that longer pathlengths can be used. This compensates for the lower absorptivity and results in the maximum noise level which can be used without predictions falling outside the A-region of the Clarke error grid being similar for both spectral regions. The error grid attained when combination region data was used to train MLP networks is shown in figure 7.7.

The MLP networks could provide clinically accurate predictions with slightly higher noise levels than the RBF's.

7.5.4 Instrument Performance and RMS Noise

The determination of the maximum noise with which clinically acceptable results can be obtained is an important consideration when experimental results are obtained. The experimental setup and the performance of the spectrometer must allow readings to be taken with noise levels lower than the maximum acceptable values.

The Instrument performance is often represented as the root mean square noise on 100% lines (RMSN-100%) [53]. Two spectra for the same sample are obtained and one spectrum is divided by the other. In the ideal case, in which there is no spectral noise, the ratio will be a horizontal line at 100% transmission when plotted against wavelength. Plotting the ratio of two spectra of the same sample against wavelength clearly shows both spectrometer noise and instrument variation. The RMSN-100% value is calculated by fitting a first or second order polynomial function to the data and determining the root mean square value for the data relative to the polynomial [53]. The RMSN-100% value is normally measured in μ AU.

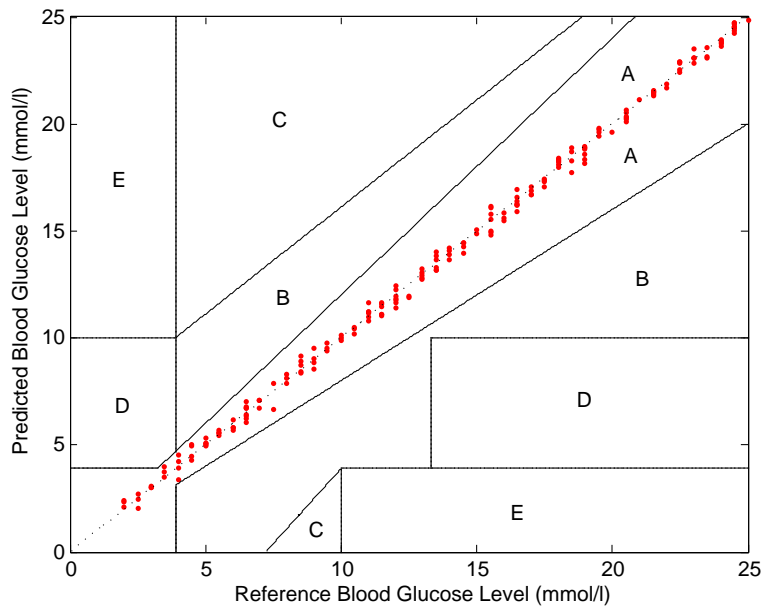


Figure 7.7: Clarke Error Grid obtained when combination region spectral data containing random noise was used to train an MLP network

Figure 7.8 illustrates the results when the RMSN-100% procedure is applied to the simulated spectral data with random noise. The random spectral variations account for the differences between two spectra of the same sample. Since simulated data has been used, no instrument variations are present which means that the ratio of two spectra can be approximated by a horizontal line at 100% transmission.

The RMSN-100% value with the maximum noise level which resulted in predictions being in the A-region of the Clarke error grid, is approximately $560 \mu\text{AU}$ for both the combination region and the first overtone region. This shows that the neural network calibration models have good noise rejection properties. It is expected that a RMSN-100% level, of less than $50 \mu\text{AU}$ would be required for in vivo glucose measurement due to the greater complexity.

7.6 Simulations with Variable Pathlengths

When in vivo measurements are performed, it is not possible to keep the pathlength constant. Pathlength fluctuations result in multiplicative variations in the absorption spectrum, since, according to Beer's Law, absorbance is directly proportional to pathlength. Simulating the effects of pathlength variations has added importance since the laboratory equipment available does not allow for the effect of pathlength variations to be determined experimentally.

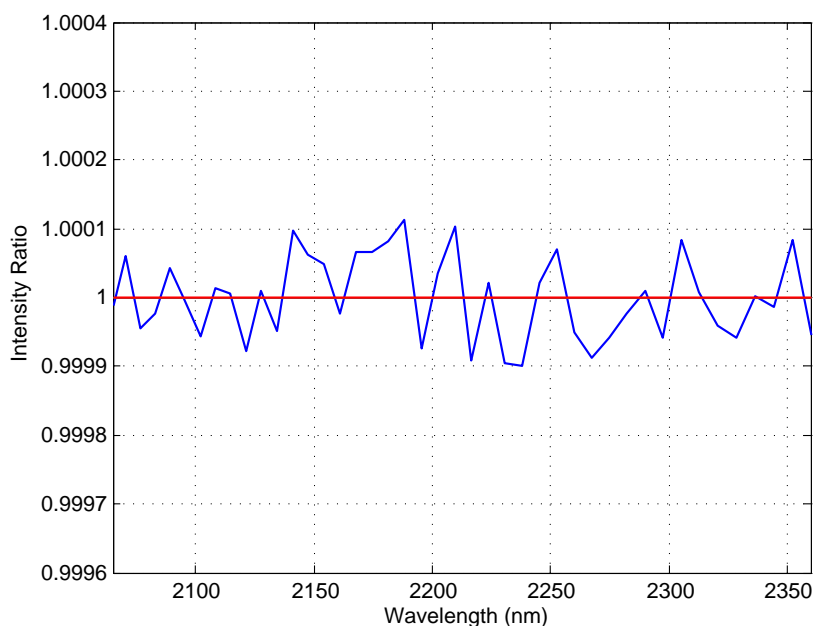


Figure 7.8: Representative 100% lines for simulated data with random noise. The ratio of two spectra of the same sample are plotted against wavelength to illustrate the effect of noise.

7.6.1 Generation of the Spectral Data

The simulated spectral data was created using a similar process to that described in section 7.2. The data was generated using equation 7.2. It differs from that in section 7.2 in that the pathlength for each sample is a randomly chosen value within 10% of the original pathlengths of 2mm for the combination region and 10mm for the first overtone region.

7.6.2 Development of the Calibration Models

RBF and MLP networks were trained with both combination and first overtone data containing variable pathlengths. Network parameters are adjusted in an attempt to improve performance. The results using the different network architectures and spectral regions were compared. Clarke error grid analysis and the SEP are used to determine the predictive ability of the networks.

7.6.3 Results and Discussion

Pathlength variations lead to a constant shift in magnitude across the measurement range. This is illustrated in figure 7.9 in which first overtone spectra of the same sample, with three different pathlengths, is given.

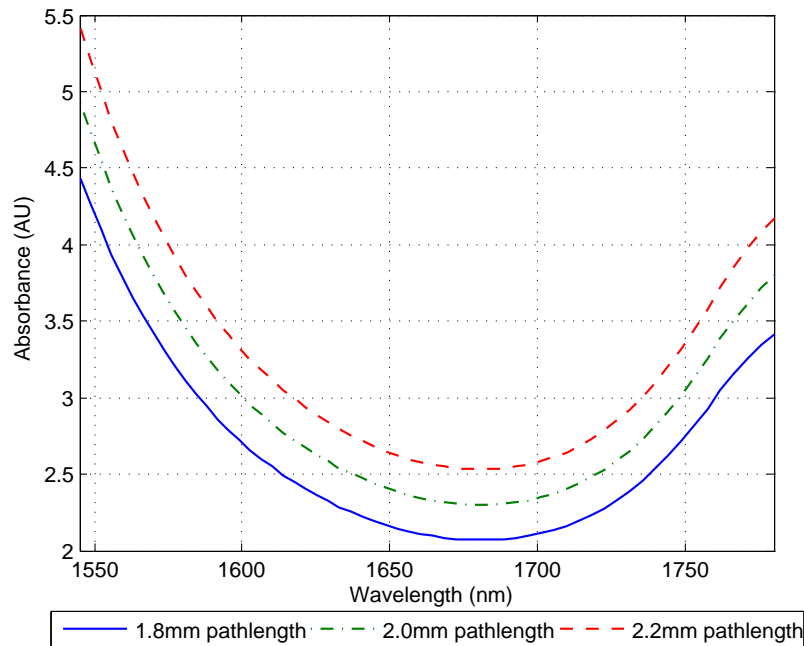


Figure 7.9: The NIR spectra of a single sample with three different pathlengths

The results obtained when MLP and RBF ANN's are used to generate the calibration models are provided in table 7.5.

Network Type	Spectral Region	Training Cycles	Hidden Layer Nodes	Time to Train (s)	Time to Run (s)	SEC (mmol/l)	SEP (mmol/l)
MLP	combination	400	8	11.328s	<0.01s	3.20E-2	3.36E-2
MLP	first overtone	300	8	14.29s	<0.01s	2.49E-2	2.75E-2
RBF	combination	1500	18	9.08s	<0.01s	3.90E-2	4.46E-2
RBF	first overtone	1500	20	11.02s	<0.01s	3.93E-2	3.82E-2

Table 7.5: Results for simulations with random pathlength variations of $\pm 10\%$

The neural networks were capable of performing highly accurate predictions when the pathlength was variable. In order to compensate for the increased complexity, the RBF's required an increased number of hidden layer nodes. The MLP networks performed marginally better than the RBF networks and the two spectral regions provided almost identical accuracy. The neural networks all attained prediction accuracies far greater than that required for in vivo measurement. All predictions fell within the A-region of the Clarke error grid. The error grid is shown in figure 7.10.

7.7 Simulations with Multiple Sources of Interference

The next phase of the simulations involved the determination of glucose concentrations in aqueous solutions containing analytes of varying concentrations, temperature fluctuation,

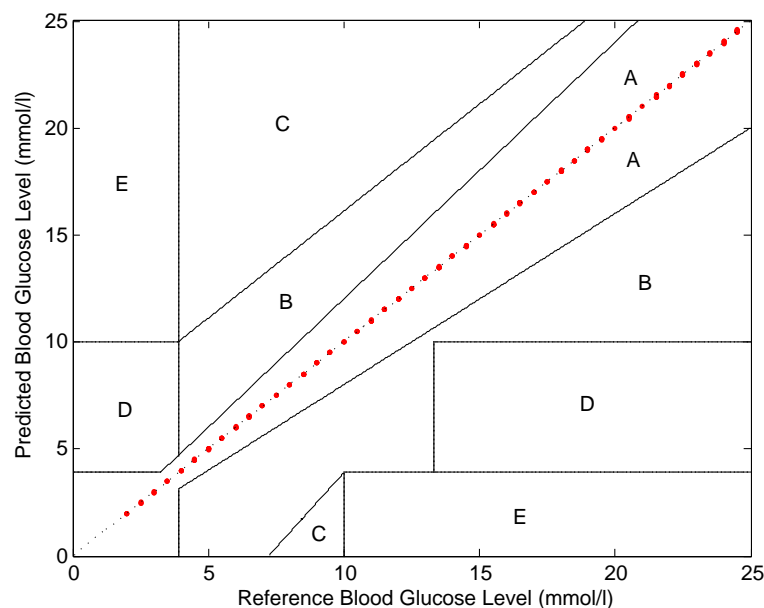


Figure 7.10: Clarke Error Grid obtained when combination region spectral data for simulated aqueous solutions with variable pathlengths was used to train an MLP network

random noise and pathlength variations. The simulation tests the ability of the calibration models in a complex environment with multiple sources of interference and thereby provides insight into the performance which is likely to be attained when in vivo spectroscopic measurements are performed. The simulations will demonstrate whether neural networks are capable of producing calibration models that provide clinically useful predictions from spectroscopic data with complexity approaching that which will be encountered when measurements are performed in human blood.

The spectra from the aqueous solution approximates those of human blood and the interferences caused by the simulated temperature and pathlength variations are representative of the types of interferences which will be present when in vivo measurements are performed in the human body. The random noise represents high frequency instrument noise which occurs during the spectroscopic measurement process.

The simulation includes examples of all the major sources of interference which will occur during in vivo measurement, multiplicative variations, low frequency fluctuations, high frequency changes and interfering spectra from other analytes.

7.7.1 Generation of the Spectral Data

The complex simulated spectra required for this phase of the project were generated by adding interferences caused by temperature variations, pathlength variations and high frequency random noise to the simulated data described in section 7.2. The interferences were added using the techniques discussed in the preceding sections.

7.7.2 Development of the Calibration Models

The ANN calibration models were developed using the same approach as in the previous sections. RBF and MLP neural networks were trained with data from the combination and first overtone regions of the spectrum and the results were compared. The standard error of prediction and Clarke EGA were used to judge the predictive ability of the networks.

7.7.3 Results and Discussion

Figure 7.11 depicts the process which occurs during the training of the neural networks. The figure illustrates the change in the error for the training and validation data as the number of training cycles is increased for an MLP neural network trained using spectral data from the combination region. When the training process begins, the training and validation errors both decrease as the network iteratively improves the calibration model. The network error is very similar for the training data and the validation data. After approximately 170 iterations the network begins to overfit the training data. This leads to the training error continuing to decrease but the predictive ability of the network becomes worse. The error when the validation data is applied to the network starts to increase. This is the point at which the training should be terminated.

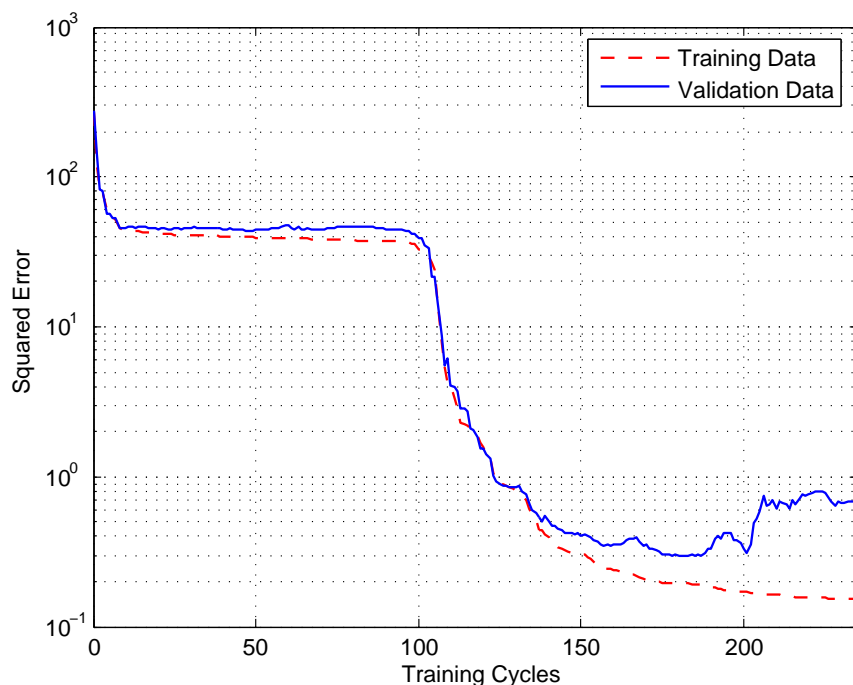


Figure 7.11: The square error versus the number of training cycles for an MLP network trained using combination region data.

The results obtained when the data containing multiple sources of interference was used to train the neural networks, are given in table 7.6.

Network Type	Spectral Region	Training Cycles	Hidden Layer Nodes	Time to Train (s)	Noise Level (AU)	SEC (mmol/l)	SEP (mmol/l)
MLP	combination	150	8	4.82s	3.0e-4	0.357	0.426
MLP	first overtone	200	10	14.828s	2.0e-4	0.502	0.531
RBF	combination	2000	16	11.76s	2.6e-3	0.442	0.455
RBF	first overtone	2000	20	14.20s	2.0e-4	0.591	0.643

Table 7.6: Glucose measurement results for the analysis of simulated aqueous solutions with random noise, pathlength variations and temperature changes. The concentrations of glucose, urea, alanine, triacetin and lactate are varied within their physiologically relevant ranges

The results in table 7.6 show that MLP and RBF networks are capable of making clinically relevant predictions in complex environments in which there are multiple sources of interference. The accuracy obtained in these simulations is lower than that for simulations with a single source of interference. This is expected due to the increased complexity of the system. With the noise levels given in table 7.6, the ANN's were capable of making predictions with sufficient accuracy for all the predictions made with the testing data to fall within the A-region of the Clarke error grid which indicates that correct treatment action would be taken.

The MLP networks proved to be slightly more accurate than the RBF networks. During the training process, the RBF networks frequently became trapped in poor local minima resulting in them producing an inaccurate model of the training data. This problem occurred less frequently when using MLP networks. The increased number of interfering signals in this simulation, led to the neural networks becoming more likely to over-fit the training data, resulting in a decreased ability to predict the glucose concentration from the unseen testing data.

Networks trained using the combination region data were less susceptible to noise than those in the combination region and produced calibration models with greater predictive ability. The better predictive ability of networks trained with combination region data can be explained by referring to a comparison between the combination and first overtone regions performed by Chen et al. [33]. Chen et al. quantify the difference between the spectral regions by measuring the net analyte signal (NAS) of glucose. NAS is the portion of the solute spectrum which is orthogonal to all other sources of spectral variance in the data matrix. Selectivity in multivariate analysis corresponds to differences in spectral shape. NAS indicates the degree of uniqueness of the glucose spectrum in the system containing multiple overlapping spectra. The length of the NAS vector per unit concentration and unit pathlength represents the degree of difference between the glucose spectrum and other sources of spectral information. The NAS vector for the combination region was found to be 3.8 times longer which indicates that the selectivity in the combination region is 3.8 times better in the combination region than in the first overtone region [33]. This suggests that the performance of multivariate calibration models using combination region data should have superior predictive abilities to those using first overtone data.

Figure 7.12 shows the Clarke error grid for an MLP trained with combination region data with a noise level of $300\mu\text{AU}$ and interferences due to temperature fluctuations between 34°C and 38°C and pathlength variations of $\pm 10\%$. The data points are scattered further from the diagonal representing a perfect prediction than during the simulations with a single form of interference, but the errors for all of the data points fall within the A-region of the grid and all the prediction errors are less than 20%. The percentage errors are greater for low glucose concentrations than for high concentrations. This is due to the signal-to-noise ratio being lower.

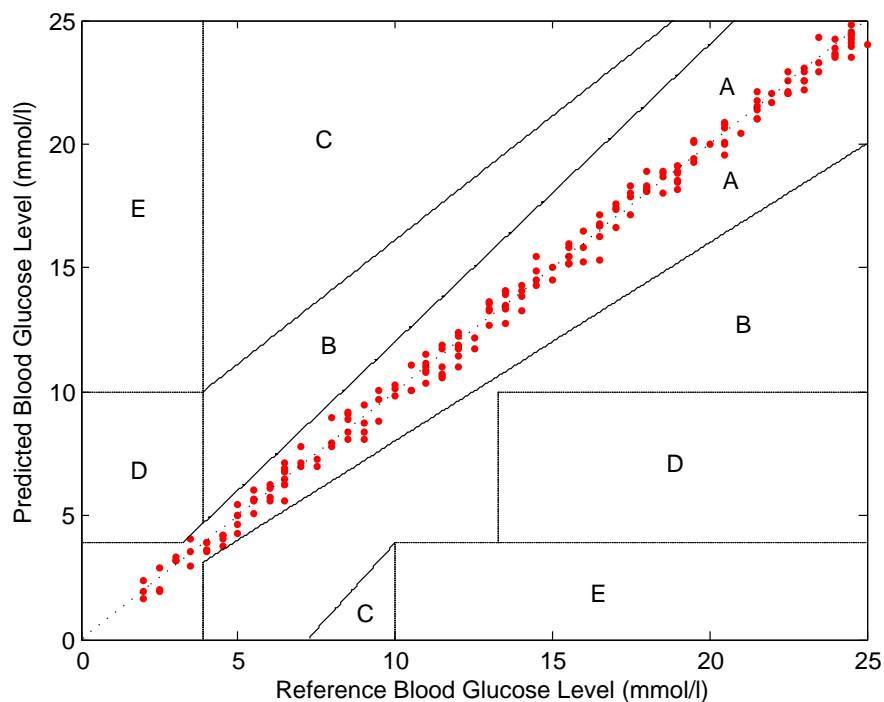


Figure 7.12: Clarke error grid for MLP network trained with combination region data with multiple sources of interference.

When human blood samples are being analysed, the time and difficulty involved in performing the spectroscopic measurements becomes an important consideration. An attempt was therefore made to determine the minimum number of spectral samples required in order to make sufficiently accurate predictions.

As the number of training samples is reduced, over-fitting of the input data becomes increasingly problematic. The network can fit the input data with very high accuracy but the generalisation becomes poor. By reducing the number of training cycles, this problem could be partially overcome. The number of training samples could be decreased significantly with only a small reduction in performance. Both MLP and RBF networks were able to make clinically relevant predictions (zone A of the Clarke error grid) with as few as 30 spectral samples.

7.8 The Use of Data Preprocessing Techniques

The preceding sections have shown that ANN's are capable of compensating for the major forms of interference which will be found during the in vivo spectroscopic measurement of glucose. The use of more advanced data preprocessing techniques can improve the performance of ANN's, thus providing improved prediction of blood glucose. The sections which follow, demonstrate the improvements which can be gained through the effective use of data preprocessing and provide a comparison between several different data preprocessing techniques. The spectral data with multiple sources of interference, discussed in section 7.7, is used in the simulations performed in this chapter. The following data preprocessing techniques are discussed:

- Moving average filters
- Normalisation by closure
- Time-domain filtering
- Fourier transform filtering
- First derivatives
- Second derivatives
- Multiplicative scatter correction
- Principal component analysis

Background information relating to these data preprocessing techniques is provided in section 5.5.

7.8.1 Preprocessing for the Removal of Multiplicative and Low Frequency Effects

As discussed previously, the multiplicative and high frequency variations caused by factors such as temperature variations, changes in pathlength and light scattering, must be overcome in order to make clinically relevant in vivo measurements. Several data preprocessing techniques can be used to reduce the effect of these interferences in multivariate calibration models. The techniques discussed in this document are normalisation by closure, filtering, the use of derivatives and multiplicative scatter correction (MSC).

Figure 7.13 shows computer-generated spectra of an aqueous solution simulating human blood. Due to temperature fluctuations and pathlength variations, the spectra from different samples have noticeable baseline offsets. As shown in section 7.7, ANN's are capable of overcoming these offsets during the modelling process. The use of preprocessing to remove

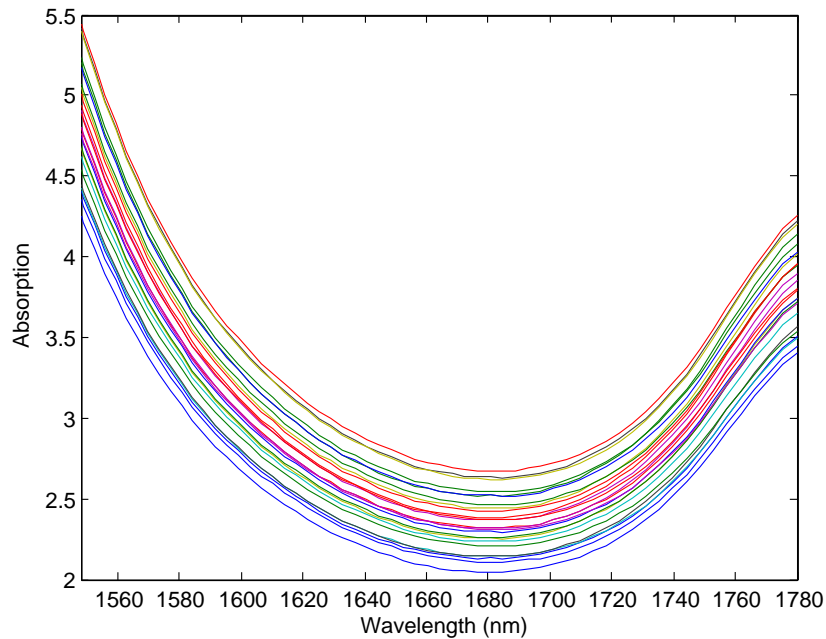


Figure 7.13: First overtone spectra of aqueous samples containing multiple sources of interference before data preprocessing has been performed

these unwanted effects before the network training commences, can simplify the modelling process thereby improving the performance of the calibration model.

Figure 7.14 shows the same spectra after data preprocessing has been performed. All the techniques shown in the figure greatly reduce low frequency differences between the spectra. This allows the calibration algorithm to focus on the meaningful higher frequency spectral signatures of the analytes. Figure 7.14 shows that normalisation by closure and MSC are able to remove the unwanted low frequency variations from spectroscopic data more effectively than the derivative-based methods and filtering techniques. The figure shows the output of the data preprocessing stage when using data from the first overtone region. Applying the various techniques to the combination region data produced similar results.

7.8.2 Preprocessing for the Removal of High Frequency Noise

Preprocessing techniques can also effectively remove interferences which occur at much higher frequencies than the characteristic peaks of glucose. Simulations were performed using three filtering techniques, a high pass fifth-order butterworth filter, moving average filter and a Fourier transform filter with a Gaussian window function.

Figure 7.15 shows the effect of using a moving average filter, which averages 5 adjacent spectral samples, for the removal of high frequency random noise. The water spectrum has been subtracted from both curves for the purpose of illustration. Similar results can be obtained using the other filtering techniques.

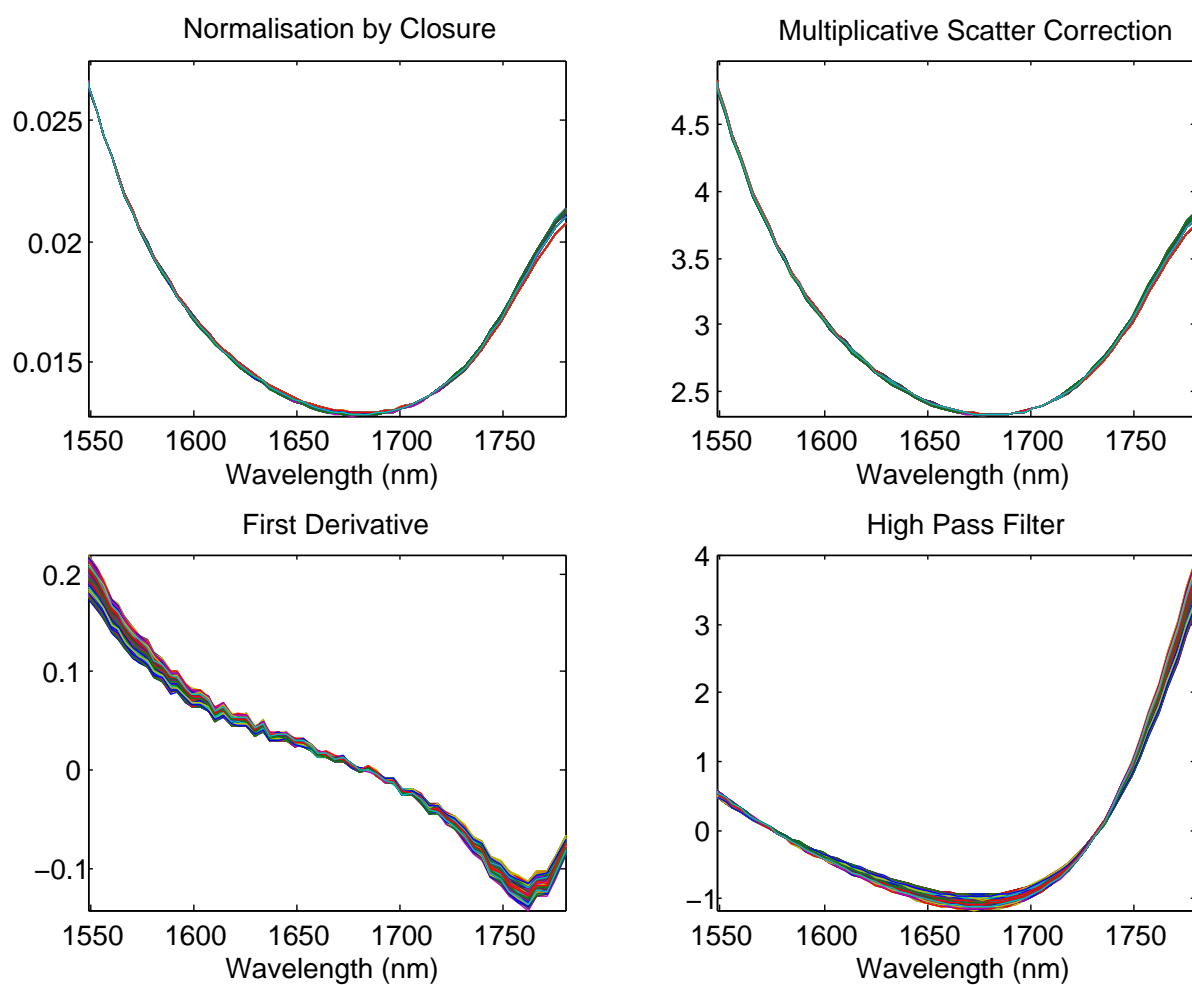


Figure 7.14: The effect of applying pre-processing techniques to the first overtone spectral data shown in figure 7.13. The resultant spectra when normalisation by closure, multiplicative scatter correction, first derivative and high pass filters are used, are shown.

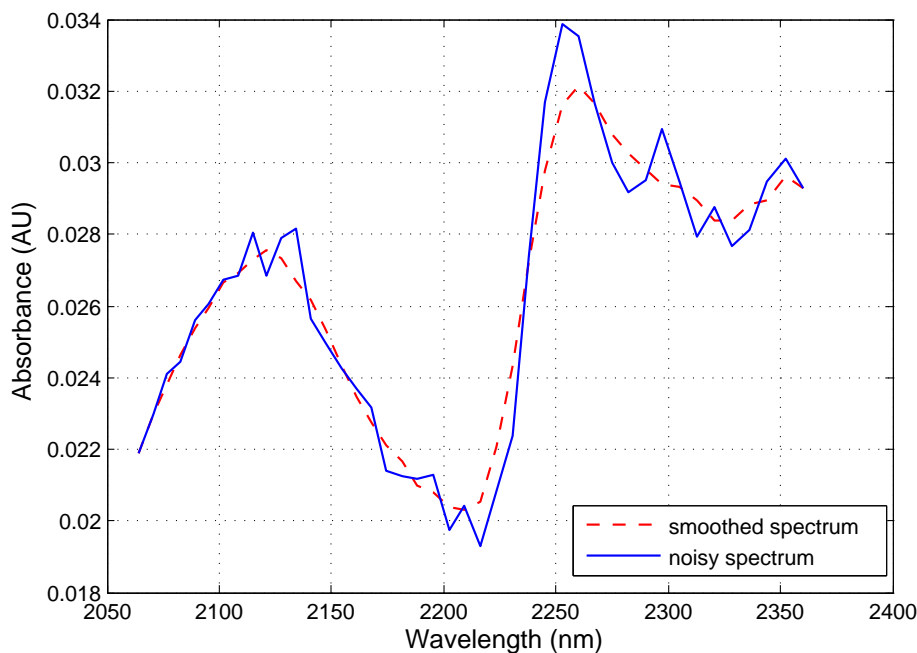


Figure 7.15: The removal of random noise with a moving average filter

7.8.3 Preprocessing for Data Reduction

Principal component analysis can incorporate the majority of the variance in the data set in a few principal components thereby decreasing the number of inputs to the ANN's. This can potentially lead to faster network training and better generalisation. The reduced number of variables can result in unwanted effects such as random fluctuations being discarded.

Due to the significant redundancy in spectral data, the input samples can be approximated with a high degree of accuracy with relatively few principal components. This is illustrated in figure 7.16. The original computer-generated spectrum is plotted along with an approximate spectrum generated from only 6 principal components. The NIR spectrum of water has been subtracted from the two spectra for clarity.

7.8.4 Performance of the Preprocessing Techniques

The results obtained when the various preprocessing techniques were used in conjunction with the ANN calibration models, are discussed below. The networks were trained using the data with multiple sources interference discussed in section 7.7.

Taking the first derivative of the input spectra resulted in networks of comparable performance to those discussed in section 7.7. The derivative spectra are less affected by low frequency variations as illustrated in figure 7.14 but have the disadvantage that high frequency noise is accentuated. The use of second derivative spectra decreased performance of the calibration models. The lowest standard error of prediction obtained is 0.7 mmol/l.

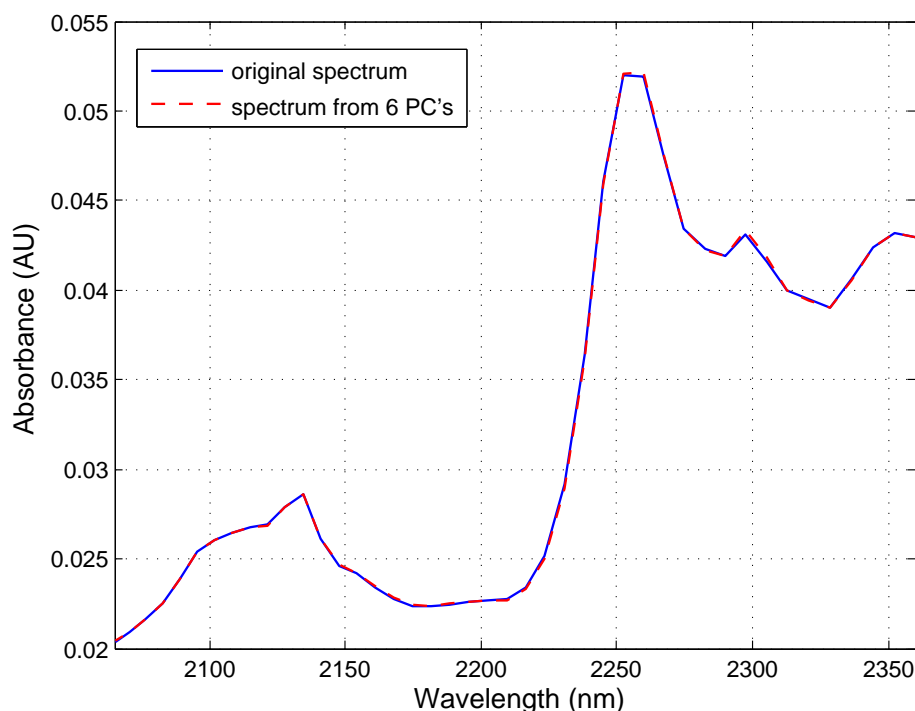


Figure 7.16: Approximation of spectrum with six principal components

The poor performance is thought to be due to the amplification of the high frequency random noise. The results of the simulations indicate that no significant advantages can be gained by obtaining the derivatives of the input data.

Normalisation by closure proved to be the most effective technique for the removal of low frequency and multiplicative variations. An SEP of 0.38 mmol/l was attained for an MLP network trained using combination region data. As shown in figure 7.14, normalisation by closure is an effective method of removing multiplicative effects which simplifies the modelling process and leads to improved generalisation when unseen data is used. This improved generalisation allows the network to be trained for more iterations without overfitting occurring.

Multiplicative scatter correction also improved the performance of the calibration models. An SEP of 0.41 mmol/l was attained. MSC, like normalisation by closure, removes low frequency variations resulting in good generalisation. MSC is commonly used for preprocessing of data for PLS calibration models [81]. These results show that it can also be effectively implemented with calibration models generated using neural networks.

Filtering the spectral data with a high-pass fifth order Butterworth filter resulted in a slight decrease in accuracy. The cut-off frequency was adjusted in an attempt to improve the performance but no filters were generated which could improve on the results obtained without the use of the filter. This suggests that techniques such as MSC and normalisation by closure, which take into account the similarities between the spectral samples, are more effective at removing low frequency and baseline variations than filtering techniques which attenuate components below the cut-off frequency. It is thought that required spectral

information was attenuated as well as unwanted variations, leading to the poor performance.

Moving average filters which calculated the average of three or five adjacent spectral samples were applied for data preprocessing. The filters provided improvements in performance with spectral data with noise as the only form of interference but failed to improve the predictive ability of models with multiple forms of interference. The greatest accuracy attained was an SEP of 0.44 mmol/l.

Several researchers have obtained promising results using Fourier filtering techniques to remove high frequency noise and low frequency base-line fluctuations [14, 45, 52, 79]. Band-pass Fourier filtering using rectangular window functions was initially applied but failed to improve the performance of the calibration models. The use of Fourier filters with Gaussian window functions did, however, lead to the development of improved calibration models. An SEP of 0.40 mmol/l was attained using a Gaussian function with a mean of 0.125 and a standard deviation of 0.08.

PCA was used to determine the effect of data reduction on the predictive ability of the neural networks. PCA provides the potential benefit that information at a greater number of frequencies can be considered without the network complexity becoming too great. The redundancy in the simulated spectral data, meant that the majority of the variance could be described by 6-8 principal components. The use of PCA for data reduction resulted in only minor reductions in accuracy and reduced the time taken to train the networks by approximately 50%. MLP's with data preprocessed using 8 principal components resulted in an SEP of 0.45 mmol/l compared to the SEP of 0.43mmol/l when PCA was not used. Normalisation of the input data to have a mean of zero and a standard deviation of unity is frequently performed before PCA is applied. This did not improve performance, resulting in a SEP of 0.47 mmol/l. Using a combination of PCA and normalisation by closure led to an accuracy of 0.43 mmol/l.

The use of various combinations of preprocessing techniques to remove both high and low frequency interferences were attempted. The highest accuracy was attained by applying a combination of normalisation by closure and Gaussian Fourier filtering. Combining these techniques resulted in an SEC of 0.370 mmol/l and a SEP of 0.374 mmol/l. The noise rejection was better than any of the other processing techniques used and the generalisation was excellent which enabled the network to be trained for more cycles without over-fitting of the training data occurring. Using moving average filters along with normalisation by closure and moving average filter followed by first derivatives also provided acceptable results, with SEP's of 0.394 mmol/l and 0.401 mmol/l respectively.

7.9 Findings from the Simulations

The simulations in this chapter provide insight into several aspects of glucose measurement which have not been studied by previous researchers. This is the first study which makes use of artificial neural networks to determine glucose concentrations using spectral data from the combination and first overtone spectral regions. It also provides a comparison of the performance of neural networks trained with data from each of these regions under a number

of different conditions. The various forms of interference which affect spectroscopic glucose measurement are analysed individually to determine how they impact on the measurement process. A detailed investigation is performed into the use of data pre-processing techniques to reduce the effect of interfering spectral features. The suitability of using various pre-processing techniques, which have been successfully applied to other spectroscopic measurement problems, for the analysis of NIR spectral data for glucose measurement, is determined. This enables appropriate preprocessing techniques to be selected, thereby improving the performance of the calibration models.

The results from the simulations are promising as they suggest that NIR spectroscopic measurement in the combination or first overtone region could be successfully used for the measurement of blood glucose. The use of ANN's for the multivariate calibration provides sufficient selectivity to detect glucose levels in the presence of multiple interfering analytes and to differentiate between changes in glucose concentrations and changes in concentrations of other analytes. The ANN calibration models are capable of overcoming the three major forms of interference which are likely to occur during in vivo measurements, namely low frequency variations, multiplicative effects and high frequency noise.

The use of simulated spectral data enables the effects of individual forms of interference to be isolated and studied independently which is not a possibility for researchers performing in vivo measurements. This approach enables a better understanding of the factors influencing the spectral measurements to be gained than the empirical approach which the majority of researchers have favoured. The analysis of the various forms of interference has shown that high frequency spectral features, such as random noise, are more difficult to compensate for in the calibration model than low frequency variations.

The simulations in section 7.7 have shown that slightly better accuracy can be obtained when data from the combination region is used than when first overtone region data is used. This is thought to be due to the greater differences between the shapes of the glucose spectral peaks and those of other analytes. Networks trained using combination region data proved to be less susceptible to random noise. The performance of the MLP networks was superior to that of the RBF networks in both the first overtone and the combination regions.

The performance of neural network calibration models can be improved with the use of data preprocessing techniques. Techniques that remove low frequency and multiplicative variations provided greater improvements than those which compensate for high frequency noise. None of the techniques designed to remove the random noise could significantly improve the network performance. This is attributed to the signal-averaging effect of neural networks caused by the summations at the network nodes and the decentralised storage of information. The internal structure of the networks ensures that random variations are inherently minimised. Additional techniques to minimise the high frequency noise therefore provide little or no improvement. Normalisation by closure and Gaussian Fourier filtering proved to be the preprocessing techniques, that when used in conjunction with the neural networks, provided the calibration models with the greatest predictive ability.

Due to the greater complexity involved when performing measurements in human blood, it is expected that the performance of the calibration models would be worse than the results attained using the simulated spectral data. Even with significant deterioration in

performance, it is likely that the spectroscopic measurement technique will be able to make clinically relevant predictions, provided that the spectrometer can provide results with an acceptably high signal-to-noise ratio.

7.10 Limitations of Simulations using Computer-generated Spectral Data

The computer-based spectral data, used for the simulations in the preceding sections, provides a good representation of the spectral data which would be obtained if laboratory measurements were performed using an NIR spectrometer. There are, however, several factors which are not considered by the model which generates the simulated spectral data.

The effect of interferences, such as scattering and instrumentation drift, are not included in the model. Since the neural network calibration models have been able to compensate for many of the major sources of interference which affect spectroscopic measurements including baseline variations, low frequency fluctuations, high frequency noise and interferences cause by other analytes, it is expected that calibration models will be able to handle other sources of interference successfully.

The dispersion of NIR radiation will occur at interfaces where there is a change in refractive index. Since this model only aims to model the spectral absorbance of aqueous solutions, this effect is not considered. A consideration of the dispersion at interfaces would be required if a complete model was created to simulate the data which would be obtained during in vivo measurements.

Another shortcoming of the model is that only six components of human blood are considered. The analytes used in the simulations were selected due to their strong influence on the NIR spectrum of blood. The consideration of these components is therefore of importance for glucose measurement. Many other components with lower concentrations, that have not been considered for the simulated data also have spectral features in the NIR region which could interfere with the ability of the neural network calibration models to predict blood glucose levels. The good predictive ability of ANN's in the presence of the analytes used in the simulations suggests clinically relevant predictions could still be made in the presence of additional analytes.

Chapter 8

Towards In Vivo Measurement of Blood Glucose

The simulated results from the preceding chapter show that NIR spectroscopy could potentially be used for in vivo glucose measurement. Although this research provides information about the feasibility of using spectroscopic techniques for in vivo measurement, the experiments performed are not continuous and relate to measurements in blood rather than in tissue.

In order for a measurement device to provide clinically acceptable in vivo results, several additional aspects must be considered including further research into the effects of skin composition, light scattering and tissue properties on NIR measurements. Practical aspects such as the cost of the measurement, the reduction of accuracy of the device over time, the need for individual calibration for each user and the optimal choice of measurement site must also be taken into account. In order for the device to perform continuous glucose measurement rather than providing episodic results, further issues must be addressed including the speed with which the device can perform measurements, the processing power of the device and the size of the device.

This chapter discusses some of the major aspects which require further work in order for the long term goal of a continuous non-invasive glucose measurement to be achieved. Section 8.1 identifies aspects which require additional research and discusses further experimentation that is required before in vivo measurements can be performed. Section 8.2 discusses the specifications for various components of a spectroscopic glucose monitor based on the findings from the simulations in chapter 7. Section 8.3 discusses some of the additional issues which must be resolved before the long-term goal of developing a continuous non-invasive glucose monitor can be attained.

8.1 Recommendations for Future Work

The research performed provides insight into some of the key aspects relating to spectroscopic glucose measurement. There are, however, several other aspects which require further work

before the design of a clinically useful glucose monitor could be performed.

The ability of the calibration models to make accurate predictions with spectra from increasingly complex samples must be evaluated. The simulations performed in this project focus on the measurement of glucose in aqueous solutions containing the components in human blood which have a major effect on the NIR absorption spectrum. It is necessary to determine the effects of tissue on the NIR spectrum through experimentation with tissue phantoms and in vivo measurements. Burmeister et al. [41] have shown that in vivo spectra from human subjects can be accurately simulated with phantoms containing layers of fat, water and muscle. Since the NIR spectrum of animal fat and muscle tissue are almost identical to that of human tissue, animal tissue is adequate for use in the phantoms. The use of tissue phantoms allows for greater levels of reproducibility than performing in vivo measurements and allows experimental parameters to be controlled so that experiments can be performed in a systematic manner [41]. Once acceptable results can be obtained using the phantoms, the next experimental phase of the project would involve collecting spectral data from in vivo measurements. The specificity and sensitivity of the measurement device must be determined.

Investigations into techniques such as pulsatile spectrometry could be performed in an attempt to minimise the spectroscopic effects of tissue. This technique eliminates the role of complex non-pulsatile components by observing the changes in transmission which occur during an arterial pulse. This technique has been used successfully in pulse oximetry. Genetic Algorithms, or other optimisation techniques could be used to select wavelength ranges which are least affected by interfering spectra.

The issue of specificity of glucose monitors requires further research, as in vivo glucose measurement studies have not yet been able to determine whether the measured changes in the NIR absorption spectrum result from changes in glucose concentrations or from indirect measurements of other physiological factors [5, 14].

Investigations must be performed using spectral data from several patients to determine if it is possible to develop a universal calibration model, which can make accurate predictions on any patient, or if it is necessary to generate a calibration model for each user. Multi-patient calibration would require a thorough understanding of physical and physiological factors affecting measurements, differences between patients and the effects of noise [5].

A better understanding of the light propagation in tissue is required. Light transport can be modelled using techniques such as Monte Carlo simulations which model the optical path of individual photons and determine the probability of absorption or scattering. Absorption and scattering coefficients are dependent on several factors including glucose concentrations, water concentrations, temperature and scattering due to connective tissue fibres and erythrocytes [5].

Studies have shown that circulation and skin structural effects differ in diabetics and non-diabetics. The skin cannot be considered as a passive optical window for non-invasive measurements as skin properties will depend on the state of the disease and on environmental factors [5]. Time-dependent physiological effects in the human body could also potentially interfere with glucose measurement [5]. Further research into these issues is required.

The choice of the measurement site is a vital consideration. Several aspects must be taken into account when selecting a measurement site, including the optical pathlength, the composition of the tissue and the discomfort that placing the measurement device at this site may cause to the user. A compromise is required when determining a suitable pathlength as the noise levels are lowest with short pathlengths, but according to Beer's Law, the glucose spectral features are larger when longer pathlengths are used, resulting in better sensitivity. Burmeister and Arnold have evaluated various measurement sites and suggest that the highest signal-to-noise ratios can be obtained at measurement sites with a low percentage of body fat [53].

Development of a spectrophotometer which is customised to meet the requirements of NIR glucose measurement is also a priority. In order to meet the cost requirements, the device must be optimised to work over a narrow spectral range, but must provide sufficient resolution and noise rejection to provide clinically useful information. The requirements of the spectrometer are discussed in 8.2.

8.2 Requirements of a Spectroscopic Glucose Monitor

The work performed in the preceding chapters has provided insight into spectroscopic glucose measurement and allows many of the requirements and specifications for a glucose monitor to be stated.

The specifications of the spectrometer is a vital consideration, as without accurate spectral data, it will not be possible to make valid predictions of the glucose concentration. The simulations in chapter 7 suggest that clinically relevant accuracy can be obtained in both the combination and first overtone regions of the spectrum. The simulations suggest that the best accuracy can be obtained using the wavelength range $2.064\ \mu\text{m}$ to $2.360\ \mu\text{m}$. The use of multiple wavelength ranges, incorporating both these regions is a possibility and the development of a suitable spectrometer which could operate over this wider spectral range could be considered. Due to practical considerations such as the potential need for multiple radiation sources, and the increased cost and size of the spectrometer, the use of a wide spectral range is not recommended.

A trade-off must be made when determining the optimal resolution of the spectrometer. A high resolution should provide improved predictive ability, but will increase the cost and size of the device. The simulations in previous chapters have shown that a resolution of 5-10 nm should be sufficient to make clinically relevant predictions.

Since the use of sample preparation techniques to reduce the effects of interferences is not possible for in vivo measurements, the signal-to-noise ratio is a vital consideration when choosing a measurement site and developing the measurement equipment. Hazen suggests that a signal-to-noise ratio of greater than unity is required in order to make clinically accurate predictions [14]. Clinically relevant accuracy could be obtained during the simulations with a signal-to-noise ratio of 0.25. It is, however, expected that a greater signal-to-noise ratio would be required for in vivo measurement due to the greater complexity. Designing a spectrometer which can attain a signal-to-noise ratio of unity is therefore considered to be a

minimum requirement. The simulations in section 7.5 show that acceptable results can be obtained with an RMSN-100% value of 285 μ AU in the combination region and 193 μ AU in the first overtone region.

As discussed in section 8.1, the throughput and the magnitude of the spectral features must be considered when determining the optimal pathlength of the glucose spectral information. The simulations suggest that a pathlength of approximately 10mm is optimal for measurements in the first overtone region and a 2mm pathlength is recommended for measurements in the combination region.

The processing requirements of the device must be sufficiently low to allow for the use of small, low cost components. This can be achieved, since even though the process of training the neural networks is resource intensive, the resources required to execute the neural networks once training is complete, are low.

Other issues of long term importance for a continuous monitor are the size, comfort, aesthetics, price and power consumption. A device that can perform glucose measurements with clinically relevant accuracy will only be accepted by the public if the price is sufficiently low and causes only minimal inconvenience to the user.

8.3 Continuous In Vivo Glucose Measurement

Even though the research discussed in this report focusses on episodic measurements, the measurement technique can be adapted to continuous measurement. When a continuous monitor is developed, aspects such as the time taken to perform the measurements, drift of the measurement device with time and the usability of the device become more important.

The short period of time required to obtain results is one of the major advantages of using spectroscopic measurement techniques. The time taken to perform the calculations required to calculate the glucose concentration from the spectral data is also minimal. It should therefore be feasible to develop a measurement device which can calculate the blood glucose concentration in a few seconds. This is acceptable for a continuous glucose monitor.

A continuous monitor has the advantage that both the current glucose reading and the rate of change of blood glucose levels can be used to determine the course of action which the patient should take. In order for the correct treatment decisions to be taken, a continuous monitor must perform additional calculations which enable a suggested treatment to be provided. Size, weight, cost and comfort of the device are critical considerations if the device is to be accepted by diabetic patients. Methods of miniaturisation and cost reduction are required in order to develop a device which would be accepted by the market.

The sensor drift must be sufficiently low for the device to operate for extended periods of time without recalibration. The need for frequent recalibration with an invasive device would prevent a continuous device from being widely accepted by diabetic patients. Long term studies would be required in which the accuracy of the device is compared to readings from a conventional glucose monitor in order to determine whether there is a deterioration in performance.

Chapter 9

Conclusion

The use of NIR spectroscopy for the measurement of blood glucose levels could overcome many of the problems associated with conventional episodic “finger-prick” glucose monitors. The non-invasive nature of a spectroscopic measurement device ensures that measurements can be taken with minimal discomfort and inconvenience to the user. This promotes frequent monitoring of blood glucose levels which would help patients to achieve tight glucose control and delay the onset of the severe late complications of diabetes.

From the research performed, it appears that glucose measurement using NIR spectroscopic measurements is feasible, but there are many technical issues which must be overcome before the development of a home spectroscopic blood glucose monitor could be achieved.

The simulations suggest that glucose measurement in complex aqueous solutions can be performed with clinically relevant accuracy. The use of artificial neural networks for developing a multivariate calibration model, has proved to be an effective method of modelling the relationship between the spectral information available at various frequencies and the glucose concentration. The calibration models are capable of compensating for the effects of several different types of interferences, including those resulting from changes in the concentrations of other analytes, pathlength variations, temperature changes and high frequency measurement noise. The combination and first overtone regions of the NIR spectrum can both provide spectral information suitable for use in a glucose monitor. Neural networks trained using data from the combination region provided marginally better predictions and were less susceptible to the effects of interferences. The benefits of using pre-processing techniques to improve the performance of neural network calibration models have been determined.

Although these results suggest that NIR spectroscopic glucose measurement is promising, further work is required in order to determine the viability of performing in vivo measurements in human tissue. Further research must be performed to gain a better understanding of factors affecting the NIR spectrum of human tissue and to determine the long-term performance of spectroscopic measurement devices. Several issues must be resolved, including determining methods of improving the signal-to-noise ratio of spectroscopic measurements and reducing the size and cost of glucose measurement devices.

There are currently no NIR spectroscopic glucose monitors available which are approved by

the FDA or similar regulatory bodies for home glucose monitoring. Even though a large number of researchers are working on spectroscopic glucose monitors and much progress has been made in recent years, it is expected that it will be several years until a spectroscopic glucose measurement device, suitable for home glucose monitoring, is developed. Only once this is achieved, will it be possible to work towards the development of a continuous measurement device. Once non-invasive glucose monitors have reached the required stage of development, they will greatly improve the quality of life of diabetics and could form part of the long term aim of creating a closed loop insulin delivery system.

The NIR spectroscopic measurement and multivariate calibration techniques used to extract quantitative information from spectral data could be applied to many applications other than glucose measurement. These include similar biomedical applications, involving the measurement of other analytes in biological fluids, as well as industrial applications.

References

- [1] K. J. C. Wientjes, “Development of a glucose sensor for diabetic patients,” Ph.D. dissertation, Rijksuniversiteit, Groningen, Apr. 2000.
- [2] M. Ganz, “Diabetes mellitus: an oncoming avalanche,” in *New Approaches in Diabetes Care*, S. Päufer, Ed. Basel, Switzerland: F. Hoffmann-La Roche Ltd, Sept. 2002, ch. 1, pp. 7–15, Translated by David Playfair.
- [3] World Health Organization, “Diabetes: the cost of diabetes,” Sept. 2002, Fact Sheet No 236. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs236/en/print.html>
- [4] New England Healthcare Institute, “Continuous glucose monitoring: Innovation in the management of diabetes,” *NEHI Innovation Series*, Mar. 2005.
- [5] O. S. Khalil, “Non-invasive glucose measurement technologies: An update from 1999 to the dawn of the new millennium,” *Diabetes Technology & Therapeutics*, vol. 6, no. 5, pp. 660–697, 2004.
- [6] S. Laskowski and S. Balicki, *Anatomie normale du corps humain : atlas iconographique de XVI planches / par le docteur S. Laskowski ; dessinee?s dapre?s les pre?parations de lauteur par S. Balicki*. Geneva: Braun, 1894, courtesy of the National Library of Medicine [<http://www.nlm.nih.gov/exhibition/historicalanatomies>].
- [7] World Health Organization, “Diabetes mellitus,” Apr. 2002, Fact Sheet No 138. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs138/en/print.html>
- [8] P. Zimmet, K. G. M. M. Alberti, and J. Shaw, “Global and societal implications of the diabetes epidemic,” *Nature*, vol. 414, pp. 782–787, Dec. 2001.
- [9] M. Naylor, “Image:blankmap-world-continent-coloured.png,” June 2006. [Online]. Available: <http://commons.wikimedia.org/wiki/Image:BlankMap-World-Continents-Coloured.PNG>
- [10] R. W. Waynant and V. M. Chenault, “Overview of non-invasive fluid glucose measurement using optical techniques to maintain glucose control in diabetes mellitus,” *IEEE Lasers and Electro-Optics Society Newsletter*, vol. 12, no. 2, Apr. 1998.

- [11] K. Youcef-Toumi and V. A. Saptari, "Noninvasive blood glucose quantitation using spectroscopic-based optical technique," Mar. 1998. [Online]. Available: <http://darbelofflab.mit.edu/ProgressReports/HomeAutomation/98%20Reports/Youcef-Toumi.pdf>
- [12] R. Kotulla, "Towards an artificial pancreas," in *New Approaches in Diabetes Care*, S. Päufer, Ed. Basel, Switzerland: F. Hoffmann-La Roche Ltd, Sept. 2002, ch. 6, pp. 59–69, Translated by David Playfair.
- [13] L. C. Clark Jr. and C. Lyons, "Electrode systems for continuous monitoring in cardiovascular surgery," *Annals of the New York Academy of Sciences*, vol. 102, pp. 29–45, Oct. 1962.
- [14] K. H. Hazen, "Glucose determination in biological matrices using near-infrared spectroscopy," Ph.D. dissertation, University of Iowa, Iowa City, Iowa, Aug. 1995.
- [15] M. J. Tierney, "Transdermal glucose monitoring opens a new age of diabetes management," *IVD Technology*, May 2003. [Online]. Available: <http://www.devicelink.com/ivdt/archive/03/05/008.html>
- [16] G. L. Coté and R. J. McNichols, *Biomedical Photonics Handbook*. CRC Press, 2003, ch. 18, Glucose Diagnostics, pp. 18.1–18.19.
- [17] D. A. Gough and J. C. Armour, "Development of the implantable glucose sensor: what are the prospects and why is it taking so long?" *Diabetes*, vol. 44, no. 9, pp. 1005–1009, Sept. 1995.
- [18] J. J. Robert, "Continuous monitoring of blood glucose," *Hormone Research*, vol. 57, pp. 81–84, 2002.
- [19] A. P. Kretz and D. Styblo, "Toward continuous blood glucose monitoring," *Medical Device & Diagnostic Industry*, p. 78, June 2003.
- [20] D. C. Klonoff, "Noninvasive blood glucose monitoring," *Clinical Diabetes*, vol. 16, pp. 43–45, Jan. 1998.
- [21] J. Lambert, M. Storrie-Lombardi, and M. Borchert, "Measurement of physiologic glucose levels using raman spectroscopy in a rabbit aqueous humor model," *IEEE Lasers and Electro-Optics Society Newsletter*, vol. 12, no. 2, Apr. 1998.
- [22] J. J. Burmeister, H. Chung, and M. A. Arnold, "Phantoms for noninvasive blood glucose sensing with near infrared transmission spectroscopy," *Photochemistry and Photobiology*, vol. 67, pp. 50–55, Sept. 1998.
- [23] RamanRxn Systems, "Raman spectroscopy - an overview," Kaiser Optical Systems Inc, MI, USA, Raman Products Technical Note 1101, June 2002. [Online]. Available: www.kosi.com/raman/resources/technotes/1101.pdf
- [24] R. Liu, W. Chen, X. Gu, R. K. Wang, and K. Xu, "Chance correlation in non-invasive glucose measurement using near-infrared spectroscopy," *Journal of Physics D: Applied Physics*, vol. 38, pp. 2675–2681, July 2005.

- [25] R. Flewelling, *The Biomedical Engineering HandBook, Second Edition*. CRC Press LLC, 2000, ch. 86.
- [26] J. Severinghaus and P. Astrup, "History of blood gas analysis. VI. Oximetry," *Journal of Clinical Monitoring*, vol. 2, no. 4, pp. 270–288, Oct. 1986.
- [27] G. W. Small, "Data handling issues for near-infrared glucose measurements," *IEEE Lasers and Electro-Optics Society Newsletter*, vol. 12, no. 2, Apr. 1998.
- [28] M. A. Arnold, "Non-invasive glucose monitoring," *Current Opinion in Biotechnology*, vol. 7, pp. 46–49, 1996.
- [29] J. J. Burmeister and M. A. Arnold, "Spectroscopic considerations for noninvasive blood glucose measurements with near infrared spectroscopy," *IEEE Lasers and Electro-Optics Society Newsletter*, vol. 12, no. 2, Apr. 1998.
- [30] C. P. Sherman Hsu, "Infrared Spectroscopy," in *Handbook of Instrumental Techniques for Analytical Chemistry*, F. Settle, Ed. New Jersey: Prentice Hall, 1997.
- [31] R. Abbink and C. Gardner, "Getting under the skin," *SPIE'S OE magazine*, pp. 18–20, Sept. 2003.
- [32] M. Cope, "The application of near infrared spectroscopy to non invasive monitoring of cerebral oxygenation in the newborn infant," Ph.D. dissertation, Department of Medical Physics and Bioengineering, University College London, Apr. 1991.
- [33] J. Chen, M. A. Arnold, and G. W. Small, "Comparison of combination and first overtone spectral regions for near-infrared calibration models for glucose and other biomolecules in aqueous solutions," *Analytical Chemistry*, vol. 76, pp. 5405–5413, 2004.
- [34] J. T. Olesberg, "Non-invasive blood glucose monitoring in the 2.0-2.5 μ m wavelength range," 2001. [Online]. Available: <http://ostc.physics.uiowa.edu/~olesberg/research/talks/leos-2001-blood-glucose-monitoring.pdf>
- [35] National Institute of Standards and Technology, "NIST chemistry webbook," NIST Standard Reference Database Number 69, June 2005. [Online]. Available: http://webbook.nist.gov/cgi/cbook.cgi?Name=*beta-d-glucose*&Units=SI&cIR=on
- [36] J. W. Hall and A. Pollard, "Near-infrared spectrophotometry: A new dimension in clinical chemistry," *Clinical Chemistry Clin Chem*, vol. 38, no. 9, pp. 1623–1631, 1992.
- [37] P. S. Jensen, J. Bak, and S. Andersson-Engels, "The influence of temperature on water and aqueous glucose absorption spectra in the near- and mid-infrared regions at physiologically relevant temperatures," *Applied Spectroscopy*, vol. 57, no. 1, pp. 28–36, 2003.
- [38] A. K. Amerov, J. Chen, and M. A. Arnold, "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*, vol. 58, no. 10, pp. 1195–1204, Oct. 2004.

- [39] O. S. Khalil, "Spectroscopic and clinical aspects of noninvasive glucose measurements," *Clinical Chemistry*, vol. 45, no. 2, pp. 165–177, 1999.
- [40] G. G. Dull and R. Giangiaco, "Determination of individual simple sugars in aqueous solution by near infrared spectrophotometry," *Journal of Food Science*, vol. 49, pp. 1601–1603, 1984.
- [41] M. A. Arnold, J. J. Burmeister, and G. W. Small, "Phantom glucose calibration models from simulated noninvasive Human Near-Infrared Spectra," *Analytical Chemistry*, vol. 70, pp. 1773–1781, 1998.
- [42] M. A. Arnold and G. W. Small, "Determination of physiological levels of glucose in an aqueous matrix with digitally filtered fourier transform near-infrared spectra," *Analytical Chemistry*, vol. 62, pp. 1457–1464, 1990.
- [43] L. A. Marquardt, M. A. Arnold, and G. W. Small, "Near-infrared spectroscopic measurement of glucose in a protein matrix," *Analytical Chemistry*, vol. 65, pp. 3271–3278, 1993.
- [44] M. J. Mattu, G. W. Small, and M. A. Arnold, "Determination of glucose in a biological matrix by multivariate analysis of multiple band-pass-filtered fourier transform near-infrared interferograms," *Analytical Chemistry*, vol. 69, pp. 4695–4702, 1997.
- [45] G. W. Small, L. A. Marquardt, and M. A. Arnold, "Strategies for coupling digital filtering with partial least-squares regression: Application to the determination of glucose in plasma by fourier transform near-infrared spectroscopy," *Analytical Chemistry*, vol. 65, pp. 3279–3289, 1993.
- [46] K. H. Hazen, M. A. Arnold, and G. W. Small, "Measurement of glucose and other analytes in undiluted human serum with near-infrared transmission spectroscopy," *Analytica Chimica Acta*, vol. 371, pp. 255–267, april 1998.
- [47] F. M. Ham, G. M. Cohen, K. Patel, and B. R. Gooch, "Multivariate determination of glucose using NIR spectra of human blood serum," *Engineering in Medicine and Biology Society*, pp. 818–819, 1994.
- [48] F. M. Ham, G. M. Cohen, I. Kostanic, and B. R. Gooch, "Multivariate determination of glucose concentrations from optimally filtered frequency-warped NIR spectra of human blood serum," *Physiol. Meas.*, vol. 17, pp. 1–20, 1996.
- [49] F. M. Ham, I. N. Kostanic, G. M. Cohen, and B. R. Gooch, "Determination of glucose concentrations in an aqueous matrix from NIR spectra using optimal time-domain filtering and partial least-squares regression," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 6, pp. 475–485, June 1997.
- [50] M. Tarumi, M. Shimada, T. Murakami, M. Tamura, M. Shimada, H. Arimoto, and Y. Yamada, "Simulation study of in vitro glucose measurement by NIR spectroscopy and a method of error reduction," *Physics In Medicine And Biology*, vol. 48, pp. 2373–2390, July 2003.

- [51] K. Youcef-Toumi and V. A. Saptari, “Noninvasive blood glucose analysis using near infrared absorption spectroscopy,” Progress Report 2-4, MIT Home Automation and HealthCare consortium, Oct. 1999. [Online]. Available: <http://darbelofflab.mit.edu/ProgressReports/HomeAutomation/Report2-5/Chapter04.pdf>
- [52] —, “Noninvasive blood glucose analysis using near infrared absorption spectroscopy,” Progress Report 2-5, MIT Home Automation and HealthCare consortium, Mar. 2000. [Online]. Available: <http://darbelofflab.mit.edu/ProgressReports/HomeAutomation/Report2-5/Chapter04.pdf>
- [53] J. J. Burmeister and M. A. Arnold, “Evaluation of measurement sites for noninvasive blood glucose sensing with near-infrared transmission spectroscopy,” *Clinical Chemistry*, vol. 45, no. 9, pp. 1621–1627, 1999.
- [54] J. Burmeister, M. Arnold, and G. Small, “Noninvasive blood glucose measurements by near-infrared transmission spectroscopy across human tongues,” *Diabetes Technology and Therapeutics*, vol. 2, no. 1, pp. 5–16, 2000.
- [55] C. D. Brown, H. T. Davis, M. N. Ediger, C. M. Fleming, E. Hull, and M. Rohrscheib, “Clinical assessment of near-infrared spectroscopy for noninvasive diabetes screening,” *Diabetes Technology And Therapeutics*, vol. 7, no. 3, pp. 456–466, 2005.
- [56] Y. P. Du, Y. Z. Liang, S. Kasemsumran, K. Maruo, and Y. Ozaki, “Removal of interference signals due to water from in vivo near infrared (NIR) spectra of blood glucose by region Orthogonal Signal Correction (ROSC),” *Analytical Sciences*, vol. 20, pp. 1339–1345, Sept. 2004.
- [57] G. W. Hopkins and G. R. Mauze, “In-vivo NIR diffuse-reflectance tissue spectroscopy of human subjects,” Hewlett-Packard Company, CA, USA, Tech. Rep. HPL-1999-13, Jan. 1999.
- [58] S. F. Malin, T. L. Ruchti, T. B. Blank, S. N. Thennadil, and S. L. Monfre, “Noninvasive prediction of glucose by near-infrared diffuse reflectance spectroscopy,” *Clinical Chemistry*, vol. 45, no. 9, pp. 1651–1658, 1999.
- [59] M. Robinson, R.P.Eaton, D. Haaland, G. Koepp, E. Thomas, B. Stallard, and P. Robinson, “Non-invasive glucose monitoring in diabetic patients: a preliminary evaluation,” *Clinical Chemistry*, vol. 38, pp. 1618–1621, 1992.
- [60] H. M. Heise, A. Bittner, and R. Marbach, “Clinical chemistry and near infrared spectroscopy: technology for non-invasive glucose monitoring,” *Journal of Near Infrared Spectroscopy*, vol. 6, pp. 349–359, 1998.
- [61] H. Heise, A. Bittner, and R. Marbach, “Near-infrared reflectance spectroscopy for noninvasive monitoring of metabolites,” *Clinical Chemistry and Laboratory Medicine*, vol. 38, no. 2, pp. 137–145, 2000.
- [62] K. Jagemann, C. Fischbacher, K. Danzer, U. A. Muller, and B. Mertes, “Application of near-infrared spectroscopy for non-invasive determination of blood/tissue glucose using neural networks,” *Zeitschrift fur Physikalische Chemie*, vol. 191S, pp. 179–190, 1995.

- [63] P. Bhandare, Y. Mendelson, E. Stohr, and R. A. Peura, "Glucose determination in simulated blood serum solutions by Fourier transform infrared spectroscopy: investigation of spectral interferences," *Vibrational Spectroscopy*, vol. 6, pp. 363–378, 1994.
- [64] P. Bhandare, Y. Mendelson, R. Peura, G. Janatsch, J. Kruse-Jarres, R. Marbach, and H. Heise, "Multivariate determination of glucose in whole blood using partial least-squares and artificial neural networks based on mid-infrared spectroscopy," *Applied Spectroscopy*, vol. 47, no. 8, pp. 1214–1221, 1993.
- [65] F. M. Ham, G. M. Cohen, and B. Cho, "Neural network based real-time detection of glucose using a non-chemical optical sensor approach," *Annual International Conference of IEEE Engineering in Medicine and Biology Society*, vol. 12, no. 2, pp. 480–482, 1990.
- [66] —, "Glucose sensing using infrared absorption spectroscopy and a hybrid artificial neural network," in *Annual International Conference of IEEE Engineering in Medicine and Biology Society*, vol. 13, no. 4, 1991.
- [67] C. Lin, T. Hsiao, M. Zeng, and H. Chiang, "Quantitative multivariate analysis with artificial neural networks," in *2nd International Conference on Bioelectromagnetism*, Melbourne, Australia, Feb. 1998, pp. 59–60.
- [68] C. Fischbacher, K. U. Jagemann, K. Danzer, U. A. Mller, L. Papenkordt, and J. Schler, "Enhancing calibration models for non-invasive near-infrared spectroscopical blood glucose determination," *Fresenius Journal of Analytical Chemistry*, vol. 359, pp. 78–82, 1997.
- [69] P. Bhandare and Y. Mendelson, "Neural network based spectral analysis of multicomponent mixtures for glucose determination," in *Proceedings of the 1991 IEEE Seventeenth Annual Northeast Bioengineering Conference*, Hartford, CT, USA, Apr. 1991, pp. 249–250.
- [70] K. H. Hazen, M. A. Arnold, and G. W. Small, "Temperature-insensitive near infrared spectroscopic measurement of glucose in aqueous solutions," *Applied Spectroscopy*, vol. 48, no. 4, pp. 477–483, Apr. 1994.
- [71] C. Pasquini, "Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications," *Journal of the Brazilian Chemical Society*, vol. 14, no. 2, pp. 198–219, 2003.
- [72] F. Despagne and D. L. Massart, "Neural networks in multivariate calibration," *The Analyst*, vol. 123, pp. 157R–178R, 1998.
- [73] S. D. Brown, S. T. Sum, F. Despagne, and B. K. Lavine, "Chemometrics," *Analytical Chemistry*, vol. 68, no. 12, pp. 21R–61R, June 1996.
- [74] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems*, vol. 39, pp. 43–62, 1997.

- [75] J. R. Long, V. G. Gregoriou, and P. J. Gemperline, "Spectroscopic calibration and quantitation using artificial neural networks," *Analytical Chemistry*, vol. 62, pp. 1791–1797, 1990.
- [76] Y. C. Shen, A. G. Davies, E. H. Linfield, T. S. Elsey, P. F. Taday, and D. D. Arnone, "The use of Fourier-transform infrared spectroscopy for the quantitative determination of glucose concentration in whole blood," *Physics In Medicine and Biology*, vol. 48, pp. 2023–2032, June 2003.
- [77] Z. Wang, T. Dean, and B. R. Kowalski, "Additive background correction in multivariate instrument standardization," *Analytical Chemistry*, vol. 67, no. 14, pp. 2379–2385, July 1995.
- [78] A. Hagman and P. Sivertsson, "The use of NIR spectroscopy in monitoring and controlling bioprocesses," *Process Control and Quality*, vol. 11, no. 2, pp. 125–128, 1998.
- [79] S. Pan, H. Chung, and M. A. Arnold, "Near-infrared spectroscopic measurement of physiological glucose levels in variable matrices of protein and triglycerides," *Analytical Chemistry*, vol. 68, pp. 1124–1135, 1996.
- [80] S.-Y. B. Hu, M. A. Arnold, and J. M. Wiencek, "Temperature-independent near-infrared analysis of lysozyme aqueous solutions," *Analytical Chemistry*, vol. 72, pp. 696–702, 2000.
- [81] H. Martens and T. Naes, *Multivariate Calibration*. Chichester: John Wiley & Sons, 1991.
- [82] I. S. Helland, T. Naes, and T. Isaksson, "Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data," *Chemometrics and Intelligent Laboratory Systems*, vol. 29, pp. 233–241, 1995.
- [83] J. Forshed, F. O. Andersson, and S. P. Jacobsson, "NMR and Bayesian regularized neural network regression for impurity determination of 4-aminophenol," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 29, pp. 495–505, 2002.
- [84] K. Danzer, M. Otto, and L. Currie, "Guidelines for calibration in analytical chemistry part 2: Multispecies calibration," *Pure and Applied Chemistry*, vol. 76, no. 6, pp. 1215–1225, 2004.
- [85] W. L. Clarke, D. Cox, L. A. Gonder-Frederick, W. Carter, and S. L. Pohl, "Evaluating clinical accuracy of systems for self-monitoring of blood glucose," *Diabetes Care*, vol. 10, no. 5, pp. 622–628, 1987.
- [86] W. L. Clarke, "The original Clarke error grid analysis (EGA)," *Diabetes Technology and Therapeutics*, vol. 7, no. 5, pp. 776–779, Oct. 2005.
- [87] A. K. Amerov, J. Chen, G. W. Small, and M. A. Arnold, "Scattering and absorption effects in the determination of glucose in whole blood by near-infrared spectroscopy," *Analytical Chemistry*, vol. 77, no. 14, pp. 4587–4594, July 2005.

- [88] A. Brugger, K. Burton, A. Engeler, A. F. Essellier, L. P. Hollander, P. Jeanneret, D. B. Keech, H. L. Kornberg, H. Krebs, H. Levi, J. Lowenstein, H. Luthy, J. R. Quayle, and C. Rhonheimer, *Documenta Geigy Scientific Tables*. Basle, Switzerland: J. R. Geigy, 1962, ch. Composition of the Body, Body Fluids and Secretions, pp. 516–603.
- [89] D. A. Cirovic, “Feed-forward artificial neural networks: applications to spectroscopy,” *trends in analytical chemistry*, vol. 16, no. 3, pp. 148–155, 1997.
- [90] I. Nabney, *Netlab: Algorithms for Pattern Recognition*. New York: Springer-Verlag, 2001.
- [91] E. Stohr, P. Bhandare, R. A. Peura, and Y. Mendelson, “Quantitative FTIR spectrophotometry of cholesterol and other blood constituents and their interference with the in-vitro measurement of blood glucose,” in *Proceedings of the Eighteenth IEEE Annual Northeast Bioengineering Conference*, 1992, pp. 105–106.
- [92] H. Varley, A. H. Gowenlock, J. R. McMurray, and D. M. McLaughlan, *Varley’s Practical Clinical Biochemistry*, 6th ed. London: Heinemann Medical Books, 1988.
- [93] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley Interscience, 2001, ch. 6, pp. 3–54.
- [94] J. Coates, *Encyclopedia of Analytical Chemistry*. Chichester: John Wiley & Sons Ltd, 2000, ch. Interpretation of Infrared Spectra, A Practical Approach, pp. 10 815–10 837.
- [95] M. Volmer, “Infrared spectroscopy in clinical chemistry, using chemometric calibration techniques,” Ph.D. dissertation, Rijksuniversiteit Groningen, Sept. 2001.
- [96] Analytical Spectral Devices, Inc, “Introduction to NIR technology,” ASD.Document 600510 Rev. 2, Boulder, CO, USA, Mar. 2004.
- [97] G. Reich, “Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications,” *Advanced Drug Delivery Reviews*, vol. 57, pp. 1109–1143, 2005.
- [98] Analytical Spectral Devices, Inc., “Glossary of NIR terminology,” ASD.Document 600520 Rev. 1, 2003.
- [99] K. Xu, Q. Li, Z. Lu, and J. Jiang, “Fundamental study on non-invasive blood glucose sensing,” *Journal of X-Ray Science and Technology*, vol. 10, pp. 187–197, 2002.
- [100] D. Ozdemir and B. Ozturk, “Genetic multivariate calibration methods for near infrared (NIR) spectroscopic determination of complex mixtures,” *Turkish Journal of Chemistry*, vol. 28, pp. 497–514, 2004.
- [101] P. Bhandare, E. Stohr, Y. Mendelson, and R. Peura, “IR spectrophotometric measurement of glucose in phosphate buffered saline solutions: effects of temperature and pH,” in *Proceedings of the Eighteenth IEEE Annual Northeast Bioengineering Conference*, 1992, pp. 103–104.
- [102] R. Bro, “Multivariate calibration: What is in chemometrics for the analytical chemist?” *Analytica Chimica Acta*, vol. 500, pp. 185–194, 2003.

- [103] B. K. Lavine, “Chemometrics,” *Analytical Chemistry*, vol. 72, no. 12, pp. 91R–97R, June 2000.
- [104] B. K. Lavine and J. Workman, Jr., “Chemometrics,” *Analytical Chemistry*, vol. 74, no. 12, pp. 2763–2770, June 2002.
- [105] B. Lavine and J. J. Workman, Jr., “Chemometrics,” *Analytical Chemistry*, vol. 76, no. 12, pp. 3365–3372, June 2004.
- [106] V. S. Hollis, “Non-invasive monitoring of brain tissue temperature by near-infrared spectroscopy,” Ph.D. dissertation, Department of Medical Physics and Bioengineering University College London, Sept. 2002.
- [107] M. Sordo, “Introduction to neural networks in healthcare,” Oct. 2004. [Online]. Available: <http://www.openclinical.org/docs/int/neuralnetworks011.pdf>
- [108] S. Haykin, “Feedforward neural networks: An introduction,” Mar. 2004. [Online]. Available: <http://media.wiley.com/productdata/excerpt/19/04713491/04713491191.pdf>
- [109] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [110] R. D. De Veaux and L. H. Unger, “A brief introduction to neural networks,” Sept. 2002. [Online]. Available: <http://www.williams.edu/Mathematics/rdeveaux/papers/amstat.pdf>
- [111] J. A. Bullinaria, “Radial basis function networks,” Nov. 2004. [Online]. Available: <http://www.cs.bham.ac.uk/~jxb/NN/I12.pdf>
- [112] J. R. M. Smits, W. J. Melssen, L. M. C. Buydens, and G. Kateman, “Using artificial neural networks for solving chemical problems,” *Chemometrics and Intelligent Laboratory Systems*, vol. 22, pp. 165–189, 1994.

Appendix A

Basic Principles of Near Infrared Spectroscopy

Infrared spectroscopy is the study of the infrared spectra formed by the absorption of electromagnetic radiation at frequencies relating to the vibration of specific chemical bonds within a molecule [94]. The infrared region of the electromagnetic spectrum lies in the wavelength range between 750nm and 1000 μ m. The narrow band adjacent to the visible region of the spectrum is known as the near-infrared (750nm to 2500nm).

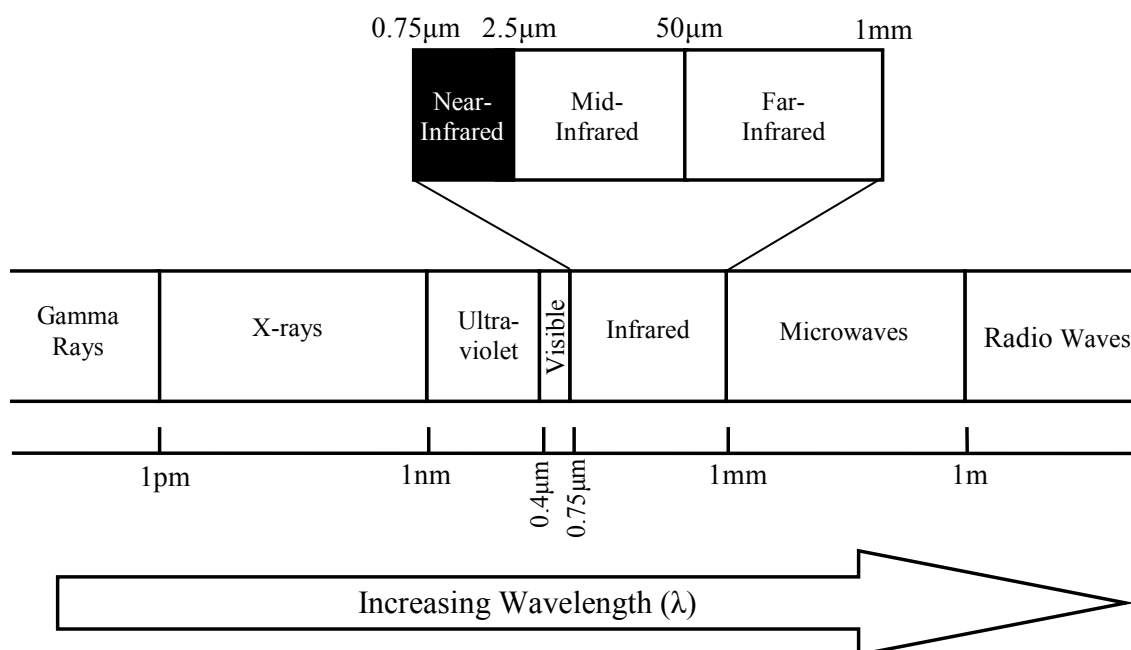


Figure A.1: The electromagnetic spectrum. Adapted from [95]

Infrared spectroscopy is frequently used in analytical chemistry for the study of organic compounds. The objective of infrared spectroscopy is to probe a sample in order to gain information from the interaction of near-infrared electromagnetic waves with its constituents

[71]. It's applications include qualitative identification of unknown compounds and quantitative measurements in which the quantities of known substances are determined [95].

A.1 Theoretical Models for IR Spectroscopy

A brief discussion of the theory relating to the origins of the infrared spectrum is given below. More detailed theoretical background is available in [71] and [94].

The total energy possessed by a molecule is defined as [94]

$$E_{total} = E_{electronic} + E_{vibrational} + E_{rotational} + E_{translational} \quad (\text{A.1})$$

The translational energy relates to the displacement of molecules as a function of the normal thermal motion of molecules. The rotational energy results from the absorption of microwaves and is observed as a tumbling motion of a molecule. The electronic energy relates to the energy transitions of electrons and is observed when visible or ultraviolet radiation is applied to the molecule [94].

The vibrational energy is the form which is of interest in infrared spectroscopy. The vibrational energy corresponds to the absorption of energy as the component atoms vibrate about the mean centre of their chemical bonds. The fundamental requirement for the absorption of infrared radiation to occur, is that a net change in dipole moment of the molecule must occur during the vibration [94].

The simplest model which can provide useful information about the origins of the near infrared spectrum is the model of the simple harmonic oscillator. A diatomic molecule can be considered to consist of two spherical masses separated by a spring with a force constant k . Application of Hooke's Law and Planck's Law shows that the energy of this system is given by [71]:

$$E = \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad (\text{A.2})$$

where,

h is the Planck constant

$\mu = \frac{m_1 m_2}{m_1 + m_2}$ is the reduced mass

m_1 and m_2 are the masses of the atoms in the molecule.

Since the energy of a photon (E_p) is given by [71]:

$$E_p = h\nu = \frac{hc}{\lambda} \quad (\text{A.3})$$

where,

c is the velocity of light

ν is the fundamental vibrational frequency.

It can therefore be shown that according to the classical model, the fundamental vibrational frequency is given by [71]:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (\text{A.4})$$

This model is capable of predicting the absorptions of diatomic molecules fairly accurately. It provides a link between the strength of the covalent bond (k) between the atoms, the mass of the interacting atoms and the frequency of vibrations [94]. It does not take into account the surrounding effects for polyatomic molecules, such as overlapping absorption spectra, hydrogen bonding, and repulsion and attraction of the electron clouds at the extremes of vibration. Bond dissociation that occurs at high energy levels and the quantum effects which only allow discrete energy levels, are not considered [96, 94, 71]. The harmonic model is not capable of predicting overtone and combination bands which means that, according to this model, most of the observable phenomena in the NIR region should not exist [71].

The anharmonic model overcomes many of the shortfalls of the harmonic model and provides a better prediction of the positions of the peaks in the near-infrared. The model still considers a molecular bond to consist of two spherical objects connected by a spring but also takes into account non-ideal behaviours which account for the repulsion between electron clouds, the variable behaviour of the bond force when the atoms move apart from each other and the rupturing of bonds when the atoms move far apart [71].

The Morse function approximates the anharmonic behaviour of diatomic molecules. It describes the potential energy of the molecule (V) as [71]:

$$V = D_e \left(1 - e^{-a(r-r_e)}\right)^2 \quad (\text{A.5})$$

where,

a is the constant for a given molecule

D_e is the spectral dissociation energy

r_e is the equilibrium distance between the atoms

r is the distance between the atoms at a particular instant.

Using quantum mechanics, the vibrational energy levels are described by [71]:

$$E = h\nu\left(v + \frac{1}{2}\right) - x_m h\nu\left(v + \frac{1}{2}\right)^2 \quad (\text{A.6})$$

where,

ν is the frequency of vibration

v is the vibrational quantum number

x_m is the anharmonicity constant of vibration ($0.005 < x_m < 0.05$).

The anharmonic model predicts the occurrence of transitions with $\Delta v \geq 2$ (overtones) and the existence of combination bands between vibrations. The fundamental vibration, which involves an energy transition from the ground state to the first vibrational quantum level, is affected very little by the inclusion of anharmonicity terms [71, 94].

The anharmonic model takes into account the interaction between vibrations. The total vibrational energy (E_ν) includes cross-terms from the various vibrations of the molecule [71]:

$$E_\nu = \sum h\nu_r(v_r + \frac{1}{2}) + \sum \sum hx_{rs}(v_r + \frac{1}{2})(v_s + \frac{1}{2}) + \dots \quad \text{for } r \leq s \quad (\text{A.7})$$

where,

ν_r is the frequency of vibrational mode r

v_r is the quantum number of vibrational mode r

x_{rs} is the anharmonicity constant for the interaction between vibrational modes r and s .

The theoretical models show that radiation of a certain frequency can be absorbed by a molecule leading to an excitation to a higher energy level. The radiation energy must match the energy difference between two vibrational levels of the molecule. This causes a selective response as the radiation at some wavelengths is absorbed, some is partially absorbed and some is not absorbed. The varied absorption at different wavelengths is responsible for forming the unique absorption spectrum of a particular molecule [71].

The energy match between the radiation and the vibrational levels only results in absorption if a change in dipole moment of the molecule, or a group of atoms within the molecule, occurs. The intensity of an absorption band depends on the magnitude of the dipole change and the degree of anharmonicity [71].

A.2 Features of the Near Infrared Spectral Region

The fundamental stretching and bending vibrations of organic molecules occur in the mid-infrared region (MIR) of the electromagnetic spectrum. The MIR region is characterised by relatively sharp absorption peaks and is commonly used for the identification of organic components [96, 95].

The NIR region of the spectrum is dominated by overtone and combination absorption bands. This is illustrated in figure A.2. The intensities of the absorption in this region are between 10 and 100 times lower than the absorption in the MIR resulting from the fundamental vibrations [71]. The NIR region is characterised by broad, super-imposed and weak absorption bands primarily due to overtone and combination bands of O-H, C-H, N-H

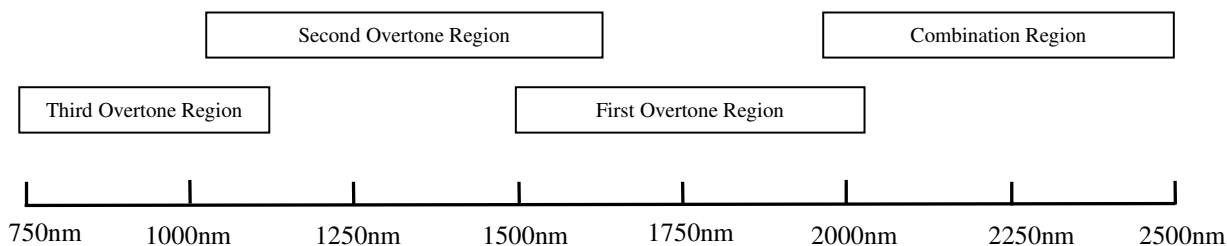


Figure A.2: The NIR region of the electromagnetic spectrum [96]

and C=O [71, 96]. Coupling and resonance effects, which are not described in this document, contribute to the complexity of the NIR spectrum [71].

The weak absorption in the NIR region offers several major advantages for non-invasive measurements; longer pathlengths can be used, minimal sample preparation is required, deep penetration is possible and measurements can be performed rapidly [95, 71]. The major disadvantage is that extracting useful information from the broad overlapping peaks of the NIR region is significantly more difficult than obtaining similar information in the MIR spectral region. Complex data handling techniques are therefore required. Scattering of light can also be problematic when NIR measurements are performed [96].

A.3 Measurement Modes

The appropriate NIR measurement mode to be used in an experiment will depend on the optical properties of the sample. An incorrect choice of measurement mode can lead to insufficient signal strength and a poor signal-to-noise ratio [97]. Three of the most commonly used modes are shown in figure A.3.

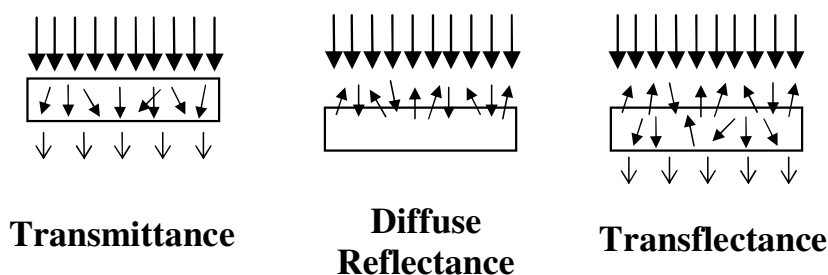


Figure A.3: Measurement modes for NIR spectroscopy [97]

Transmission measurements are performed by placing the sample in front of the light source and measuring the intensity of the light that passes through the sample. Materials with low absorptivities are usually measured using this technique. Diffuse reflectance measurements often provide a stronger signal when measurements are performed on turbid liquids and solids

which scatter light and absorb strongly. The radiation reflected off the sample is measured. Transflectance techniques measure both the reflected and transmitted radiation.

A.4 Instrumentation

NIR spectrophotometers are capable of providing sufficiently accurate spectral information for quantitative measurements to be performed. The most common forms of spectrophotometers are dispersive and Fourier Transform spectrometers but filter-based and LED-based instruments are gaining popularity due to their portability and suitability to low cost applications [71]. A brief discussion relating to dispersive and Fourier Transform spectrometers is given below.

A.4.1 Dispersive Spectrometers

A dispersive spectrometer contains three basic components: a radiation source, a monochromator and a detector [30]. Common detectors include silicon, PbS and InGaAs photoconductive materials. InGaAs materials have a particularly high detectivity and a fast response time [71]. These detectors are used in conjunction with high powered tungsten or halogen radiation sources in order to obtain the very high signal-to-noise ratios required for NIR measurements [71].

The monochromator is responsible for isolating a very narrow wavelength region [98]. It consists of gratings and prisms which are used in conjunction with variable-slit mechanisms or filters [98, 30]. The gratings or prisms are responsible for focussing a narrow band of frequencies on a mechanical slit. Narrow slits enable the detector to better distinguish closely-spaced frequencies of radiation resulting in good resolution, while wider slits enable more light to reach the detector and therefore provide better sensitivity [30].

Most dispersive spectrometers have a double-beam design. Two equivalent beams from the same source pass through the sample and reference chamber. An optical chopper focuses the reference and sample beams alternatively on the detector so that unwanted interferences are removed [30].

Dispersive spectrometers are less expensive than Fourier transform spectrometers. The main disadvantages are the slow scanning speed and a lack of wavelength precision [71].

A.4.2 Fourier Transform Spectrometers

Fourier Transform spectrometers examine all frequencies simultaneously rather than viewing each frequency sequentially like a dispersive spectrometer. Fourier Transform instruments offer superior speed, wavelength precision, signal-to-noise ratio and sensitivity to dispersive spectrometers [30, 71].

A Fourier Transform spectrometer makes use of an interferometer, rather than a monochromator. The interferometer divides radiant beams and generates a path difference between the beams. The beams are then recombined to produce repetitive interference signals. The interference signals are passed through the sample and infrared spectral information is picked up by the detector. The Michelson interferometer is used in the majority of Fourier Transform spectrometers [71].

Appendix B

Data Handling and Processing of Spectral Data

NIR spectra are composed of broad, overlapping absorption bands containing information from all sample components. Complex mathematical and statistical methods are required in order to extract relevant quantitative information from the spectral data while reducing the effect of interfering parameters [97].

Due to the complex nature of the NIR spectral region, univariate calibration models based, on information obtained from a single wavelength, are seldom useful in attaining quantitative information [71]. The measurement error for univariate calibration is large and a high demand is placed on the precision of the measurement instruments [99]. Univariate techniques are only capable of providing accurate readings in interference-free systems in which there is only one variable. They are unable to compensate for changes in temperature, pH or variations in the concentration of other analytes [100, 101].

Due to the problems associated with univariate techniques, the process of obtaining information from NIR spectra relies on the field of chemometrics and the use of multivariate calibration techniques. Chemometrics is the use of mathematical and statistical techniques to extract useful information from analytical data [71]. The reader is referred to [102, 73, 103, 104, 105] for discussions on the field of chemometrics. Multivariate calibration techniques are discussed below, these techniques overcome many of the problems associated with univariate techniques and can therefore be very useful for the quantitative analysis of NIR spectral data.

B.1 Multivariate Calibration

Several techniques have been developed that enable information at several different wavelengths to be used in the process of extracting quantitative information from NIR spectra. Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) are two of the most common calibration methods used for NIR spectroscopy. These two techniques are suited to situations in which information from a large number of wavelengths is

to be used as they avoid co-linearity problems and can therefore be used when the number of variables is greater than the number of available samples [71]. These two methods assume a linear relationship between the spectral data and the quantitative value which is to be measured [71]. PCR is discussed in section B.1.1 and PLS is briefly described in section B.1.1.

Artificial Neural Networks (ANN) are emerging as an alternative to PCR and PLSR for NIR calibration. ANN's are not as widely used as PCR and PLSR but show promise in certain applications since they may provide better results when a non-linear relationship exists between the spectral data and the quantitative value of interest [71]. Background information relating to ANN's is given in section B.1.2.

B.1.1 Linear Calibration Techniques

Principle Component Regression

Principle Component Regression (PCR) is a calibration technique that aims to reduce the dimensionality of the data set by discarding parameters which contribute noise or redundant information while keeping those that contribute the majority of the useful information [106, 81]. Since NIR spectral data contains many correlated variables, it is possible to reduce the number of variables and describe the data using fewer uncorrelated variables which contain the relevant spectral information [97].

The first step in PCR is known as Principal Component Analysis (PCA). PCA resolves the spectral data into orthogonal components whose linear combination approximates the original data. The original data set is compressed into a smaller number of variables known as principle components [97, 106]. These principle components correspond to the largest eigenvalues of the co-variance matrix and therefore contain the largest possible variance in the data set. The first principal component will represent the maximum variance among all linear combinations and all subsequent components will represent the largest possible portions of the remaining variability [97].

The transformation procedure is represented graphically in figure B.1 using a simple spectrum containing three wavelengths. Figure B.1.a shows the original spectrum with three wavelengths. The spectrum is transformed to a new co-ordinate system in which the spectrum can be represented as a single point in three dimensional space (figure B.1.b). Figure B.1.c represents a set of spectral data in the new co-ordinate system. Figure B.1.d shows the mean centring stage and figure B.1.e illustrates the development of the principal components [97]. The reader is referred to [81] for a detailed explanation of PCA.

PCA can be described mathematically by:

$$A = S \cdot L + E_A \quad (\text{B.1})$$

where A is the $(n \times m)$ data set comprised of many independent variables, S is an $(n \times h)$ matrix, L is an $(h \times m)$ matrix, E_A is the residual matrix and h is the number of principal

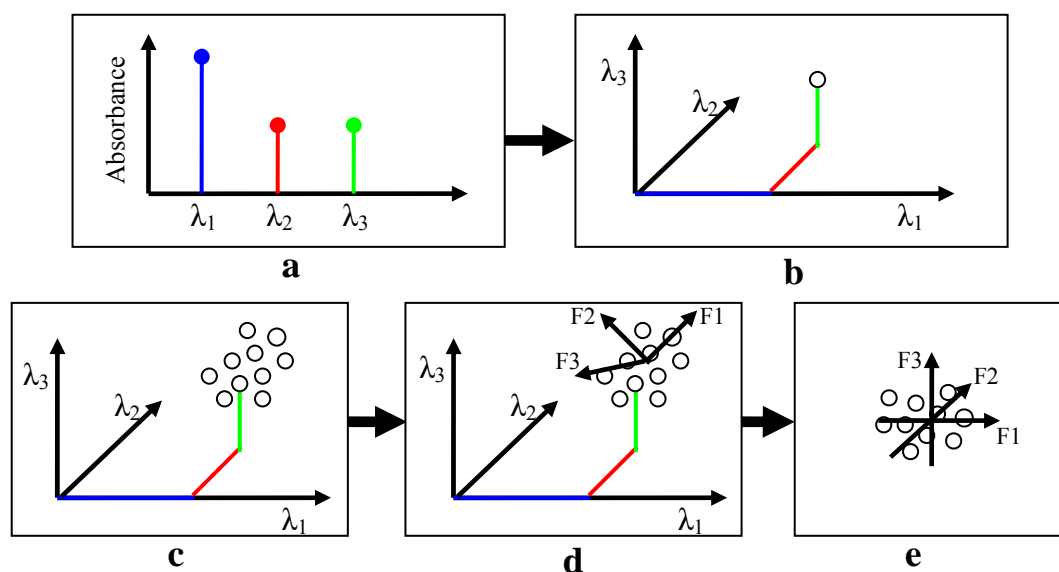


Figure B.1: PCA transformation procedure [97]

- The original spectrum with data points at three wavelengths
- The spectrum represented as a single point in the transformed co-ordinate system
- A set of spectral data in the transformed co-ordinate system
- The mean centring stage
- Development of the principal components

components. The column vectors of S are known as the scores and the row vectors of L are called the loadings of the principal components [106].

Once the Principal Components Analysis has been performed, the calibration model is generated by establishing a linear relationship between the PCA scores and the dependent variable of interest [106].

Partial Least Squares Regression

Partial Least Squares Regression (PLSR), like PCR, is a bilinear calibration technique that composes the data matrix into two smaller matrices. PLSR differs from PCR in its approach to the reduction of the original data set [106]. The major difference between the two techniques is that the PLS method generates loading vectors that find the direction of greatest variability by comparing the spectral variables and the target property information, while the loading vectors in PCA only explain the variance in the spectral variables [97, 27].

The objective of PLSR is to improve the correlation between the dependent variables and the spectral scores, rather than focussing on the minimisation of the residuals in equation B.1. This makes PLSR more robust than PCR. PLSR performs the calibration process in one step while PCR requires a two stage process, this ensures that the probability of discarding useful information in PLSR is reduced [106].

Certain studies have shown that PLSR and PCR provide similar prediction performance

when the optimal number of principal components are used [71] while others claim that the performance of PLSR is better for NIR calibration [106, 62]. PLSR usually requires a lower number of components in order to generate a good calibration model [71]. A detailed mathematical discussion of PLSR is beyond the scope of this document, the reader is referred to [81] for further information.

B.1.2 Artificial Neural Networks

An Artificial Neural Network (ANN) is an information processing system inspired by the structure and operation of biological nervous systems [107]. Neural networks consist of a large number of interconnected processing elements known as neurons. These neurons process information in parallel in response to external stimuli [107]. The network obtains knowledge through a learning process. The weights of the interconnections of neurons are adjusted in order to store this knowledge [108]. Neural networks that use a supervised learning scheme are well-suited for calibration purposes [62].

Neural networks are particularly useful for finding the relationship between the inputs and outputs of non-linear systems. In spectroscopic applications, the assumptions which lead to the creation of linear models, are frequently violated. Real or apparent non-linear spectral responses occur due to the instrumental, physical and chemical properties of the system. Under these circumstances, the non-linear modelling performed by ANN's can provide better results than linear techniques such as PCR and PLSR [62].

The two most common neural network topologies are multi-layer perceptron and radial basis function networks. A brief description of these techniques is given in the sections which follow. A discussion of the procedure of implementing neural networks is also given.

Multi-Layer Perceptrons

The multi-layer perceptron (MLP) architecture is the most widely used neural network topology [90]. MLP's make use of a layered feed-forward topology. Each layer of the network consists of several basic processing units known as neurons. Each neuron receives an input signal which it manipulates before outputting a signal to neurons in the subsequent layer [107]. This leads to a one-way flow of information through the system. The operation of a neuron, the basic unit from which the network is constructed, is given in figure B.2.

The synaptic weights represent the strength of the connection between two neurons. These weights are adjusted during the training process so that the optimal values can be determined. The bias is a constant value which can modify the input value by a fixed amount. The neuron in an MLP network can be described mathematically by equation B.2 [109].

$$y = \Phi \left(\sum_{i=0}^n w_i x_i + \theta_k \right) \quad (\text{B.2})$$

where x_1, \dots, x_n are inputs, θ_k is the bias, y is the output and Φ is the activation function.

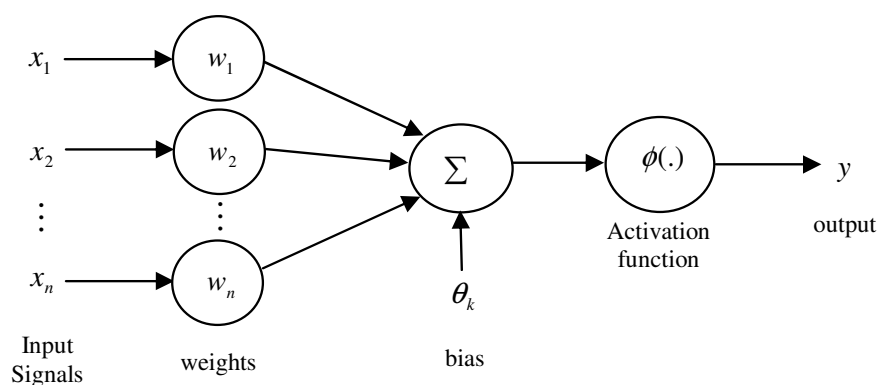


Figure B.2: A single neuron [109]

An MLP network consists of an input layer, one or more hidden layers and an output layer [108]. The role of the input layer is to receive the input stimuli and propagate the information to the first hidden layer. The hidden layers receive a biased weighted sum of the inputs and process them using an activation function [107]. Commonly used activation functions include the saturation, sigmoid and hyperbolic tangent function [90]. The output layer receives a biased weighted sum of the output from the last hidden layer and processes it by means of an activation function. A linear output activation function is commonly used, although softmax and logistic functions are also used in certain applications [90]. MLP's with only one hidden layer are normally used. It can be shown mathematically that an MLP with one hidden layer is capable of modelling a system of arbitrary complexity [109]. A simplified diagram of an MLP network with one hidden layer is shown in figure B.3.

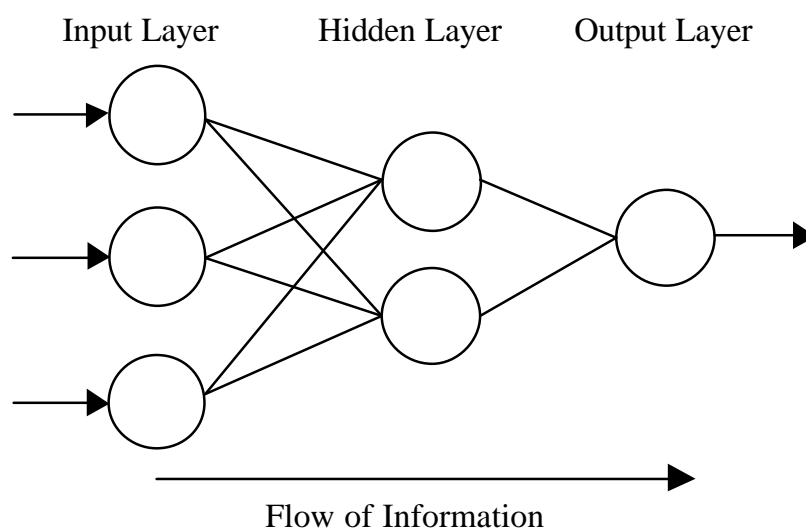


Figure B.3: Multi-layer perceptron network with one hidden layer [107]

A training process is necessary in order to teach the ANN how to perform a particular task. During the training process, the weights and biases are adjusted iteratively in order to minimise an error function [109]. In a supervised learning paradigm, the learning technique most applicable to chemometric applications, the network is supplied with the input data and

the corresponding output data. The network makes use of a suitable optimisation algorithm to minimise the error between the output predicted by the network and the target output. Popular optimisation algorithms include the Scaled Conjugate Gradient, Quasi-Newton and Hybrid Monte-Carlo algorithms [90]. A well trained network will be able to make accurate predictions of the output when new input data is supplied to the system.

Radial Basis Function Networks

Radial Basis Function (RBF) neural networks are a popular form of layered feedforward network partially inspired by the receptive fields found in animal vision systems [110]. An RBF network consists of an input layer, a single hidden layer and an output layer. The hidden nodes are known as basis functions. They compute an activation function on the inputs received from the input layer. The activation function in the hidden layer calculates the Euclidean distance between the input signal vector and the parameter vector of the system [108]. The activation function of the output layer is linear. Bias parameters can be introduced into the output layer to compensate for constant differences between the predicted value and the target value. A diagram of an RBF network is given in figure B.4 [108].

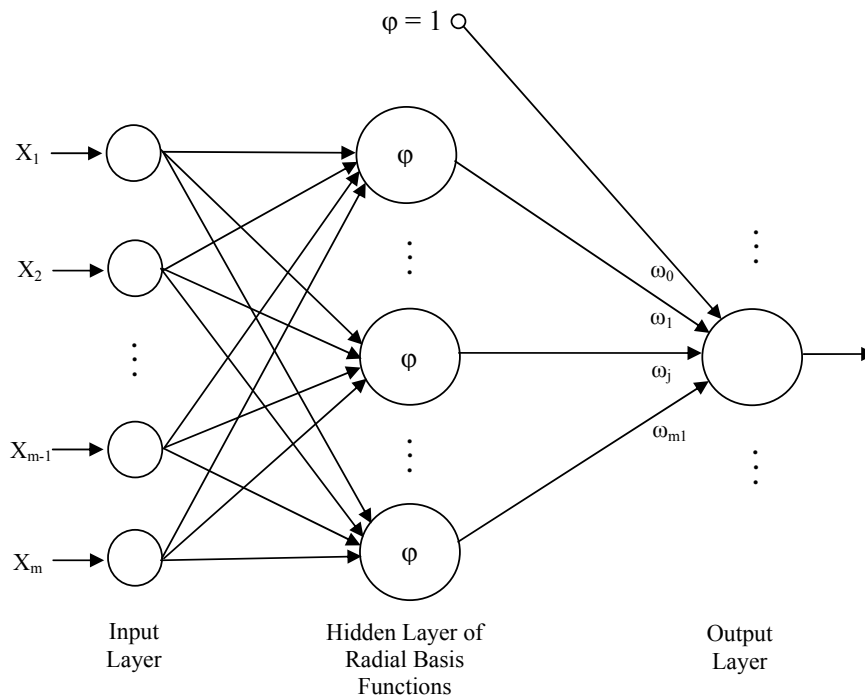


Figure B.4: Radial basis function network [108]

The RBF network can be described mathematically by equation B.3 [111]:

$$y_k(x) = \sum_{j=0}^M w_{kj} \phi_j(x) \quad (\text{B.3})$$

where y is the output, M is the number of hidden units, x is the input and $\phi(\cdot)$ is the activation function.

The Gaussian activation function is given by

$$\phi_j(x) = \exp\left(-\frac{\|x - \mu_j\|^2}{2\sigma_j^2}\right) \quad (\text{B.4})$$

where μ_j are the basis centres and σ_j is the standard deviation.

The learning process for RBF's is equivalent to finding a curve in multi-dimensional space which best approximates the training data. Multi-dimensional interpolation is used during the training procedure [108]. RBF's are trained using a two-stage process. In the first stage, the weights from the input to the hidden layer are determined. The second stage involves the determination of the weights of the connections between the hidden layer and the output layer using least squares regression [111].

Procedure for implementing ANN's

In order for the neural network to perform adequately, it is necessary for the input data to be converted to a form which can be easily interpreted by the network. Data pre-processing is used for this purpose. The data pre-processing stage will involve examining the data to ensure that no outliers are present. Outliers may occur when errors are made during the process of obtaining the data or due to poor performance from the measurement equipment[72]. The presence of these erroneous data points could greatly decrease the performance of the network. Scaling of the input data is often used to ensure that certain inputs are not given greater importance by the network due to them having greater numerical values [93]. In the case of spectroscopic measurement, some regions of the spectrum may have poor signal-to-noise ratios due to high absorption of water. Removal of data from these regions before training commences may improve the performance of the network. Filtering of the input data can be useful to remove the unwanted spectral noise [72].

Once the data has been pre-processed, partitioning of the data must be performed. The purpose of the partitioning phase is to ensure that the network is thoroughly changed without over-training occurring. If the network is not trained for a sufficiently long period of time, the network will have poor predictive abilities. Over-training the network will lead to over-fitting of the source data resulting in poor performance when unseen data is applied to the network [72, 93, 74, 112].

The partitioning phase usually involves dividing the data into 3 sets, a training set, a validation set and a testing set. The training data is used by the supervised learning algorithm. The network makes use of known input-output pairs to adjust the weights of the connections appropriately. The optimal number of nodes in the hidden layer of the network, the optimal number of iterations of the training algorithm and the most effective activation function must be determined. The network will use optimisation techniques to minimise the error between the predicted output and the target output. The validation data is used periodically to ensure that the over-training does not occur. Initially, the error between the expected

and actual output when the validation data is applied to the network, will decrease. As the network begins to over-fit the training data, this error will begin to increase. The use of the training data thereby ensures that training is stopped at the correct time. The testing data is used to once the training process is complete, to determine the predictive ability of the network when unseen data is provided.

In situations where the amount of available data is limited, techniques such as cross-validation can be used. This technique does not require the validation data set during the training process. The technique is, however, not as effective at stopping the training process at the optimal time and the apparent accuracy will not necessarily match the accuracy attained when new data is applied [72, 74].