University of Nevada, Reno

Development of a Software Tool to Perform Qualitative Analysis on Spectral Data Obtained from a NIR-MOEMS Spectrometer

A thesis submitted in partial fulfillment of the

requirements for the degree of Master of Science in

Electrical Engineering

by

Tejbir Singh Bajwa

Dr. Banmali Rawat/Thesis Advisor

May, 2009

University of Nevada, Reno
Statewide • Worldwide

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

TEJBIR SINGH BAJWA

entitled

Development Of A Software Tool To Perform Qualitative Analysis On Spectral Data Obtained From A NIR-MOEMS Spectrometer

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Banmali Rawat, Ph.D., Advisor

Bruce Johnson, Ph.D., Committee Member

Carl Nesbitt, Ph.D., Graduate School Representative

Marsha H. Read, Ph. D., Associate Dean, Graduate School

May, 2009

Abstract

Quality analyses are a vital feature in process monitoring at all production stages. The objective of this thesis is to develop a software tool which can qualitatively analyze spectral data obtained from a Near Infra Red Micro Opto Electro Mechanical Systems (NIR-MOEMS) spectrometer. NIR spectroscopy is a cost effective, non evasive and faster method for analyzing the samples. NIR-MOEMS spectrometer can be used to measure the spectra of samples at any stage of sample manufacturing. Quality analyzing of the samples requires comparing of an unknown sample with samples which are already known. So the software tool developed qualitatively analyzes the data available from the spectrometer at less cost, high speed and using various mathematical operations as pretreatment methods. The software developed requires libraries to be made of known samples and unknown samples are compared with known samples stored in these libraries. Some of the advantages of this software tool over other tools are that this tool can convert the reflection data obtained from NIR-MOEMS spectrometer to absorption data. The software tool is also able to select the wavelength range for comparison depending on user choice provided that this range is a subset of original range obtained from the spectrometer. Libraries of known samples can be made at different wavelength ranges and can be used at any stage of comparison. Future work will focus on development of the tool so that it can be used more dynamically and better methods of data pretreatment are incorporated.

Keywords: NIR spectroscopy, MOEMS spectrometer, Savitzky-Golay Differentiation, Principle Component Analysis, Cluster Analysis.

Acknowledgement

I would like to thank my adviser Dr. Banmali Rawat for his constant encouragement and suggestions in completing my research work without which this project would not have reached its completion. I would also like to thank him for his extraordinary guidance and support over the years spent in this university.

I would like to thank Prof. Thomas Gessner and Prof. Thomas Otto of Fraunhofer Institute for Reliability and Micro-integration (IZM), Chemnitz, Germany for providing me with research experience as an intern. I would also like to thank Alexander Weiss of Technical University Chemnitz, Germany, for all his support and guidance.

I sincerely thank Dr. Bruce Johnson and Dr. Carl Nesbitt for agreeing to be members of my thesis committee. I am also thankful to the Department of Electrical and Biomedical Engineering of the University of Nevada, Reno for accepting me as a student.

I would like to thank my family for their love and support. And above all, I would like to thank God for making me successful in my field of study.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Quality analysis is one of the most important features of Near Infra Red (NIR) spectroscopy. For analytic chemistry, quality analysis has become an important part. The use of NIR together with the new instrumentation technology has become a leading approach for analyzing samples. Quality analysis involves knowledge of the unknown sample by comparing its spectra with the spectra of already known samples stored in the library. NIR spectroscopy is one of the best tools for measuring composition, purity and even for finding the unknown sample. NIR spectroscopy has emerged as highly cost effective, non invasive and faster method for quality analysis [1]. NIR spectroscopy has gained a great deal of attention in quality analysis due to its ability to record spectra of solid and liquid samples at low cost without any pretreatment of samples or without utilizing any destructive methods. The identification of an unknown substance with NIR spectroscopy is based on a comparison between the spectral data of the unknown substance being analyzed and the spectral data of multiple samples of batches in a reference library as similar substances have similar spectra [2]. To compare data and reach conclusions, chemometrics should be used. For the characterization of raw materials, the technique has been found to have some unique benefits. NIR spectroscopy enables analysis of the starting material in its original packaging, without the need to open the main container. Thus it reduces the risk of cross-contamination and eliminates the need for conducting analysis within a designated sampling area. This leads to a significant decrease in time involved for each analysis [3]. Important applications of quality analysis are in pharmaceuticals industry, identification of textiles, identification of polymers, identification of oil quality, medical science, food and beverages etc [4].

MOEMS (Micro-Opto-Electro-Mechanical-Systems) technology has not only helped in miniaturizing analytical device used for NIR spectroscopy but has also helped in making spectrometry more viable and cost effective. These MOEMS based spectrometer have become an alternative to large and expensive FTIR (Fourier Transform Infra Red) spectrometers which are based on diode array technology and are quite complex. MOEMS spectrometer is based on the principle of dispersive grating. The central component of this type of spectrometer is the scanning micro-mirror whose design has been optimized for the spectrometer system. This makes the MOEMS based spectrometer more compact and can be used for in-situ analysis of the materials. The MOEMS based spectrometer also uses two detectors at the output to collect light which has been scattered by the mirror inside which increases the sensitivity of the device. Due to better sensitivity of NIR over MIR (Mid Infra Red) cooling of the device is generally not required resulting in lowering the cost and the size of the device. The dual detector NIR MOEMS based spectrometer offers versatility, accuracy, high speed and sensitivity over wide spectral range [1].

The MOEMS based spectrometer helps us to record readings in ASCII format. These readings have large amount of data in it. So it becomes important to evaluate this data so that we can take out meaningful information out of these set of readings. The software tool helps in using this information to obtain meaningful results. The software performs various functions and pre-treatment on the data that is measured from NIR-MOEMS spectrometer.

Some of the important features of this software tool not found in other tools are that this tool is able to average if more than one measurement is taken for each sample. This helps in removing the variations that occur during measurement of samples. It is also able to convert the reflection readings obtained from the spectrometer to absorption readings. Another unique feature of this software tool is that it is able to select a wavelength range of our interest. Some times the original range of data contains noise at the beginning and at the end. To overcome this problem the easiest way is to discard the data in this range. This is done by selecting the wavelength range of interest. This range is a subset of the original range. This tool also helps us to make and store libraries made at different wavelength ranges. This is important as it helps in saving time. We can select any of these libraries and compare samples at any or all of these ranges. We can add and delete samples in the libraries whenever the need arises. We can make comparison of unknown samples with known samples stored in the library in different ways. These methods have been discussed in the following chapters. The second chapter explains the fundamentals of NIR spectroscopy. It discusses the basic principle of NIR, instrumentation, advantages and disadvantages. Third chapter discusses the principle, specifications and applications of NIR-MOEMS based spectrometer. Fourth chapter analyzes mathematically the multivariate techniques used to develop the software tool. Fifth chapter discusses the working of the software tool with the help of flow charts and also discusses its salient features. Chapter six presents the results of sample comparison using different methods included in the software at different wavelength ranges. It also compares some of the results with the results obtained from other software tools for establishing the validity of the software tool developed. Chapter seven concludes the research work completed in this thesis and also suggests some future work required to upgrade the software tool.

Chapter 2

Fundamentals of NIR-Spectroscopy

2.1 Introduction

Near Infrared (NIR) spectroscopy is one of the best practical techniques in industry to be used for material analysis. This technique works in near infra red spectrum range from 700 nm to 2500 nm wavelength. NIR spectroscopy offers a practical alternative to time consuming, solvent intensive wet-testing methods and liquid chromatography techniques. NIR spectroscopy is an ideal technique for quick and reliable raw sample analysis. It is a powerful analysis tool for accurate multi-component quantitative analysis. NIR is also ideal for process monitoring due to its non destructive nature and does not require any sample preparation. It can sample through glass and packaging. Remote sensing of NIR spectroscopy can also be done using optical fiber.

Infrared region was discovered in 1800 by William F. Hershel's and stated in his famous work, "Experiments on the Refrangibility of the Invisible Rays of the Sun". In this experiment he used a glass prism to disperse sunlight onto three thermometers with carbon blacked bulbs. The heating effect became apparent towards the red spectrum but he noticed the maximum heating effect just after the red spectrum where there was no visible light [5]. In 1881 Abney and Festing recorded the spectra on a photographic film and noted that the spectra were related to the presence of hydrogen atoms. They recognized the importance of hydrogen bonding in NIR region. Till now NIR spectroscopy suffered from lack of calibration methods and unavailability of computers to analyze. However in 1960's Karl Norris used NIR spectroscopy in the field of agriculture and food products which is considered as one of the pioneer works in this field. Development of chemometrics techniques have made it possible for NIR spectroscopy to be applied to a wide variety of fields.

2.2 Basic Concept and Analysis of NIR spectroscopy

NIR spectroscopy depends upon absorption occurring from molecular overtones and combination vibrations in the infrared region. These overtones are harmonics making the NIR spectra quite complex. Most of the bonding occurring in NIR spectra is due to X-H stressing modes. The formation of hydrogen bonds changes the bond lengths and the bond strength. This formation of hydrogen bonds leads to a shift in the absorption peaks.

Two main laws that form the basic of vibrational spectroscopy are Hooke's law and Frank Condon Principle. Hooke's law is stated as [6]

$$F = -kx \tag{2.1}$$

where F is the restoring force exerted by a material, k is spring constant (or force constant) and x is distance moved (compressed or stretched). For a two body harmonic oscillator, Hooke's law gives the frequency of vibration as

$$v = \frac{1}{2\pi c}\sqrt{\frac{k(m1+m2)}{m1*m2}} \tag{2.2}$$

where frequency of vibration, v is in cm^{-1} a the reciprocal of 2π times the speed of light (c) multiplied by the square root of force constant (k) times the sum of two masses divided by their product as shown in Equation 2.2 [6].

This law is only used to calculate the fundamental vibrations of diatomic molecules in mid infrared region where as NIR consists of both fundamental vibrations and molecular overtones. For NIR, Franck Condon principle is more suitable. Franck Condon principle is stated as “When a molecule vibrates, the probability of finding an atom at a certain point is inversely proportional to its velocity when it is at that point. Therefore the atoms in vibrating molecule spend most of their time in configurations where the kinetic energy is low [7].” This means that the total energy is equal to the sum of both kinetic energy and potential energy and since the potential energy is so large that the total energy is almost equal to the potential energy. Thus the photon is most likely to be absorbed when the nuclei are stationary or are moving slowly. Furthermore the excitation resulting from the absorption of photon cannot be transferred immediately to the nuclei because absorption is almost an instantaneous process as it involves the rearrangement of inertia free electrons. But it is not the same as in the case of nuclei which are heavier than the electrons. Thus nuclei will tend to continue moving slowly immediately even after the absorption process. Since light absorption takes place in femoseconds, it is possible for the electrons to move in this time but the nuclei cannot move because of their weight. Since the nuclei have no time to readjust, they have to adjust when the absorption process is over. This readjustment causes them to vibrate [6].

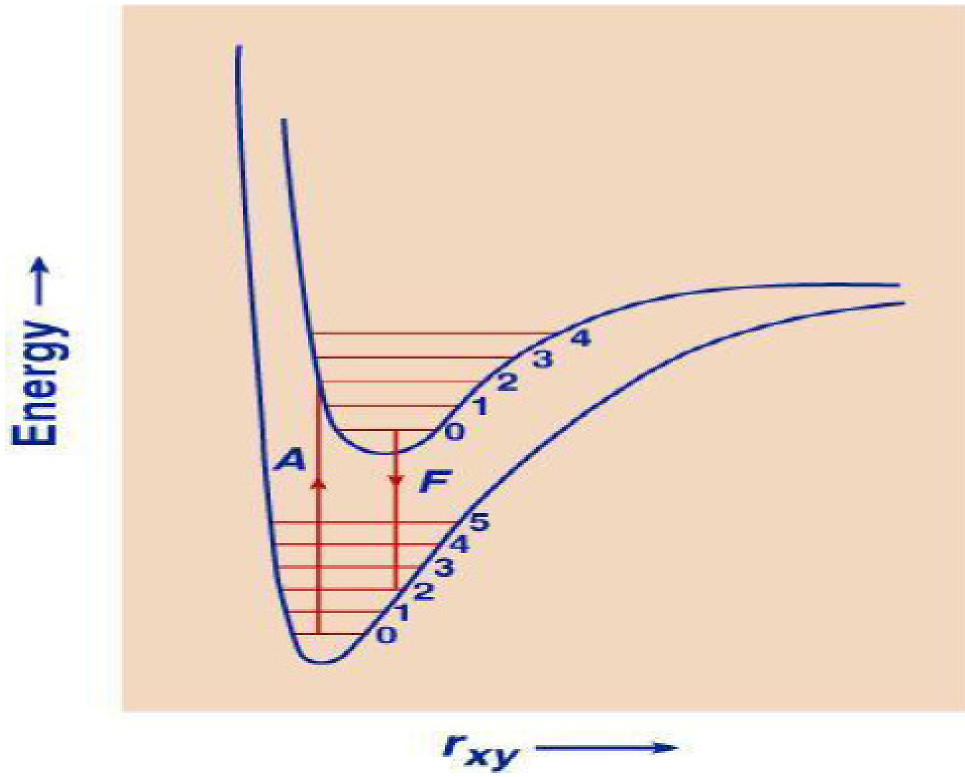

Fig 2.1 Potential energy curves for both ground as well as excited state for diatomic molecule .

As shown in Figure 2.1 [8], the distance between the nuclei, r_{xy} is represented along the x axis. The two curves represent the potential energy of the molecules both in ground state as well as in excited state. Here the vertical arrow A marks the excitation state. This arrow hits the upper curve of the excited case, unlike in special cases where it hits the lowest curve of excited case corresponding to a non vibrating state. It means that after photon absorption the molecule finds itself in excited non equilibrium state and it starts to vibrate like a spring. The molecule vibrates down, up and down again etc along the upper potential curve like a pendulum. These vibrations generally last for duration of 10^{-13} to 10^{-12} seconds, whereas the normal time period for excited state is about 10^{-9} seconds. Thus there is enough time for few thousands of vibrations to take place during the excited state. During this time all the extra vibrational energy is lost due to the energy exchange in the medium. Fluorescence occurs from the bottom of the upper potential curve and it follows the vertical path down until it reaches the lower potential curve. As shown in Figure 2.1, when returning back, the molecule does not reach the bottom most part of the curve, the available excitation energy gets converted to vibrational energy. This cycle of absorption-emission continues for two periods of energy dissipation. Due to this reason the fluorescence arrow is always smaller than the absorption arrow. It can also be interpreted that the fluorescence band always have larger wavelength than the absorption band as shown in Figure 2.2 [9]. The extension of the shift depends on the difference between two potential curves [8].

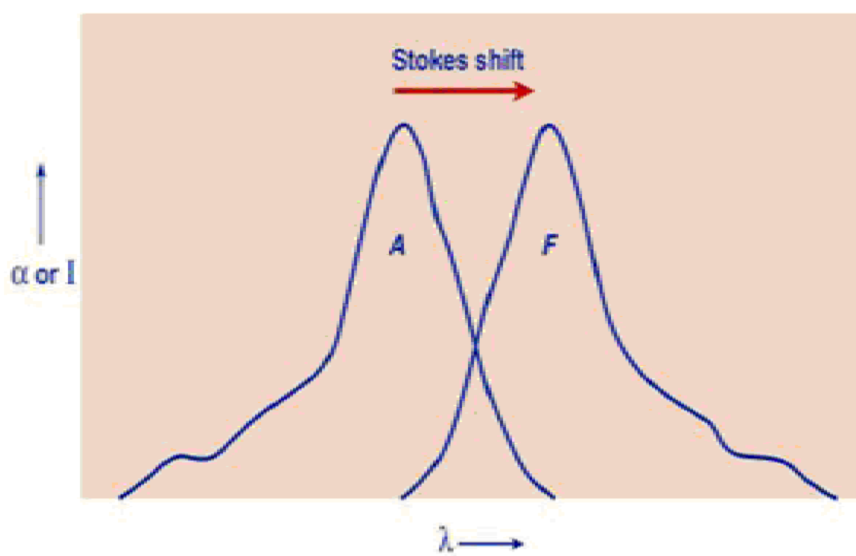

Fig 2.2 Fluorescence band displacement compared with respect to absorption band.

Another law that holds true for NIR spectroscopy is Lambert Beer's law. It states that the absorption is proportional to the concentration of the sample. A spectroscope based on NIR technology actually measures the number of photons that are absorbed at a particular wavelength. If the measurement is performed on a sample with different types of molecules, the instrument measures the number of photons absorbed by a particular molecule at a certain wavelength. As shown in Figure 2.3 [9], if I_0 and I_1 are the incident and the transmitted light through the sample respectively. The transmission T is given as

$$T = \frac{I_1}{I_0} = 10^{-\alpha l} = 10^{\varepsilon lc} \tag{2.3}$$

where l is the path length, α is absorption coefficient, c is the concentration of absorption material and ε is the molar absorptivity.

The transmission in terms of absorption can be written as

$$A = -\log_{10}\left(\frac{I_1}{I_0}\right) \tag{2.4}$$

where $A = \alpha l$

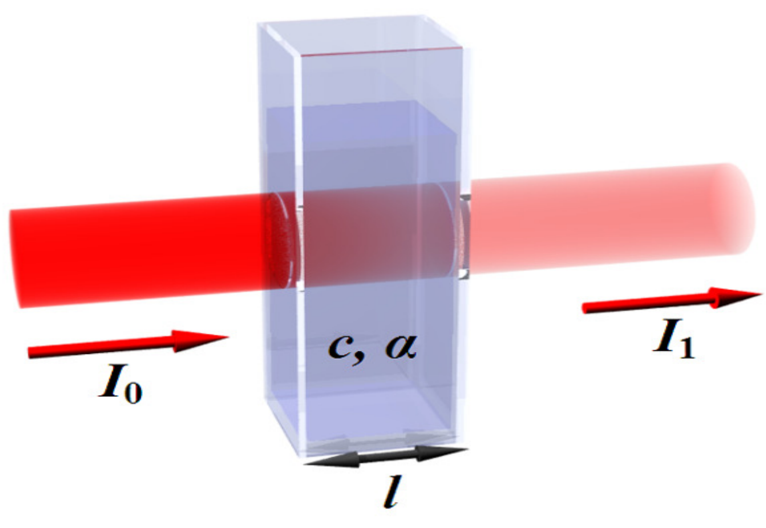

Fig 2.3 Lambert Beer's law.

2.3 Instrumentation

In the past when there were no computing facilities available, the NIR instruments were comparatively less complicated without any spectrum plot arrangement. But with the advent of computers, things have changed completely. During 1990's, the main instrument used to be a dispersive mono chromator which is still being used with some variations [7]. Instrumentation based on NIR spectroscopy includes a source, a detector and a dispersive element generally a diffraction grating. Broad wavelength sources like quartz halogen light bulbs are most commonly used for NIR analysis. Light Emitting Diodes (LEDs) are also used as they offer greater lifetime, better spectral stability and low power consumption. Detector selection depends on the wavelength range of interest. For shorter end measurement of NIR spectra, silicon based Charged Coupled Devices (CCD) are used. For remaining part of the spectrum range, these devices are not as suitable as InGaAs and PbS devices [10].

2.4 Advantages, Disadvantages and Applications

The main advantage of NIR spectroscopy is that it does not need any or little sample preparation and the data we get is in real time. NIR is a fast, reliable, as it neither requires nor produces any chemicals and is an environmental friendly technique [11]. NIR instrumentation is simple to operate and can be transported easily to any work environment. It is truly portable and can be used to do in-situ analysis. NIR is not a stand alone technique. Separate calibrations are needed and a portion of the unknown sample has always to be analyzed by the reference so that calibrations are reliable. Currently the largest growth area of NIR spectroscopy is for process analytical technology related applications where the technology is interfaced with the manufacturing process to analyze physical or chemical product qualities in real time. Major application of NIR spectroscopy has been in the analyses of gasoline, polymers and pharmaceuticals using both dispersive and Fourier transform based instruments. Pharmaceutical applications include identification of raw materials, testing the quality of products, imaging and packing of tablets. NIR spectroscopy has also proved useful in the field of agriculture where analyses of food products for adulteration are important to keep a check for quality control. NIR spectroscopy is also used in astronomy to study the atmosphere, age and masses of distant planets.

Use of NIR spectroscopy is expanding in many applications. Low absorptivity of near infrared radiations encourages this technology to be used for analysis of raw materials [11]. But still many more applications are to be discovered. NIR analysis take very little time so they have proved to be very useful in process monitoring and instruments can be installed near the process line to keep a monitor on the quality using feedback loops. NIR spectroscopy is a technology for future which still has to go far beyond.

Chapter 3

NIR-MOEMS Spectrometer

3.1 Introduction

These days instruments used to perform NIR spectroscopy are compact and are seen as an alternative to bulky and expensive FTIR (Fourier Transform Infrared) spectrometers which were based on diode array technology. These diode array based FTIR spectrometers are used at a wavelength range of 1.2 to 2.5 μm. To perform NIR spectroscopy we need a device which is not only cost effective and compact, but it should be able to perform tests without any sample preparation, hence further reducing the time required. This can be achieved using a NIR-MOEMS based spectrometer. This device uses micro-mirror to deflect the light on a diffraction grating which is then split in spectral components. NIR-MOEMS spectrometer uses two detectors at the output which is a more cost effective method compared to spectrometers based on InGaAs-Arrays. The miniaturization of the analytical devices is helping in opening opportunities for using NIR spectroscopy for process monitoring. These devices are fast, versatile, highly sensitive and accurate [1, 4].

3.2 Operating Principle

The operating principle of NIR-MOEMS based spectrometer is explained in Figure 3.1 [12]. Radiation from the sample, which is to be analyzed, is made to pass through

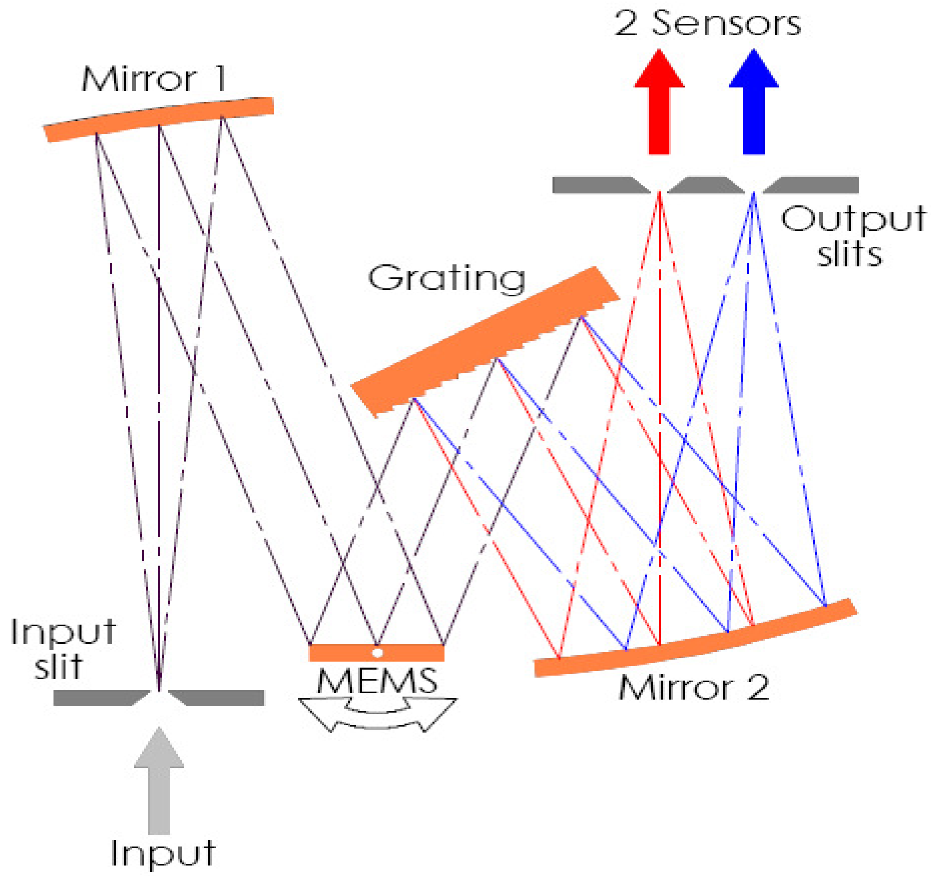

Fig 3.1 Operating principle of NIR-MOEMS spectrometer explained.

an input slit and is focused on the collimating mirror which collimates the input light on the MEMS (Micro Electro Mechanical systems). Now the radiation is reflected to a diffraction grating which divides the light into its spectral components and these components are detected by two detectors and converted to corresponding electrical signals [13].

Figure 3.2 shows how the various elements are arranged in the actual spectrometer [1]. The basic components of the NIR-MOEMS spectrometer are the fiber optics, collimation mirror, micro mirror, diffraction grating and the two detectors. The NIR-MOEMS based spectrometer is realized using simple optical Littrow alignment.

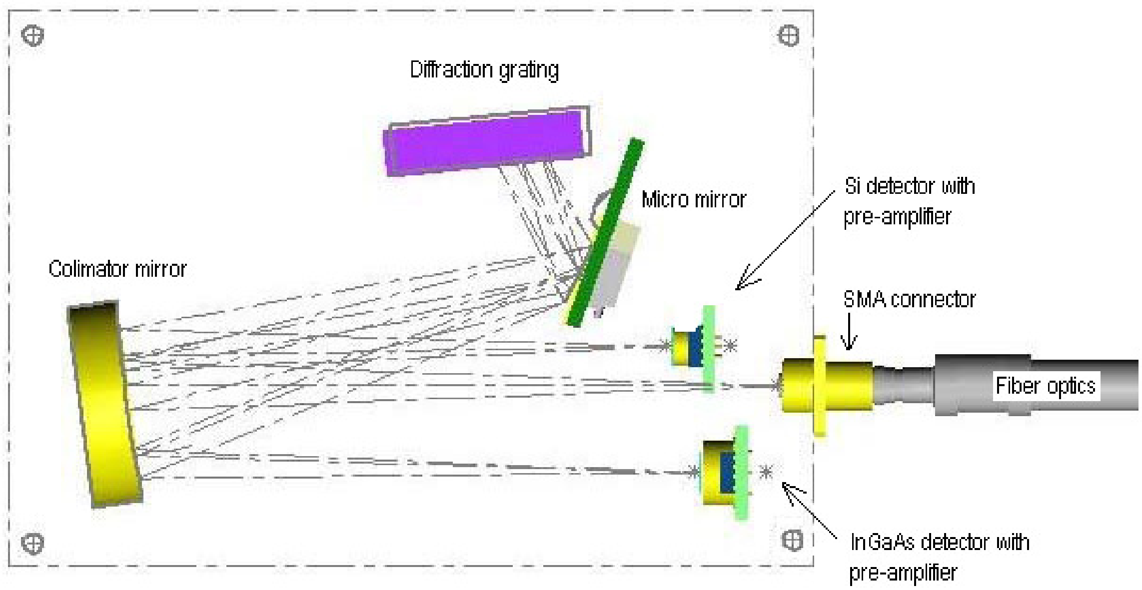

Fig 3.2 Schematic view of various components arranged in a NIR-MOEMS spectrometer.

Infrared radiation emitted by sources like halogen lamp, quartz tungsten is absorbed by the sample and then enters the fiber optics. Sample can be treated in the transmission mode using a liquid in a cuvette or in the diffusion mode where the sample surface is illuminated by incident radiation and then the diffused light is received by the fiber optics. After the light enters the fiber optics it is launched into the spectrometer through an entrance slit and is focused on the collimator. This spherical mirror collimates the light and focuses on the micro mirror which then focuses on the diffraction grating. This diffraction grating divides the light into its spectral components. The wavelength dependent reflection angle of the micro mirror transmits the desired component of the diffracted light to the exit slits using collimator. Behind these slits there are two single element detectors which detect the monochromatic radiation and with the help of trans-impedance amplifiers coverts this monochromatic radiation to electrical signals used for further data processing [4].

3.3 Specifications

The micro-mirror which is the central component of the NIR-MOEMS spectrometer should have a large optical mirror area, high reflectivity, large tilt angle and low plate wrap. The tilt of the micro-mirror can control the angle of incidence of the light that falls on the diffraction grating and separated in spectral components. Hence the tilt of the torsion mirror also affects the wavelength selection. This tilt can be controlled by the driving voltage. The torsion mirror has a resonance frequency of about 250 Hz. One complete spectrum is scanned in half period of the mirror oscillation requiring about 2 ms for one basic measurement. A high throughput and a broad working range can be achieved with a mirror surface area of $8 \times 10\ mm^2$ and a tilt angle of $\pm 6^\circ$. A plate wrap of thickness than a fraction of optical wavelength is good for getting most accurate results [14].

In order to get distinct spectral ranges, different diffraction gratings and detectors are used. For receiving a wavelength range of 660 to 1700 nm as shown in Figure 3.3, a grating of 830 lines/mm and silicon (Si) and Indium Gallium Arsenide (InGaAs) single element detectors are required. Si detectors have a detection range from 602 to 1528 nm, while InGaAs detectors have a working range of 902 to 1732 nm. But the long pass filters placed between collimator and exit slits eliminate higher diffraction orders caused by shorter wavelengths by reflecting back the short wavelength radiation. Radiations below 610 nm and 850 nm cannot reach the Si and InGaAs detectors, respectively. If we want to

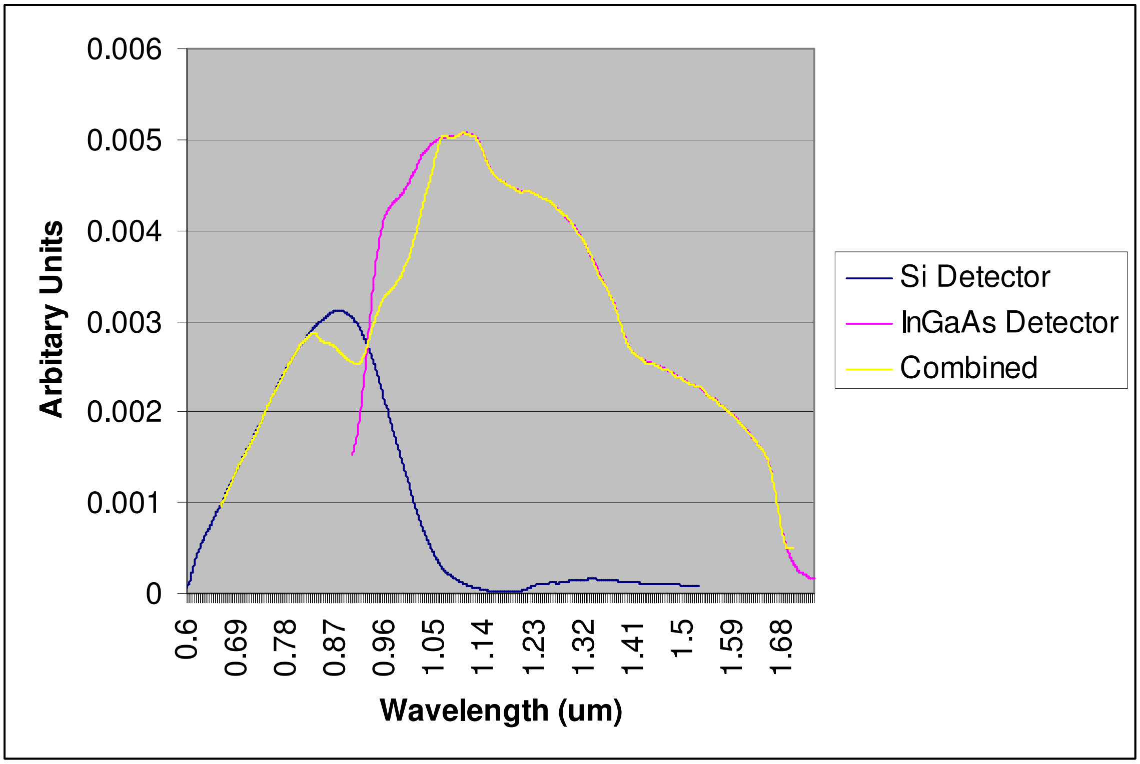

Fig 3.3 Si, InGaAs detectors and their combined readings shown with respect to wavelength.

extend the spectral range of the spectrometer to a higher level, we should use a grating with 600 lines/mm and InGaAs detector with higher cut off wavelength. For this configuration the spectral range of Si detector should be 760 to 2260 nm and for InGaAs detector this range should be 1100 to 2500 nm. Thus the combined spectral range will be of 760 to 2500 nm [1, 4]. According to the peak detectable wavelength of InGaAs detector ranging from 2100 to 2500 nm, the NIR-MOEMS spectrometer can be distinguished in various types as shown in Table 3.1. [15].

Table 3.1 Types of NIR-MOEMS spectrometers based on wavelength range.

Type	Wavelength range	Types of detectors
Type I	660 nm to 1730 nm	One Si and one InGaAs detector
Type II	910 nm to 2390 nm	Two InGaAs detectors
Type III	910 nm to 2500 nm	Two InGaAs detectors

.

. Optical resolution of the spectrometer is defined by the input and output slits and reciprocal dispersion of optical grating. Optical resolution is expressed as

$$\Delta\lambda_{1/2} = \frac{2.b.g}{F}.\cos\alpha \qquad (3.1)$$

where $\Delta\lambda$ is the optical resolution, b is the width of the exit slits (both slits are of equal distance), g is the reciprocal optical grating dispersion, F is the focal length of the collimator and α is the angle of incidence [1].

The spectrometer uses 24 V and 140 mA. It can be operated on 100 to 240 VAC, 50-60 Hz supply. The NIR-MOEMS based spectrometers are meant to be operated on room temperature. As all the sensors are not TE cooled, a substantial increase in dark current, electronic offset drift and detector noise is seen at higher operating temperatures. The effect of temperature is seen in the readings taken for the same sample at different operating temperature as the device heats up Operation temperature should not exceed 70 degree Celsius main board temperature [15]. The graph in Figure 3.4 shows the effect of temperature on a sample reading. These readings have been taken at an interval of 5 minutes for the same sample. This change in the readings is mainly due to heating up of the electronics in the spectrometer and also due to heating up of sample which results in increased absorption of light.

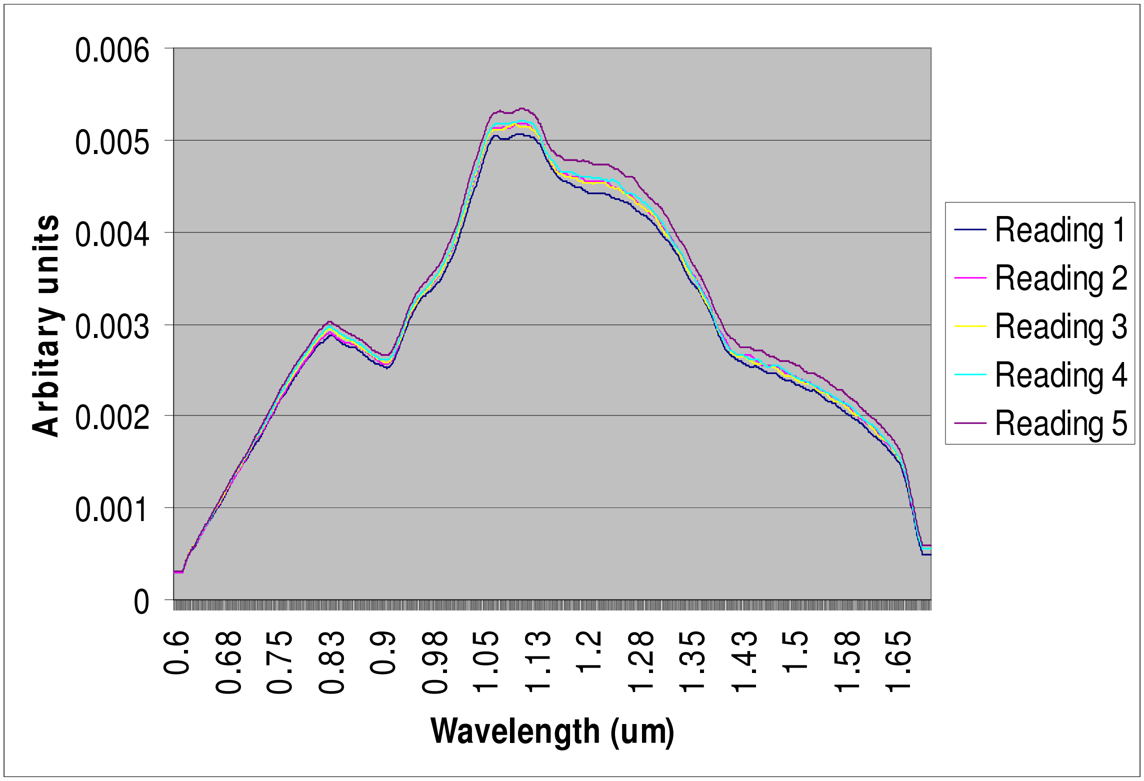

Fig 3.4 Readings of same sample at 5 minute intervals showing the effect of temperature on absorption of light.

Table 3.2 compares some of the basic specifications of NIR-MOEMS spectrometer with conventional NIR spectrometer systems [12]. It is clear that NIR-MOEMS based spectrometer has better resolution and is very fast compared to other conventional spectrometers. These spectrometers are smaller in size and lighter than other conventional counterparts. SNR (Signal to Noise Ratio) of 38 dB can be achieved by using the discussed resolution. By using rms of the original values we can greatly reduce the noise. SNR of 43 dB can be achieved by using a 5 second measurement time.

Table 3.2 NIR-MOEMS spectrometer specifications compared with conventional NIR spectrometers.

Specifications	NIR-MOEMS Spectrometer	Conventional NIR spectrometer
Wavelength range	660-1730 nm / 910-2100 nm	900-1700 nm / 1200-2100 nm
Resolution	9-12 nm	10-20 nm
Primary measuring time	4 ms	>1 sec
SNR (@single measurement)	38 dB / 30 dB	33 dB / 27 dB
Wavelength accuracy (50 single measurements)	< 0.5 nm	< 0.45 nm
Wavelength repeatability	+ 0.05 nm to –0.05 nm	+0.1 nm to -0.1 nm
Stray light	0.1%	0.3%
Dimensions (mm)	107*66*80	200*100*100
Weight	750 gm	>750 gm

3.4 Applications

NIR technique is the best technique to analyze raw materials at the starting of manufacturing of a product till the end for analyzing the finished product. Since NIR-MOEMS spectrometer is small in size, it can be used in any stage of product manufacturing for process monitoring. Not only is the size important but the capability to deliver consistent results is also important. These spectrometers are used in all the fields. Some of the potential fields of application are textile identification, pharmaceuticals, biotechnology, chemicals, food and beverages, petroleum / natural gas, polymers etc. New fields of applications are being discovered as this technique is set to change the face of spectroscopy forever. Some of these applications have been discussed briefly.

3.4.1 Textile Identification

NIR-MOEMS spectroscopy is done to ensure quality for finished goods, for textile recycling and distinguishing between real and counterfeit garments. It has been seen that there is an effect of absorption of light depending on the color of the textile. For textile applications, spectral range from 1150 nm to 1650 nm is best suited [12].

3.4.2 Medical Science

In medical science, one of the examples of using NIR-MOEMS spectrometer is the determination of fecal fat which is important in the conformation of mal-absorption and mal-digestion. Traditional methods of determining fecal fat needed sample preparation and determination which makes them both time consuming and expensive [16].

3.4.3 Food and Beverages

Applications in food and beverage industry have been using spectroscopy for a long time to keep a watch on quality control, blending and process control. NIR-MOEMS spectrometer provides accurate results in less time without or little need of sample preparation. In food industry a check has to be maintained at every level to provide healthy products to consumers. An example is to differentiate between types of condensed milks (e.g. 4%, 7.5%, 10%, 12% fat) [17].

3.4.4 Petroleum Industry

It is important to know octane number, vapor pressure, API gravity and aromatic content for petroleum related products. NIR-MOEMS spectrometers help in determining these parameters [18].

3.4.5 Waste Recycling

Recycling has become important in our day to day life. A lot of waste is recycled everyday which helps in saving the environment. NIR-MOEMS spectroscopy helps in sorting out waste depending on identification of various polymers in the waste materials. This kind of application needs a device which can provide in-situ and accurate results in a short time frame [19].

3.4.6 Forensics

NIR-MOEMS spectrometer in forensics can be used in determination of drugs e.g. heroine, amphetamine etc. Since these spectrometers are handy, accurate and can be used without any need for preparing a sample, they deliver faster results. This in turn helps to fight crime more effectively as they can be used in places like airports, even in police cars etc.

Chapter 4

Multivariate Exploratory Techniques for

NIR-MOEMS Spectroscopy

4.1 Introduction

NIR-MOEMS spectrometer is designed to perform in-situ analysis of various samples. For this purpose we need a software tool to analyze the data available from the spectrometer to perform quality analysis. The aim of qualitative analysis is not to find how much any given component is present in an unknown compound or material but rather classify it on the basis of its NIR spectral characteristics [20]. There are various methods by which this data can be analyzed and it has become a separate branch for study. This has been generally titled as Statistics in Spectroscopy (SiS) or sometimes as Chemometrics or multivariate exploratory techniques where there is more than one variable [21]. Some important methods for analyzing data obtained from NIR-MOEMS spectrometer have been discussed in detail with respect to the software tool developed. This chapter has been divided into data pretreatment, correlation analysis, cluster analysis, principal component analysis (PCA) sections.

4.2 Basic Principle

The main aim of the software tool is to recognize unknown samples. The basic principle followed here is that alike samples have similar spectra. The comparison is done in three basic steps: measurement of an unknown sample, comparison with library and identification of new sample by finding the closest match of the new sample with the data of the already known samples stored in the library. But before comparison of data, the reflection data obtained from the spectrometer has to be averaged, converted to absorption data and has to be pretreated so that we can compare the fine differences between the unknown sample data with known sample data from library. This helps in increasing the probability of getting near matches. All these points have been discussed in detail in this chapter.

4.3 Averaging of Reflection Data

A common problem faced by quality control everyday is that the reference samples can vary to a certain degree. In order to minimize this problem it is important that before we start the measurements, we should have a large number of data for the same sample so that they can be averaged. This process is done in the following steps.

1. Take the first reflection reading for sample A and store it.
2. Take the second reflection reading for sample A and store it.
3. Similarly take n reflection readings for sample A and store it.
4. Take the mean of all the readings at first wavelength for sample A and store it.
5. Repeat step 4 for all the m wavelengths for sample A and store them.

Thus the final reading is the average of n number of readings for each sample at m wavelengths and hence we obtain the final set of reflection readings with m wavelengths as shown in Table 4.1. This step is repeated for each sample. The next step is to convert these averaged reflection values into absorption values.

Table 4.1 Averaged reflection data reading for sample A.

Wavelength	Averaged Reflection reading for sample A
λ_1	$A_1^{RD} = \dfrac{A1_1^{RD} + A2_1^{RD} + \ldots\ldots\ldots\ldots + An_1^{RD}}{n}$
λ_2	$A_2^{RD} = \dfrac{A1_2^{RD} + A2_2^{RD} + \ldots\ldots\ldots\ldots + An_2^{RD}}{n}$
. . .	. . .
λ_m	$A_m^{RD} = \dfrac{A1_m^{RD} + A2_m^{RD} + \ldots\ldots\ldots\ldots + An_m^{RD}}{n}$

where A_m^{RD} is the averaged reflection data reading of sample A at wavelength m.

4.4 Conversion of Reflection Data to Absorption Data

Now that we have already got the averaged reflection data, it has to be converted to absorption data before using it for future calculations. Reflection data of Table 4.1 is converted to absorption data at each wavelength as shown in Table 4.2.

Table 4.2 Absorption data readings for sample A at m wavelengths.

Wavelength	Absorption data readings for sample A
λ_1	$A_1^{AD} = -\log\left(\dfrac{A_1^{RD}}{100\%\ \mathrm{Re}\ f_1^{RD}}\right)$
λ_2	$A_2^{AD} = -\log\left(\dfrac{A_2^{RD}}{100\%\ \mathrm{Re}\ f_2^{RD}}\right)$
. . .	. . .
λ_m	$A_m^{AD} = -\log\left(\dfrac{A_m^{RD}}{100\%\ \mathrm{Re}\ f_m^{RD}}\right)$

where A_m^{AD} is the absorption data reading of sample A and $100\%\,\mathrm{Re}\,f_m^{RD}$ is the reflection reading of reference sample at wavelength m which is used to convert reflection data readings of sample A to absorption data readings. This step is completed for all the known samples and the absorption data obtained for each sample is stored in the library which is used for data pretreatment and comparison.

4.5 Data Pretreatment

Data pretreatment is done to remove large baseline variations and to reduce noise in the observed spectra. Data acquired from the spectrometer is limited in signal bandwidth, so filtering of data is required to remove the noise. Different pretreatment methods emphasize different aspects of the data and each pretreatment method has its own merits and drawbacks. Pretreatment of data also helps in better lightening the fine differences between samples which also effectively increases the sensitivity of the software tool. The data obtained for each sample from all the pretreatment methods is saved in the library which is designed to store all the information of each sample according to sample name. In the end unknown sample X is compared with the known samples already stored in the sample library depending upon the data pretreatment method already selected for unknown sample X.

4.5.1 Absorption

In this method no pretreatment is done on the raw absorption form of data. Here the reflection data for each sample is converted to absorption data as shown in Table 4.2 and stored in the library. This step is done for all the known samples and the information is stored in the library which is used when comparison is made.

4.5.2 Centering

Centering (C) is done on the absorption data. In centering the mean of each experiment distribution is made equal to zero [21]. The absorption data for sample A as shown in Table 4.2 is added and a mean of this is calculated by dividing it with the number of wavelengths i.e. m. This mean is calculated as

$$\overline{\mathrm{A}} = \frac{A_1^{AD} + A_2^{AD} + \dots\dots\dots\dots\dots + A_m^{AD}}{m} \tag{4.1}$$

where $\overline{\mathrm{A}}$ is the mean of sample A absorption reading. Now to center the absorption data of sample A, this mean is subtracted from reflection readings of sample A at each wavelength as given in Table 4.3.

Table 4.3 Centered data readings for sample A at m wavelengths.

Wavelength	Centered readings for sample A
λ_1	$A_1^C = (A_1^{AD} - \overline{\mathrm{A}})$
λ_2	$A_2^C = (A_2^{AD} - \overline{\mathrm{A}})$
. . .	. . .
λ_m	$A_m^C = (A_m^{AD} - \overline{\mathrm{A}})$

where A_m^C is the centered absorption value of sample A at wavelength m. This centered data is stored in the library for all the known samples and is used for comparing the unknown sample.

4.5.3 Centering and Scaling

Centering and Scaling (CS) is another method of data pretreatment. When centering, a model offset is corrected but it does not work for all types of offsets. Therefore scaling is a way to fit models with weighted least squares functions [22]. In this process we center the data as shown in Table 4.3. Then we find the standard deviation of this centered data as

$$SD_A = \sqrt{\frac{(A_1^C)^2 + (A_2^C)^2 + \dots\dots\dots\dots\dots + (A_m^C)^2}{m}} \tag{4.2}$$

where m is the number of wavelengths, SD_A is the standard deviation and A_m^C is the centered absorption value of sample A. Now the centered data of Table 4.3 is divided by this standard deviation at all the m wavelengths to get the centered and scaled data as shown in Table 4.4.

Table 4.4 Centered and Scaled data readings for sample A at m wavelengths.

Wavelength	Centered and Scaled readings for sample A
λ_1	$A_1^{CS} = \frac{A_1^C}{SD_A}$
λ_2	$A_2^{CS} = \frac{A_2^C}{SD_A}$
. . .	. . .
λ_m	$A_m^{CS} = \frac{A_m^C}{SD_A}$

where A_m^{CS} is the centered and scaled data for sample A at wavelength m. This step is repeated for all the known samples. The centered and scaled values for all the samples is calculated and stored in the library which is used if the unknown sample is compared by this method.

4.5.4 Min-Max Scaling

The aim of using Min-Max Scaling (MM) is to adjust the differences in spectra by converting the data into differences relative to the maximum and minimum value of the spectra [23]. This pretreatment is done on the absorption data available for the sample as shown in Table 4.2. Min-Max Scaling is calculated at all the wavelengths as shown in Table 4.5

Table 4.5 Min-Max Scaled data readings for sample A at m wavelengths.

Wavelength	Centered and Scaled readings for sample A
λ_1	$A_1^{MM} = \frac{(A_1^{AD} - A_{\min}^{AD})}{(A_{\max}^{AD} - A_{\min}^{AD})}$
λ_2	$A_2^{MM} = \frac{(A_2^{AD} - A_{\min}^{AD})}{(A_{\max}^{AD} - A_{\min}^{AD})}$
. . .	. . .
λ_m	$A_m^{MM} = \frac{(A_m^{AD} - A_{\min}^{AD})}{(A_{\max}^{AD} - A_{\min}^{AD})}$

where $A_{\max}^{AD}$ and $A_{\min}^{AD}$ are the maximum and the minimum values of the absorption spectra for sample A respectively. A_m^{SRA} and A_m^{AD} are the scaled values relative to area and absorption values for sample A at wavelength m respectively. This process is completed at all the wavelengths for all the known samples and the readings are stored in the library as shown in Table 4.5. This data can be used for comparison of unknown sample X with the known samples when we select scaling relative to area as the method for comparison.

4.6 Savitzky-Golay Smoothing and Differentiation Filter

Initial objective of Savitzky-Golay (SVG) filter is to separate noise from raw data. Savitzky-Golay filter acts as a low pass filter and is commonly used to differentiate data. It also conserves central moments of data [24]. The most common means of estimating the true signal from a set of data is to smooth the data using moving point average. This forms the basics of SVG filter method. In this process a fixed number of points are considered and we add the data of these points and divide it by the number of points considered. We keep on repeating this process by adding the next point and dropping the last point in the group [25].

The moving point average can be applied with the concept of convolution function. This is easier to explain with the help of an example. In Table 4.6 the set of values on right is the raw data and the set of data on left is any range of wavelength. The center column represents the range of convoluting integers. In this case we have selected 5 point moving average from C_{-2} to C_2 with C_0 being the middle one. To perform the convolution of the raw data with the convoluting integers, C_i, every raw data point is multiplied by the convoluting integer corresponding to it in the table and the results of these multiplications are summed and divided by the number of convolution integers considered i.e. 5 in this particular case. The result is the desired data for the point corresponding to C_0. For the next step, we move the set of convoluting integers down by one point and repeat the process until all the data points have been normalized. For a simple moving average the value of all the convoluting integers is one. This is not as useful in case of going through a sharp peak. The average would degrade the end of slope. So it becomes very important to convolute with a function which treats the past and the future values on equal basis [25].

Table 4.6 SVG convolution concept explained.

Wavelength (nm)	Convoluting Integers	Raw data
660		0.146615
661		0.143581
662	C_{-2}	0.144675
663	C_{-1}	0.147406
664	C_0	0.148174
665	C_1	0.147145
666	C_2	0.145489
667		0.144134
668		0.144079

Least squares method evaluates the best value at a point chosen by this method for a group of points being examined. Least squares method is used to determine improved set of coefficients using polynomials rather than simple averaging. These coefficients are then used for performing the convolution operation. It can be applied for any odd number of points and to any degree of polynomials. The number of points in the filter function is chosen such that we can adjust the cut off frequency of the filter accordingly. But we should be careful in choosing the number of points for filter function such that the high frequency noise is removed without having a significant effect on the signal. Line shape can be maintained by conserving higher central moments. To keep a balance between high frequency responses of filter and conserve central moments a compromise is needed [24].

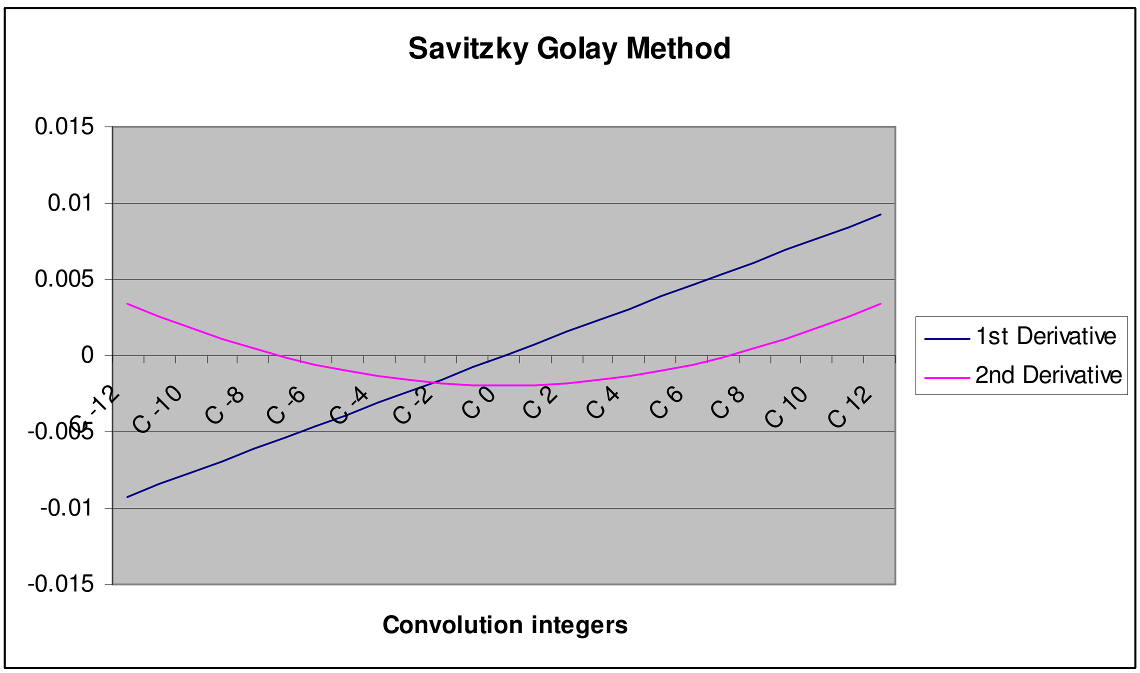

Fig 4.1 25 point 1st and 2nd derivative SVG convoluting integers.

Figure 4.1 shows the convoluting integers for the quadratic 25 point 1st and 2nd derivative functions. So the order of the filter is quadratic and the numbers of points considered are twelve. These integers are multiplied with the data set to obtain the first and the second derivative of the data set. These convolution integers are taken from the Savitzky and Golays original paper [25].

In the software tool 25 Point 1st and 2nd Drivative Savitzky-Golay filter function has been used on all the four types of pretreated data explained in Section 4.5. The results are stored in the library. Since we have taken 25 point derivatives so the first and the last twelve points have been considered to be zero by default as the values of these points cannot be calculated. Block diagram of implementation of Savitzky-Golay method in the software is shown in Figure 4.2.

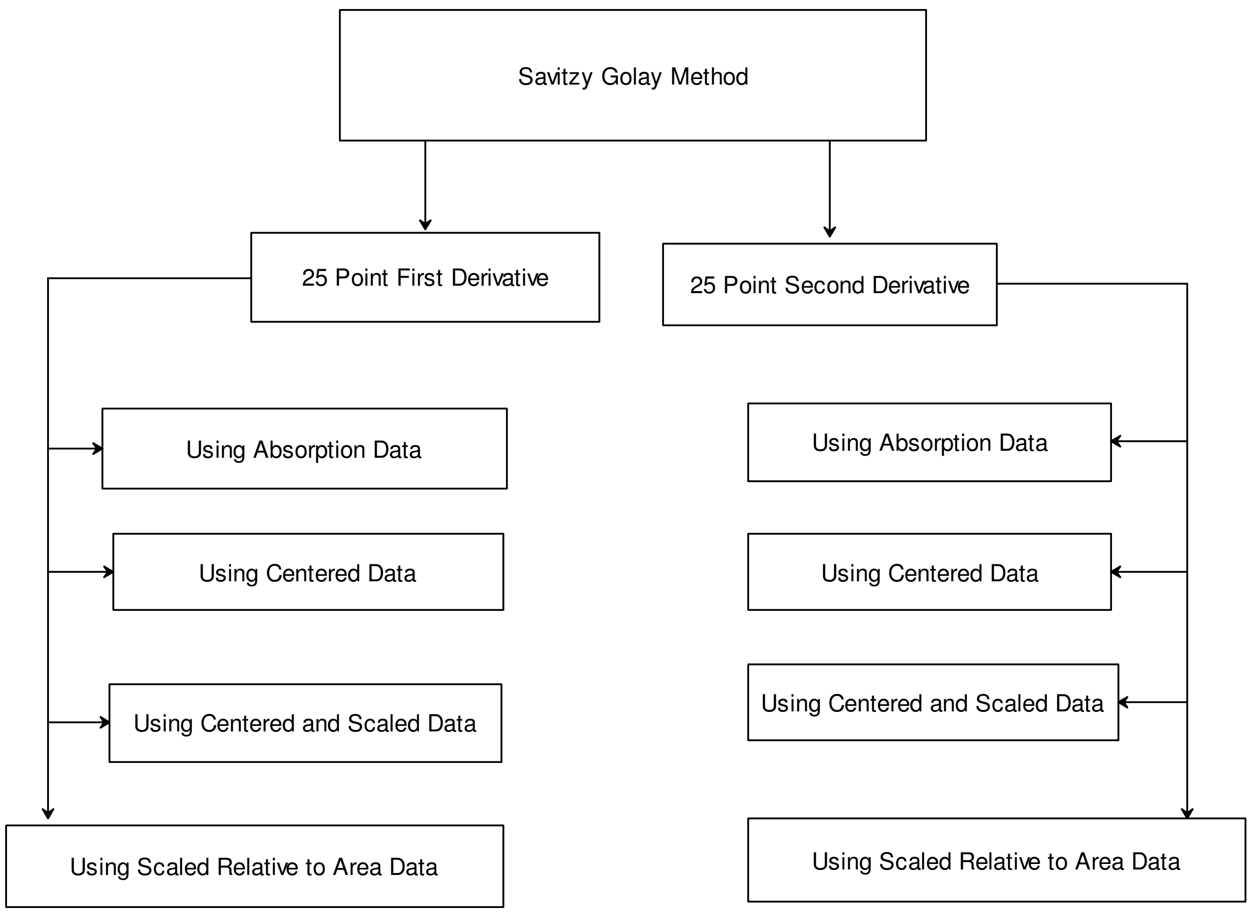

Fig 4.2 Block diagram of implementation of Savitzky-Golay method in the software tool.

4.7 Principal Component Analysis

Principal Component Analysis (PCA) is a way of identifying the patterns in data and highlighting their similarities and dissimilarities. It is used when the dimension of data is huge, making it difficult to find the patterns in the data [26]. The methodology of PCA technique has been explained in the following steps.

4.7.1 Data Acquisition

The first step is to get data. Our data is two dimensional as we have to compare the known sample with unknown one. A is the absorption data of known sample and X being the absorption readings of the unknown sample on the other side as shown in Table 4.7 [27].

Table 4.7 Data for known sample A and unknown sample X.

Data for known sample A	Data of unknown sample X
A_1^{AD}	X_1^{AD}
A_2^{AD}	X_2^{AD}
. . .	. . .
A_m^{AD}	X_m^{AD}

4.7.2 Finding Mean and Subtracting it

The next step is to find the mean by adding the readings and dividing it by the number of readings as given in Eq (4.1). This is done for both the known sample $\left(\overline{A}\right)$ and unknown sample being compared $\left(\overline{X}\right)$. These mean values are then subtracted from values of sample A and X, respectively at all wavelengths as shown in Table 4.8 [27].

Table 4.8 Centered values for known sample A and unknown sample X.

Wavelength	Centered readings for sample A	Centered readings for sample X
λ_1	$(A_1^{AD} - \overline{\mathrm{A}})$	$(X_1^{AD} - \overline{X})$
λ_2	$(A_2^{AD} - \overline{\mathrm{A}})$	$(X_2^{AD} - \overline{X})$
. . .	. . .	. . .
λ_m	$(A_m^{AD} - \overline{\mathrm{A}})$	$(X_m^{AD} - \overline{X})$

4.7.3 Finding Covariance Matrix

Since the data is two dimensional, the covariance matrix of the data will is a 2x2 matrix [27]. The first value of covariance matrix $\left(a_{11}\right)$ is calculated from known sample A as

$$a_{11} = \frac{[(A_1^{AD} - \overline{A})^2 + (A_2^{AD} - \overline{A})^2 + ... + (A_m^{AD} - \overline{A})^2]}{(m-1)} \tag{4.3}$$

Fourth value of the 2x2 matrix, (a_{22}) is calculated the same way but it utilizes the data from unknown sample X as

$$a_{22} = \frac{[(X_1^{AD} - \overline{X})^2 + (X_2^{AD} - \overline{X})^2 + ... + (X_m^{AD} - \overline{X})^2]}{(m-1)} \tag{4.4}$$

The second (a_{12}) and the third (a_{21}) values are equal and are calculated using both known sample A and unknown sample X as

$$a_{12} = a_{21} = \frac{[((A_1^{AD} - \overline{A})(X_1^{AD} - \overline{X})) + ((A_2^{AD} - \overline{A})(X_2^{AD} - \overline{X})) + ... + ((A_m^{AD} - \overline{A})(X_m^{AD} - \overline{X}))]}{(m-1)} \tag{4.5}$$

Thus the 2x2 covariance matrix (C) is obtained as

$$C = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \tag{4.6}$$

4.7.4 Calculating Eigen Values

The next step is to calculate the eigen value and it can be calculated as [27, 28]

$$Z \times I - C = Z \times \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} - \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = 0 \tag{4.7}$$

where Z is an unknown value whose roots are to be calculated using this formula. I is the identity matrix and C is the covariance matrix. Using this equation we get a quadratic equation and if we solve for determinant of this matrix we get two roots for Z. These are the eigen values for the covariance matrix. The highest eigen value of these two is called the principal eigen value. This principal eigen value is used to calculate the eigen vectors.

4.7.5 Calculating Eigen Vectors

Eigen vectors can be calculated as

$$(C - Eval \times I) \times Y = \left(\begin{vmatrix} a_{11} \, a_{12} \\ a_{21} \, a_{22} \end{vmatrix} - Eval \times \begin{vmatrix} 1\,0 \\ 0 1 \end{vmatrix} \right) \times \begin{vmatrix} Y_{11} \\ Y_{21} \end{vmatrix} = 0 \tag{4.8}$$

Here C is the covariance matrix, $Eval$ is the highest eigen value, I is the identity matrix and Y is a column matrix with same number of rows as covariance matrix C. Now solving Eq. (4.8) we get an equation containing two variables Y_{11} and Y_{21}. These are the two eigen vectors corresponding to the highest eigen value. For finding out these two eigen vectors we use the simple property of eigen vectors given as [27, 28]

$$Y_{11}^2 + Y_{21}^2 = 1 \tag{4.9}$$

By solving Eqs. (4.8) and (4.9), we are able to determine the values of Y_{11} and Y_{21} i.e. the two eigen vectors. These two eigen vectors are then compared by using cluster analysis explained in the next section. In this case unknown sample X is compared with all the known samples and for every case we get the two eigen vectors. These eigen vectors are compared by cluster analysis using Euclidean distance formula.

Principal component analysis has been implemented on raw absorption data, 25 Point 1st Derivative Savitzky-Golay absorption data and 25 Point 2nd Derivative Savitzky-Golay absorption data. A block diagram of implementation of this technique is shown in Figure 4.3.

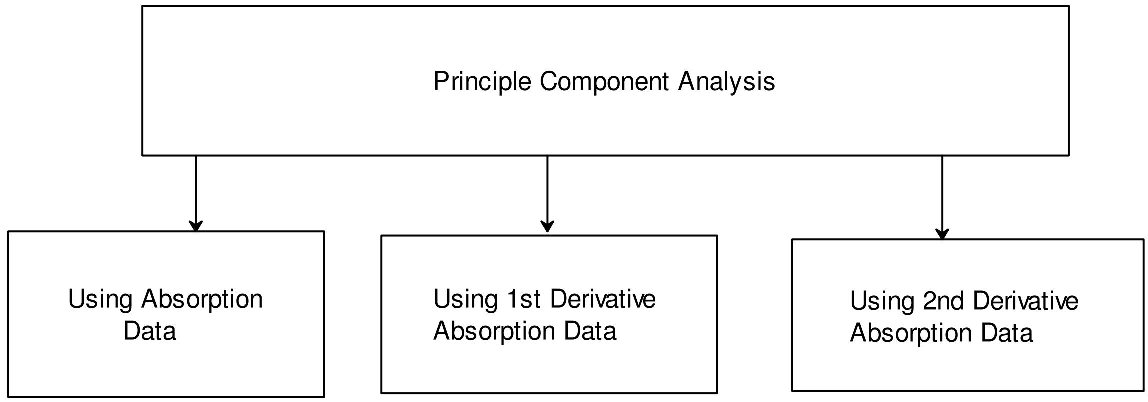

Fig 4.3 Block diagram for implementation of Principal Component Analysis on the types of data.

4.8 Cluster Analysis

Cluster analysis in general means grouping of same type of objects. Distance analysis forms an important part of cluster analysis. Euclidean distance formula is the most common formula being used to find distance between two spectra [20]. All the data that have been pretreated or not, are compared by using the Euclidean distance formula as given

$$D=\sqrt{\sum_{m=n}^{m}\left(A_m-X_m\right)^2} \tag{4.10}$$

where D is the Euclidean distance A_m and X_m are the values of known sample A and unknown sample X at wavelength m. For comparison, the distance is found at each wavelength for the unknown sample with all the known samples according to pretreatment method selected. The smallest distance in that sample is the closest match for the unknown sample. Euclidean distance representation is shown in Figure 4.4.

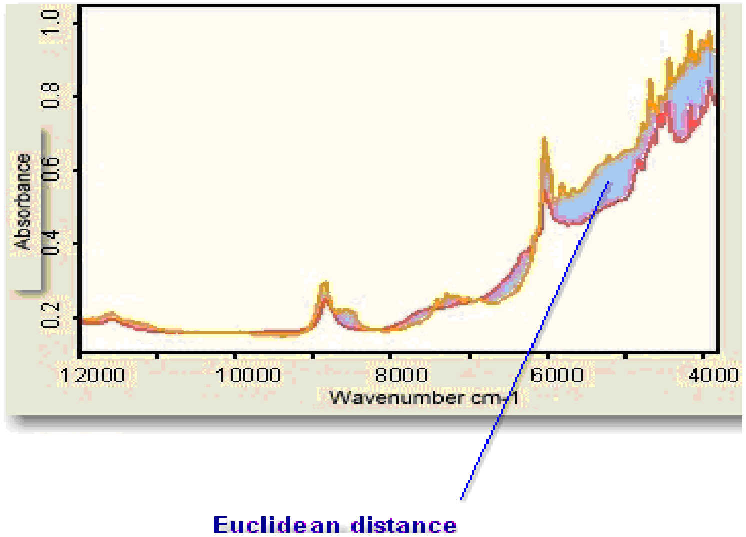

Fig 4.4 Euclidean distance representation.

Use of proper chemometrics technique can help to precisely and correctly analyze the data. This can be achieved by using the techniques discussed in the previous sections. By utilizing these techniques we can handle spectral data i.e. correct offset and achieve referencing. It also helps in minimizing the noise and is able to extract useful information to utilize it the best possible way.

Chapter 5

Measurement Setup of NIR-MOEMS Spectrometer and Working of Software Tool

5.1 Introduction

The multivariate techniques that have been discussed in the last chapter have been developed and implemented as software tool. The software developed is simple to understand and use. The aim of this software is to compare the unknown sample data with the data of the known samples stored in the library. However before we can make any comparison, it is necessary to set up a library with all the information of the known samples. This information includes all the original data and the new data obtained after pre-treating the original sample data. Thus the main focus of this chapter is the understanding of the measurement set up of the NIR-MOEMS spectrometer and working of the software tool with the help of flow charts. Some salient features of the spectrometer software tool have also been discussed.

5.2 NIR-MOEMS Measurement Setup

NIR-MOEMs spectrometer is a powerful technique for identifying different types of drugs or tablets. In order to distinguish between packaged tablets, measurements are made in diffuse reflection mode using an integrated sphere. Figure 5.1 shows the NIR-MOEMs spectrometer set up using the integrating sphere. The integrating sphere is attached on top of the NIR-MOEMs spectrometer. Inside the integrating sphere, there is a golden foil for maximum reflections to take place. There is also a halogen source located in the middle of the integrating sphere. An opening is created at the top of the integrating sphere where test samples are placed. To maximize optical performance of the described method, optimal sample configuration has to be taken. Some important points that should be considered while performing the experiment are: there should be equal distance between aperture of the integrating sphere and the samples, the illumination of the sample should be homogeneous and multiple measurements are done for each sample [15].

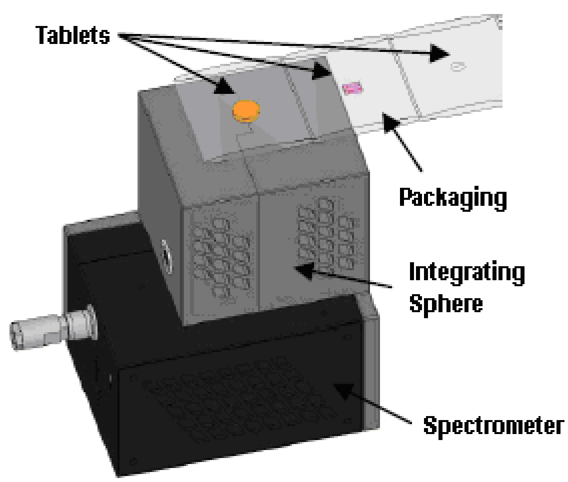

Fig 5.1 Spectrometer with integrating sphere.

The main steps for the measurement are [15]:

1. Connect the spectrometer to a power supply of 24 VAC.
2. Then connect the spectrometer to a computer using a USB cable.

3. Power is supplied to the integrating sphere containing the halogen bulb. The halogen source is given 9 volt because the signal to noise (SNR) ratio is better at this power supply. If it is set up at 12 volt, the system might get heated and below 9 volt (say 3 or 6 volt), the illumination is not sufficient to get a good SNR.
4. After connecting the spectrometer to the computer via USB cable, the software in the computer detects this port and after few routine calibrations, it is ready to perform measurements.
5. Measurements can be started by going to the menu item “Start / Single measure”. We can also perform loop measurements by selecting “Start / Loop measure” from the menu item.
6. All measurements primarily contain emission spectra which are shifted by an unknown spectral offset. This offset can be determined by means of a null measurement i.e. taking a measurement without turning the radiation source ON. It can be set by going to the menu item and selecting “Reference / Set reference 0%”. Similarly we can set the reference by taking some readings of a 100% reference sample (sample which has high reflectivity) and set it by going to the menu item and selecting “Reference / Set reference 100%”
7. Now single or multiple measurements of the samples are performed and are saved by selecting “File / Save last curve as ASCII file” from the menu item in appropriate folders. These measurements are used in the software tool to perform data analysis.

5.3 Software Tool Working with Flow Charts

The working of the software tool has been explained in this section with the help of flow charts. The software tool mainly consists of five basic utilities. These five utilities are shown in Figure 5.2.

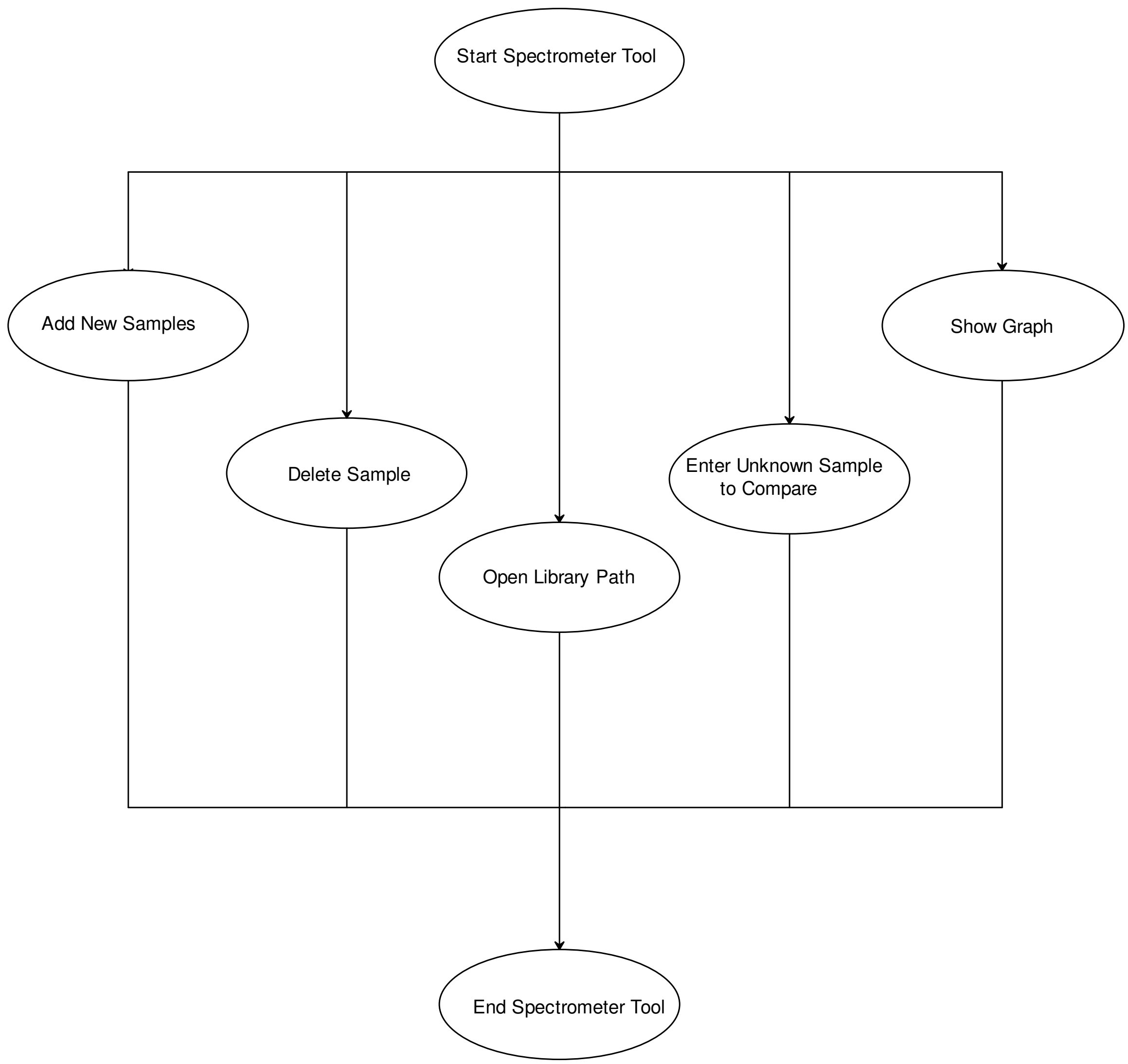

Fig 5.2 Flow chart for the basic operations performed by the software tool.

These utilities are: “Add New Samples”, “Delete Sample”, “Open Library Path”, “Enter Unknown Sample to Compare” and “Show Graph”. Each of these utilities is explained using their individual flow charts in the following sections.

5.3.1 Add New Samples Utility

To add a sample we have to follow the following steps. These steps are shown in the flow chart in Figure 5.3(a) and continue in Figure 5.3(b).

1. First of all select “Add New Samples” option from the file menu. A dialog box pops up and asks for reference sample readings and sample readings.
2. Add the reflection readings of the 100% reference sample to reference readings. These 100% reference sample readings are important to convert the reflection readings (RD) of the sample to absorption readings. Also now add the folder containing the reflection data (RD) readings of the known sample.
3. After selecting the reference and known sample readings, select the wavelength range of interest. This wavelength range depends upon the user. Once selected, all other samples which are going to be added in the library should also be saved at this wavelength range only.
4. Then select the sample name. If the sample name already exists, a pop up window informs us that a sample already exists by that name.
5. The sample is stored in the library as sample name.smd. The path of storage of this sample is given by the name of sample in “Library.smt”.
6. Then the “sample.smd” file is loaded in the spectrometer tool.
7. The User Interface (UI) is updated with the data of the sample.

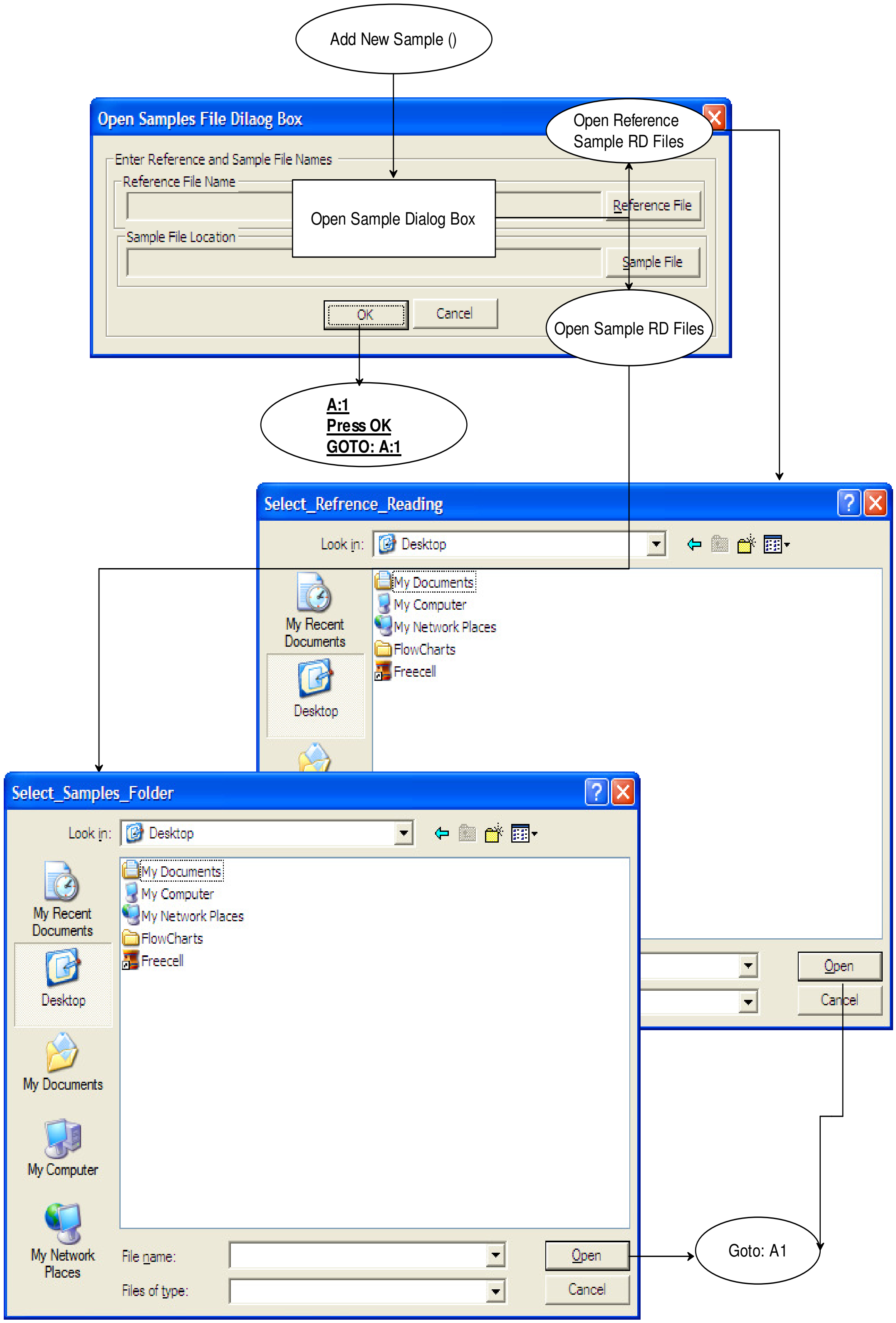

Fig 5.3(a) Flow Chart for "Add New Samples" utility.

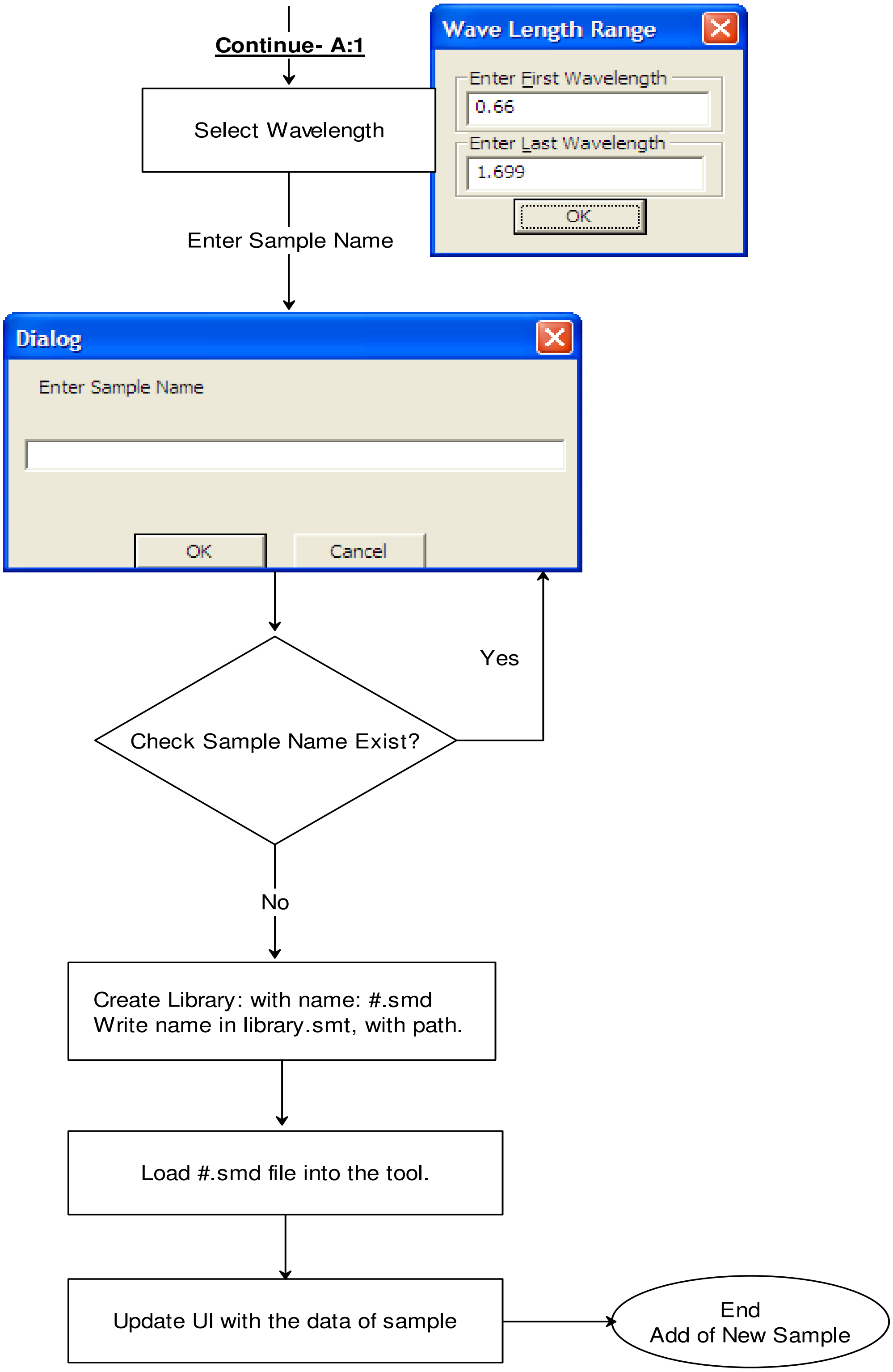

Fig 5.3(b) Flow chart for "Add New Samples" utility continued.

5.3.2 Delete Sample Utility

Some times it may be necessary to delete a sample. So “Delete Sample” utility is provided in the software tool which deletes a sample and all its corresponding data from the library. “Delete Sample” utility flow chart is given in Figure 5.4 and explained.

1. To start deleting a sample from the library we have to first click the “Delete Sample” option from the menu bar.
2. Then we have to enter the name of the sample which is to be deleted.
3. The tool checks from the name of the sample if it exists in the library.
4. If the name does not exist then it ends the process.
5. But if the sample name exists in the library then it deletes the sample asking first if you are sure that you want to delete that particular sample from library.

There is also a shortcut button to delete any sample. We have to select the sample name from the list of samples in the tool. Then we press the “Delete” button on the keyboard. Deleting sample also deletes the corresponding file number (“sample.smd”) given to the sample. This vacant number is given to the new sample which is added next and all its information is updated in “Library.smt” file.

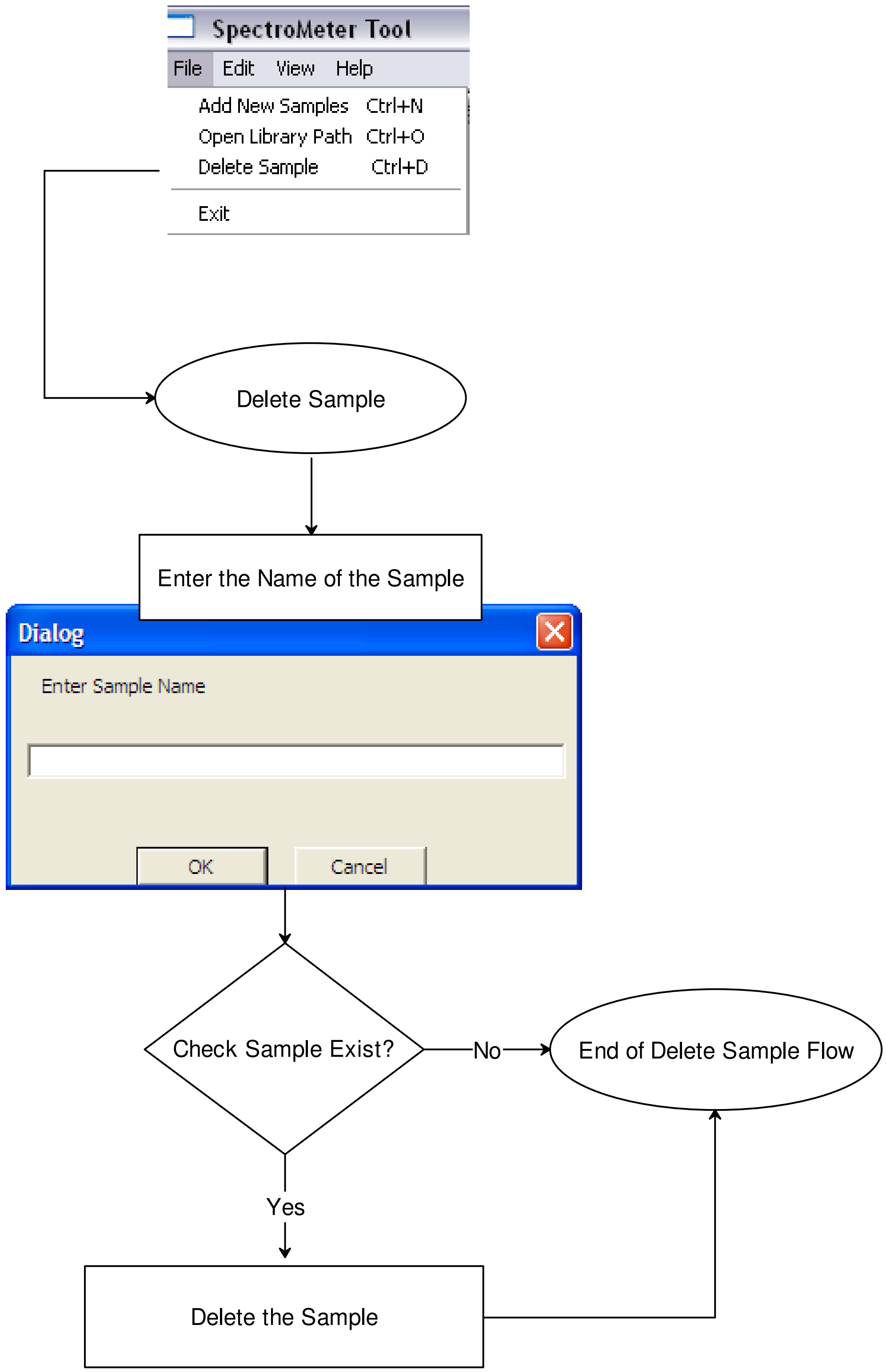

Fig 5.4 Flow chart for "Delete Sample" utility.

5.3.3 Open Library Path Utility

As we can make more than one library at different wavelength ranges, so the spectrometer tool should have an option of opening up any library of interest. The libraries can be saved in any folder on the computer and can be used as per our requirements. In order to use the libraries saved at different folders, we have to provide proper path for them as shown in Figure 5.5. The operational steps involved are:

1. From the file menu select “Open Library Path” option or press the short cut button Ctrl+O.
2. After that a dialog box opens and it asks for the library path.
3. Go to the folder which contains the library of interest and then select “Library.smt” file from the folder.
4. It then reads the contents of “Library.smt” file.
5. Then it loads the samples from that library in the user interface.

Whenever we close the spectrometer tool and start it again, it uploads the library which was selected the last time when the tool was working.

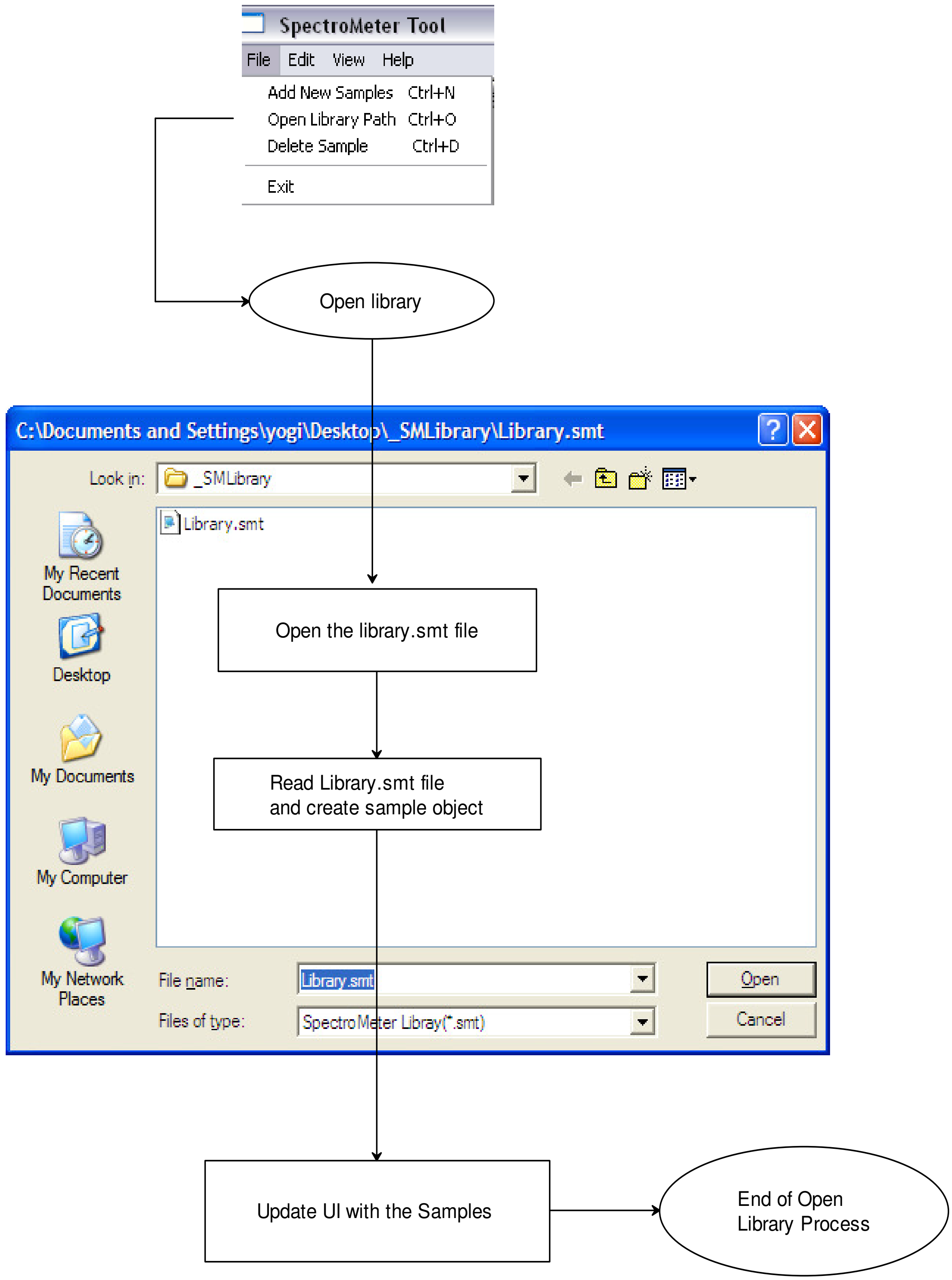

Fig 5.5 Flow chart for "Open Library Path" utility.

5.3.4 Enter Unknown Sample to Compare Utility

An unknown sample can be compared by the known samples stored in the library in different ways. The process that takes place during comparison of samples is shown in Figure 5.6(a) and Figure 5.6(b). This process is explained below.

1. First of all enter the unknown sample. This is done by first entering the reflection readings of the 100% reference sample from reference data files and then the reflection readings of unknown sample from sample data files.
2. Then select the method of comparison. There are six main methods of comparison with some of these methods having sub methods (Derivation and PCA) as shown in the flow chart of Figures 5.6(a) and 5.6(b). If we select Derivation or PCA new dialog box pops up on the screen asking about the different ways of performing Derivation or PCA.
3. After selecting the appropriate method, select the wavelength range for comparison. This selection of wavelength range should be the same as the wavelength range of the known samples in the library for best results.
4. After pressing ok the results of comparison are seen at the lower end of the user interface. It gives the first five matches of the closest samples and also tells us how close these matches with the unknown sample are.
5. In the left lower most corner, some information about the unknown sample being compared is also given example number of reflection data readings taken and the number of wavelengths each reading have.

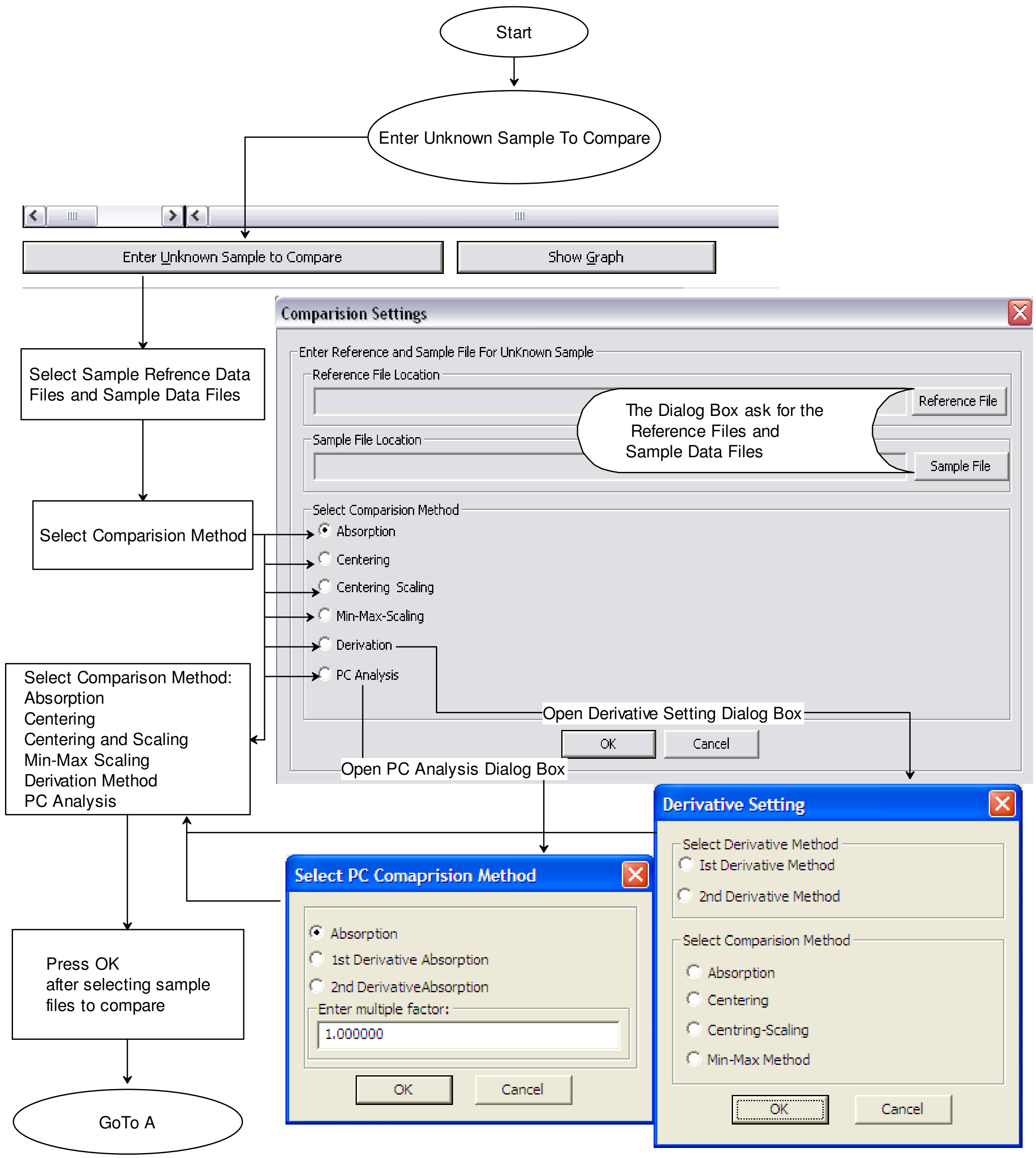

Fig 5.6(a) Flow chart for "Enter Unknown Sample to Compare" utility.

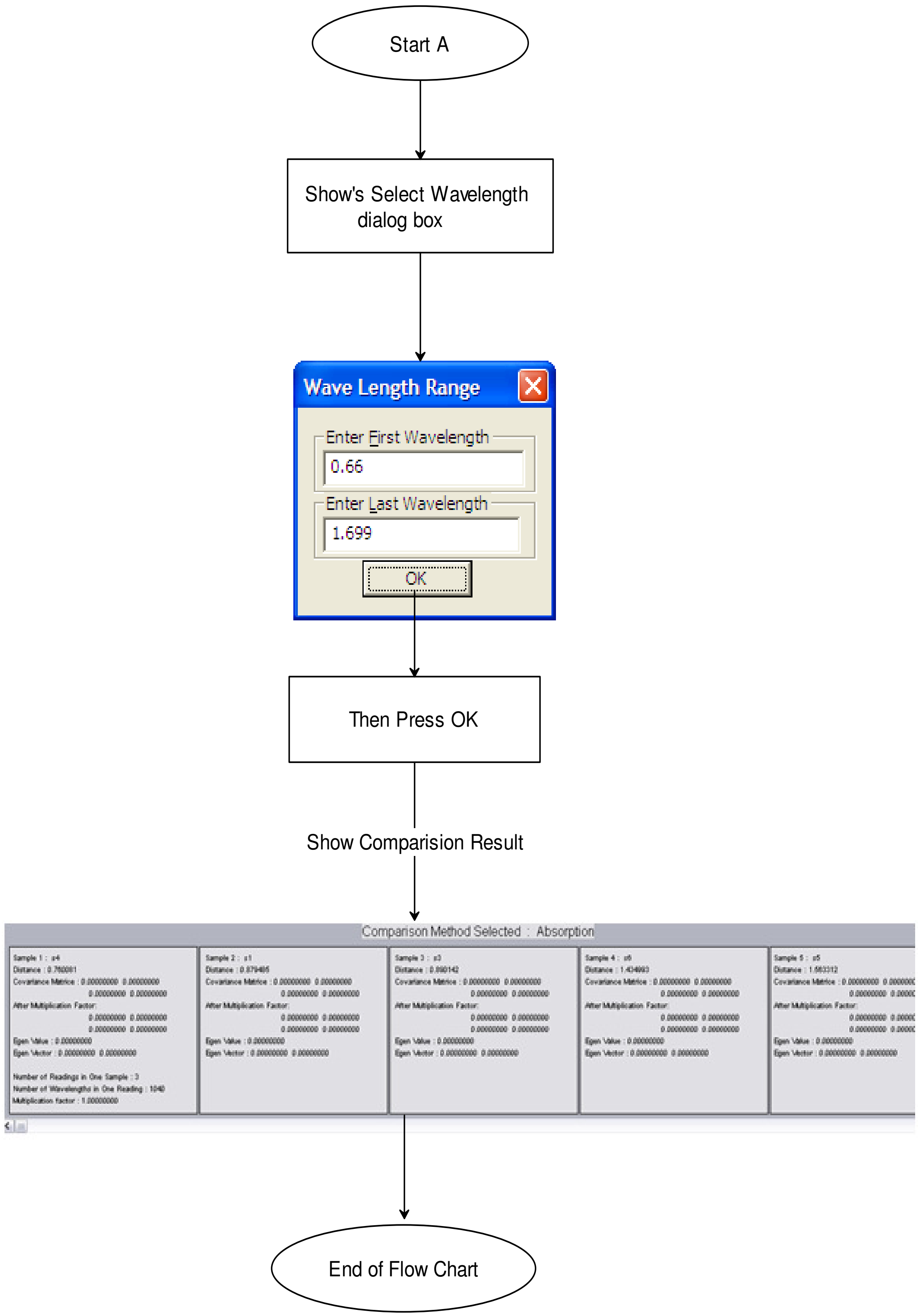

Fig 5.6(b) Flow chart for "Enter Unknown Sample to Compare" utility continued.

5.3.5 Show Graph Utility

After the comparison has been made and the results are displayed in the lower part of the tool, we have a choice of displaying these results in form of a graph. There are two types of graphs. The first one is line graph which shows the readings of the unknown sample and the readings of first five closest samples. The line graph is in accordance to the method of comparison chosen. Only in case of PCA the tool does not give any line graph because we only get two eigen vectors in that case. The second type of graph is the distance graph. The distance graph shows the distance of the closest first five samples with the unknown sample as bars. The smaller the bar, the closest the sample is to unknown sample. These graphs are colored and each color depicts a different sample shown in the index at the right hand corner of each graph. If we double click on any graph, that graph is highlighted on the entire screen. Figure 5.9 shows how the graph utility works and is explained in the following steps.

1. After the method of comparison is selected and the samples are compared, we get the result in the lower end of the tool.
2. After getting the results of comparison press the button for "Show Graph" and the graph screen pops up. This screen has line and distance graphs.

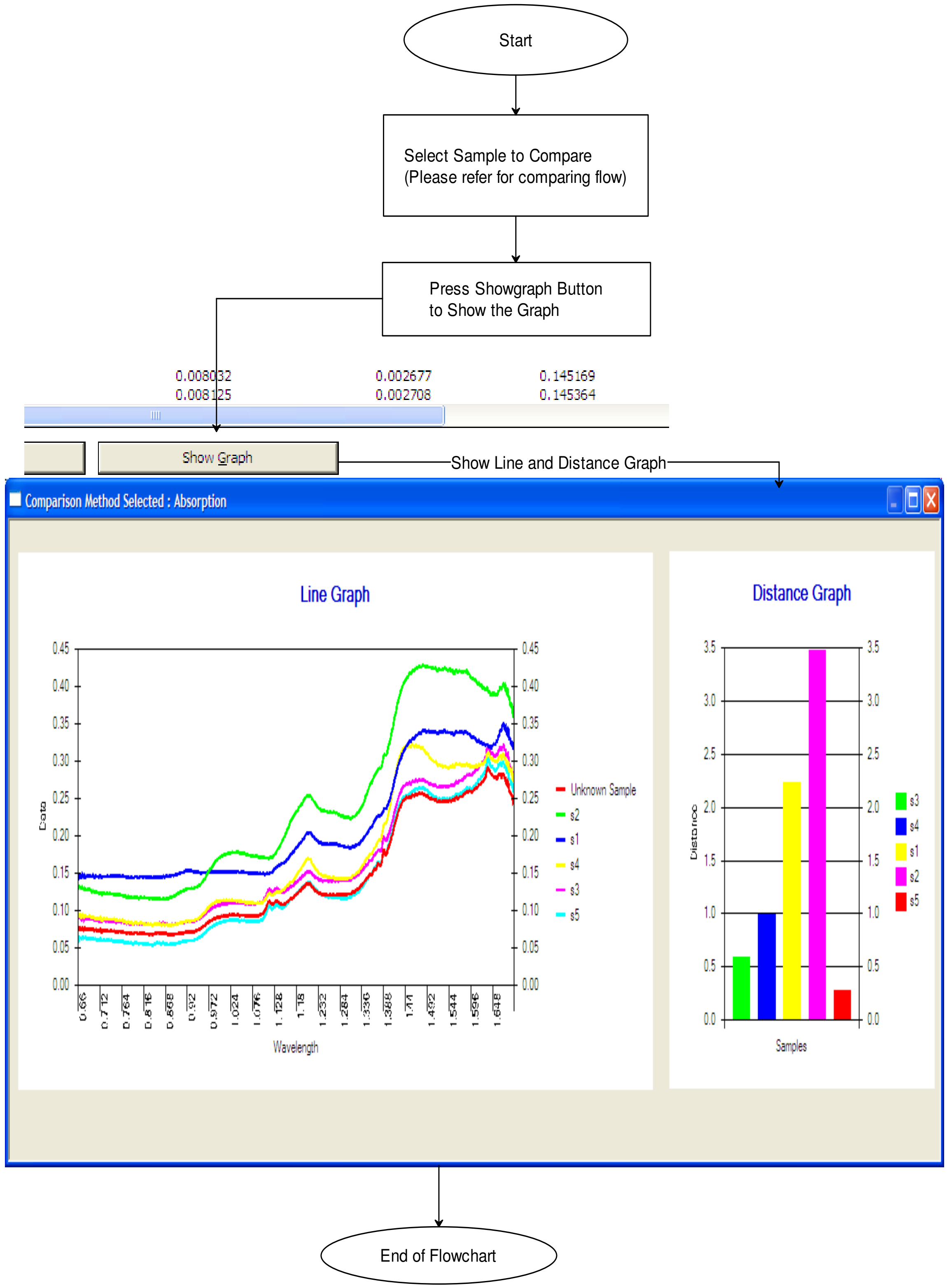

Fig 5.7 Flow chart for "Show Graph" utility.

5.4 Salient Features

The software tool to analyze the data contains some special features which help in better understanding the data patterns and finding similarities between samples. Some of these features are discussed below.

1. Same sample readings may vary up to some extent each time we take measurements. To overcome this problem the tool provides an option of averaging all the data. This is done by taking several readings for a sample and averaging them. The number of readings to be taken is user dependent and can vary from sample to sample.
2. Another important feature is that before starting any pre-treatment of data we can convert the reflection data readings obtained from the spectrometer to absorption data readings.
3. Using this software tool we can also select any wavelength range of interest. But this new wavelength range is always a subset of the initial range of wavelengths selected in the spectrometer. This helps in removing the noise at the beginning and at the end of the measurement.
4. We can make as many numbers of libraries for the samples at different wavelength ranges and save them anywhere on the computer. All these libraries are easily accessible by choosing the open library path.
5. We can add and delete samples any time according to user choice. But the care should be taken that the new sample that is being added should be at the same wavelength range as rest of the samples in the library where it is being added.

6. We can compare the readings using different methods. There are 15 different ways in total by which we can compare the samples. Among these different ways the noteworthy techniques are using Derivation and Principal Component Analysis.
7. The results obtained from the tool are shown in the lower end of the tool with all the important information like, method of comparison selected, number of reflection data readings and the number of wavelengths taken for the unknown samples. The results show us the first five samples which are closest to the unknown sample and it also shows how close these samples are to the unknown sample by giving the distances between them.
8. The results can also be displayed in the form of graphs. Software tool shows the graph of the readings of closest five samples corresponding to the comparison method selected. It also shows the distance graph of the five closest samples from the unknown sample.
9. This software tool does not make any changes to the original data files obtained from the spectrometer. It uses those files to make a library. So even if we have to make another library using different wavelength range, we can use the same original data files instead of repeating the measurements of all the samples with spectrometer.

Chapter 6

Results and Discussions

6.1 Test Samples

The design of the software is such that the drugs can be analyzed using various data pretreatment methods. A parallel utility consisting of Microsoft Excel and various other programs available online was also used to verify the results obtained from our software tool. The results were also compared with OPUS software made by Brucker Optics Inc. The results obtained from OPUS software are limited to just one wavelength range whereas our software tool can obtain results at any wavelength range provided that the new wavelength range is a subset of the original wavelength range. Before discussing the final results, it is important to know the kind of drugs that were used for testing. Almost ten known samples were chosen from a bundle of samples because of various reasons such as some were altogether different samples, some were the same tablets with different concentrations while some drugs had the same concentrations but from different manufactures. The samples whose name were already known and were stored to make a library are listed below

1. Sample 1 - AH3N Tab
2. Sample 2 - Akineton Retard Tab
3. Sample 3 - Allopurinol (100 mg)
4. Sample 4 - Allopurinol (100 mg Different manufacturer)

5. Sample 5 - Allopurinol (300 mg)
6. Sample 6 - Allopurinol (300 mg Different manufacturer)
7. Sample 7 - Allopurinol (300 mg Different manufacturer)
8. Sample 8 - Alna Ocas (0.4 mg)
9. Sample 9 - Alprazolam (0.25 mg)
10. Sample 10 - Alprazolam (0.5 mg)

The aim of the software tool is to distinguish between the drugs and to provide the closest match to the unknown sample with the known samples already stored in the library. The results are even better if the closest matches are same in concentration and from the same manufacture. The test sample or unknown sample is carefully chosen such that it is totally different from some of the samples stored in the library, similar to some stored samples and an exactly same to one of the samples. The drug that was chosen was the exact same as Sample 5 stored in the library. So the ideal hierarchy for the top five matches should be in order as

1. Sample 5
2. Sample 6 or 7
3. Sample 7 or 6
4. Sample 3 or 4
5. Sample 4 or 3

6.2 Tabulating Comparisons and Graphs

To start testing the samples, we first make the library of known samples i.e. Sample 1 to Sample 10. The libraries can be made according to user choice. In our case two libraries were made. First one was made at 660 to 1699 nm wavelength range and the second one was made at 781 to 1640 nm wavelength range which is a subset of original range. The results are shown when the unknown sample data is being compared with all the known samples in the library at these two wavelength ranges, in different ways and using different tools for comparison. It is important to note that OPUS software does not utilize as many choices as given by our software tool. It does not have the choice of averaging the sample readings and converting them to absorption readings. So the results for OPUS software are obtained using reflection data only. Since it cannot average out, so only one measurement is performed for each sample. As OPUS cannot change the wavelength range so all OPUS calculations are done in the original wavelength range i.e. 660 to 1699 nm. The Tables 6.1 to 6.12 have six rows and five columns. The first row lists the software used for calculation and the rest of the rows show sample names in order of closest to farthest match with the distance values given in some cases. The columns give information about the wavelength range selected and the type of tool chosen to analyze the data. The following sections are arranged according to different ways of data pretreatment and comparison.

6.2.1 Using Absorption Data

Table 6.1 shows the comparison of samples using Absorption data on various tools. There is no pretreatment of data done in this case.

Table 6.1 Comparison of samples using Absorption Data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 4 (0.760081)	Sample 4 (0.760081)	Sample 4 (0.693906)	NA
2nd	Sample 1 (0.879485)	Sample 1 (0.879485)	Sample 1 (0.763961)	NA
3rd	Sample 3 (0.890142)	Sample 3 (0.890142)	Sample 3 (0.828211)	NA
4th	Sample 6 (1.434993)	Sample 6 (1.434993)	Sample 6 (1.332425)	NA
5th	Sample 5 (1.563312)	Sample 5 (1.563312)	Sample 5 (1.434270)	NA

Figure 6.1(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.1(b) shows the graph for the absorption readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

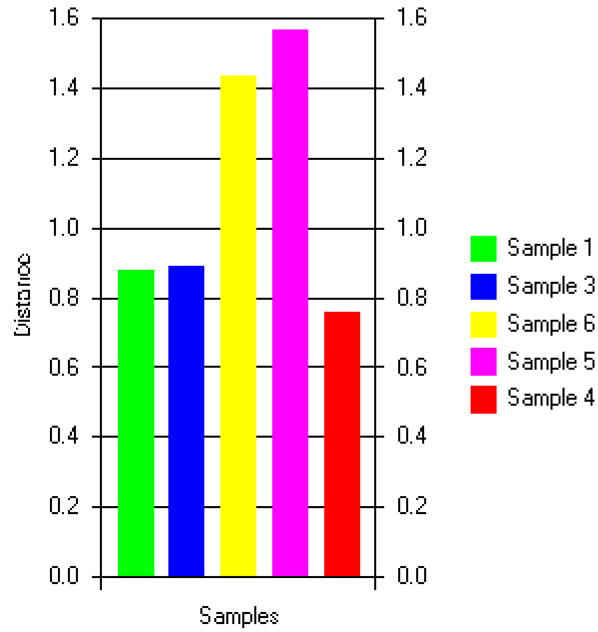

Fig 6.1(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using Absorption data.

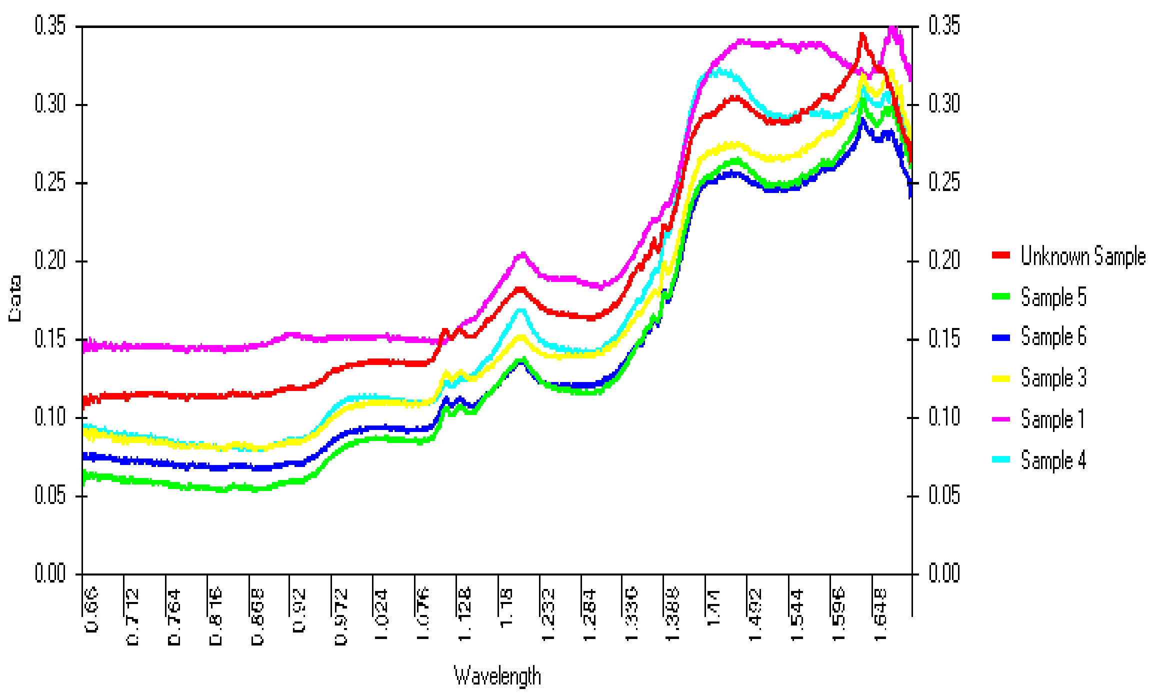

Fig 6.1(b) Absorption data readings for the first five and unknown samples at 660 to 1699 nm.

6.2.2 Using Centered Data

Table 6.2 gives the comparison results of samples using Centered data technique.

Table 6.2 Comparison of samples using Centered Data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 6 (0.144489)	Sample 6 (0.144489)	Sample 6 (0.0636309)	NA
2nd	Sample 7 (0.160857)	Sample 7 (0.160857)	Sample 3 (0.097838)	NA
3rd	Sample 3 (0.252595)	Sample 3 (0.252595)	Sample 7 (0.129695)	NA
4th	Sample 5 (0.313765)	Sample 5 (0.313765)	Sample 5 (0.192939)	NA
5th	Sample 1 (0.420347)	Sample 1 (0.420347)	Sample 1 (0.363604)	NA

Figure 6.2(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.2(b) shows the graph for the Centered readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

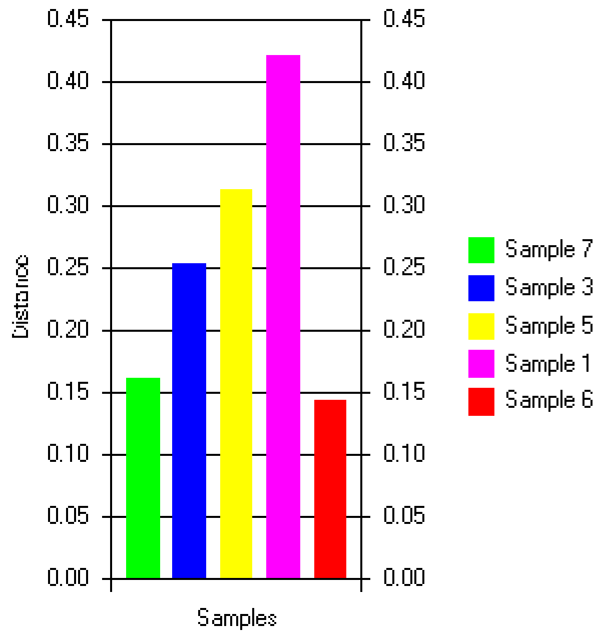

Fig 6.2(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using Centered data.

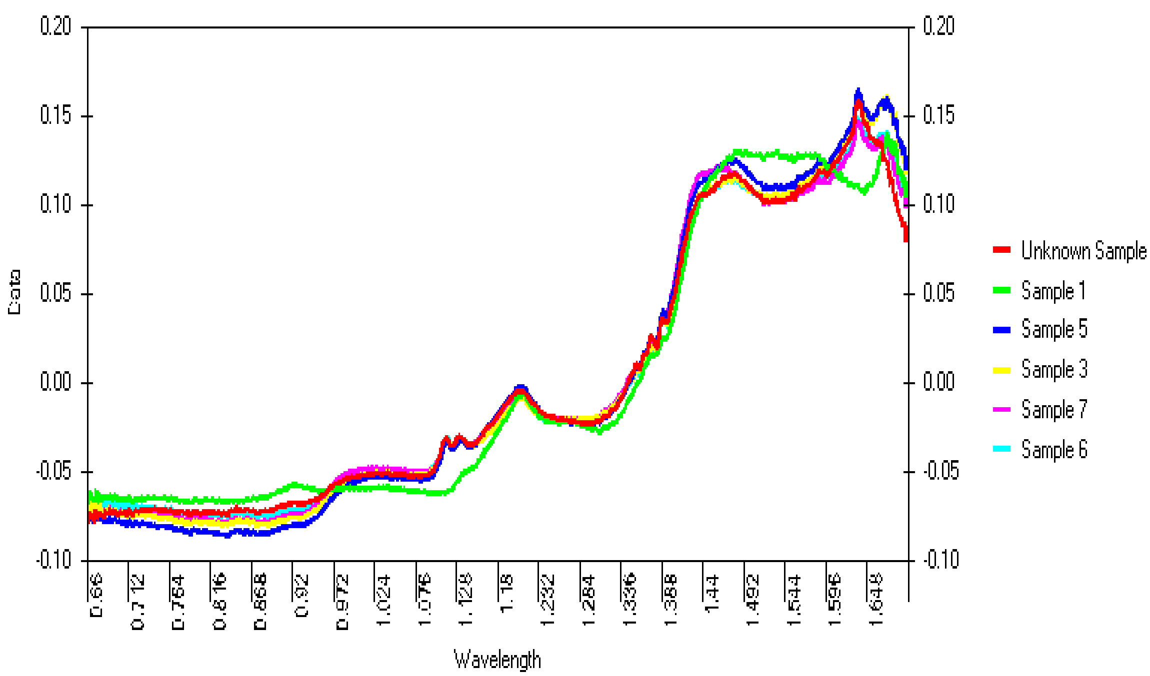

Fig 6.2(b) Centered data readings for first five and unknown samples at 660 to 1699 nm.

6.2.3. Using Centered and Scaled Data

Comparison of results using Centered and Scaled data is shown in Table 6.3.

Table 6.3 Comparison of samples using Centered and Scaled Data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 6 (1.908462)	Sample 6 (1.908462)	Sample 6 (0.863385)	Sample 3
2nd	Sample 7 (2.040833)	Sample 7 (2.040833)	Sample 3 (1.077399)	Sample 5
3rd	Sample 5 (2.404835)	Sample 5 (2.404835)	Sample 5 (1.165629)	Sample 7
4th	Sample 3 (2.79522)	Sample 3 (2.79522)	Sample 7 (1.748727)	Sample 6
5th	Sample 4 (3.91221)	Sample 4 (3.91221)	Sample 4 (3.46618)	Sample 4

Figure 6.3(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.3(b) shows the graph for the Centered and Scaled readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

.

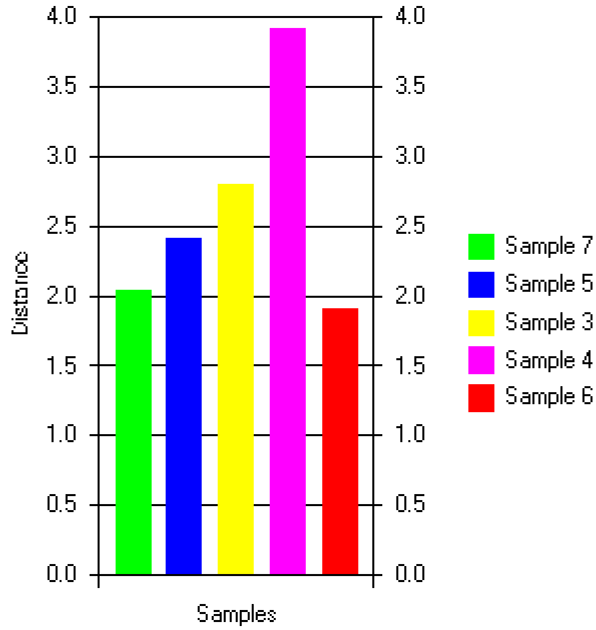

Fig 6.3(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using Centered and Scaled data.

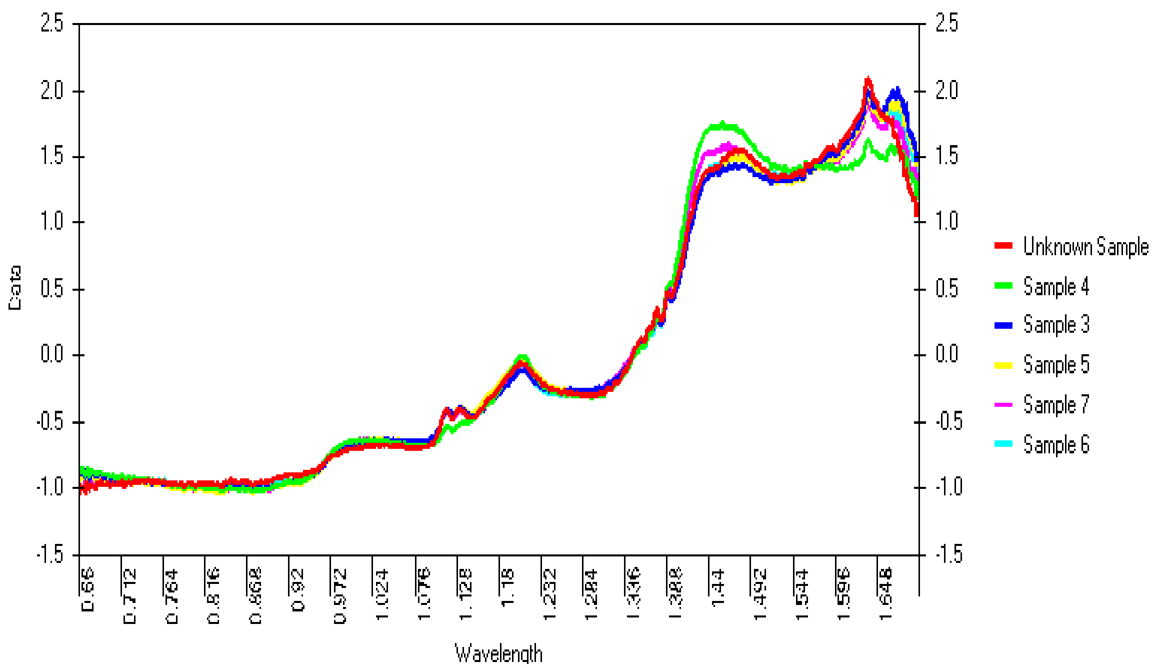

Fig 6.3(b) Centered and Scaled data readings for first five and unknown samples at 660 to 1699 nm.

6.2.4 Using Min-Max Scaled Data

Table 6.4 gives the comparison results using Min-Max Scaled data.

Table 6.4 Comparison of samples using Min-Max Scaled Data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 6 (0.876019)	Sample 6 (0.876019)	Sample 3 (0.430241)	NA
2nd	Sample 3 (0.976741)	Sample 3 (0.976741)	Sample 6 (0.537052)	NA
3rd	Sample 5 (0.985864)	Sample 5 (0.985864)	Sample 5 (0.676191)	NA
4th	Sample 7 (1.15073)	Sample 7 (1.15073)	Sample 7 (1.10524)	NA
5th	Sample 4 (2.246342)	Sample 4 (2.246342)	Sample 4 (2.22988)	NA

Figure 6.4(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.4(b) shows the graph for the Min-Max Scaled readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

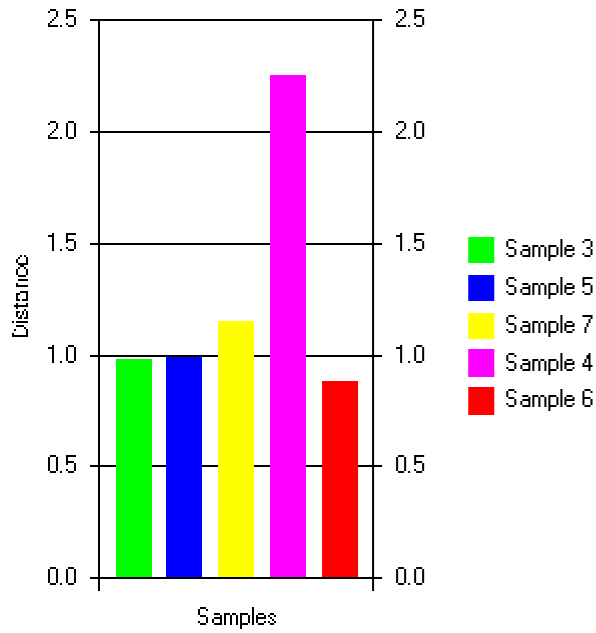

Fig 6.4(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using Min-Max Scaled data.

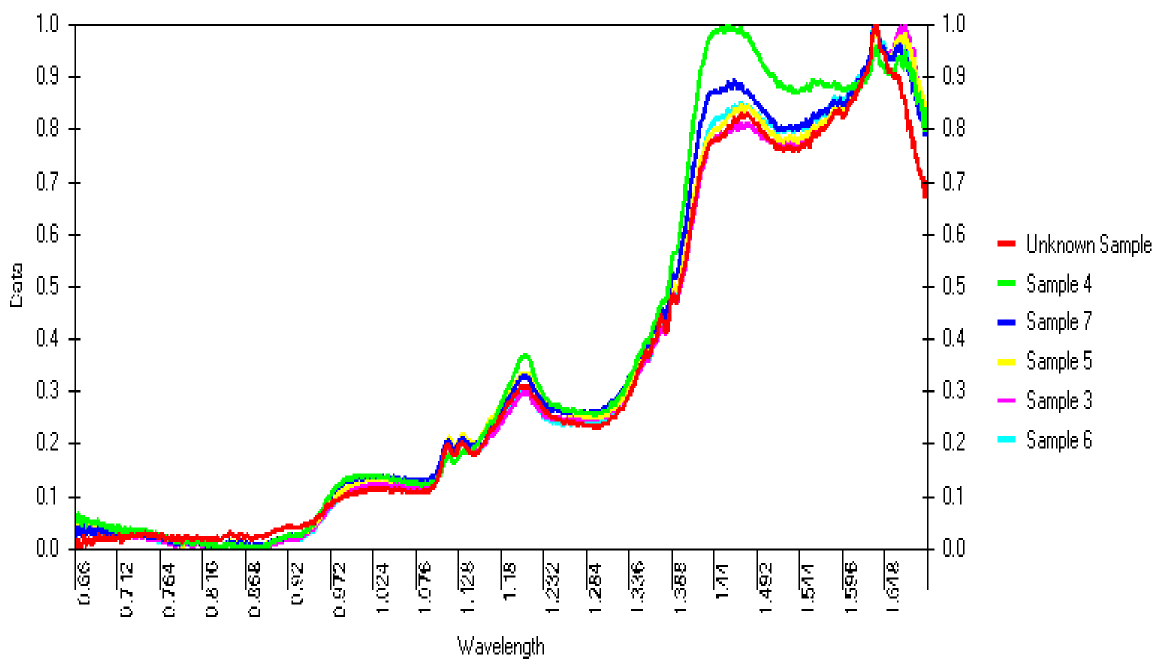

Fig 6.4(b) Min-Max Scaled data readings for first five and unknown samples at 660 to 1699 nm.

6.2.5 First Derivation Using Absorption Data

.Table 6.5 gives the results using 25 Point 1st Derivative on Absorption data.

Table 6.5 Comparison of samples using 25 Point 1st Derivative on Absorption data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 7 (0.00503)	Sample 7 (0.00503)	Sample 5 (0.001945)	NA
2nd	Sample 6 (0.005548)	Sample 6 (0.005548)	Sample 6 (0.002633)	NA
3rd	Sample 5 (0.006163)	Sample 5 (0.006163)	Sample 3 (0.002670)	NA
4th	Sample 3 (0.007004)	Sample 3 (0.007004)	Sample 7 (0.003488)	NA
5th	Sample 4 (0.01018)	Sample 4 (0.01018)	Sample 4 (0.008144)	NA

Figure 6.5(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.5(b) shows the graph for 25 Point 1st Derivative on Absorption readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

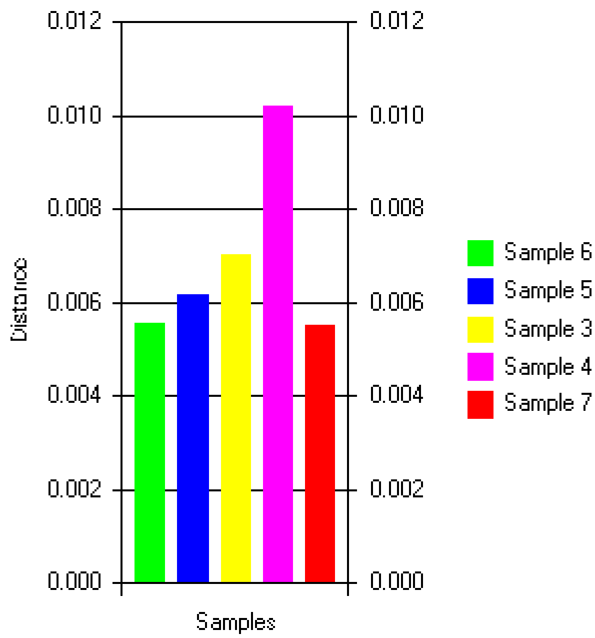

Fig 6.5(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using 25 Point 1st Derivation on Absorption data.

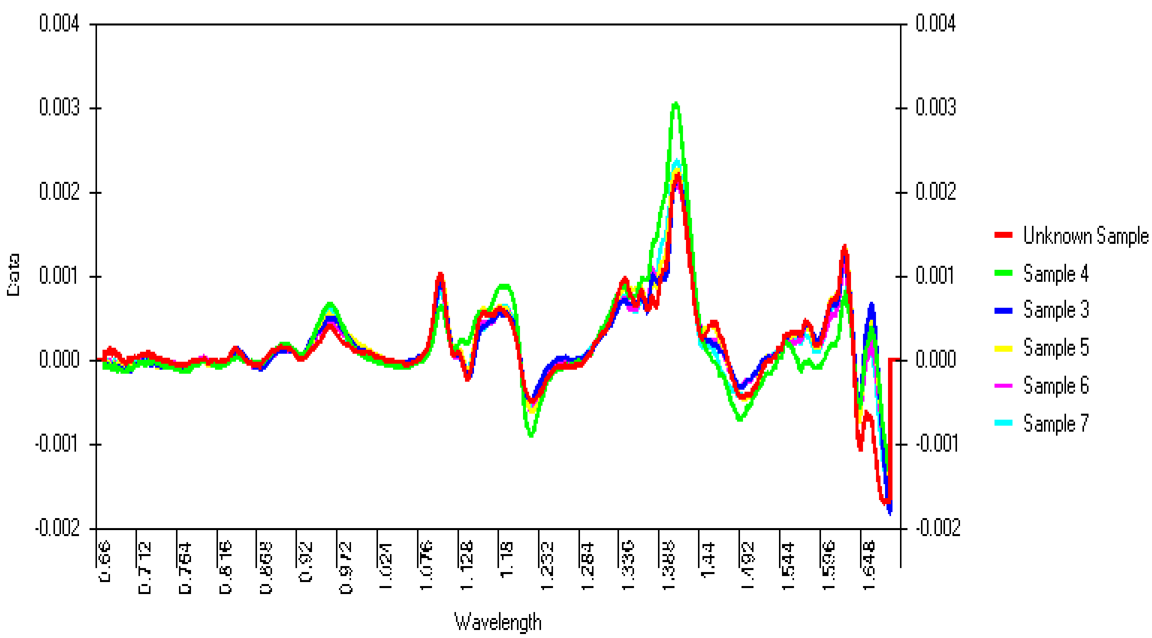

Fig 6.5(b) Graph of 25 Point 1st Derivative method using Absorption data readings for the first five samples and unknown sample at 660 to 1699 nm wavelength range.

6.2.6 First Derivation Using Centered Data

Table 6.6 gives the results using 25 Point 1st Derivative on Centered data.

Table 6.6 Comparison of samples using 25 Point 1st Derivative on Centered data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 7 (0.00503)	Sample 7 (0.00503)	Sample 5 (0.001945)	NA
2nd	Sample 6 (0.005548)	Sample 6 (0.005548)	Sample 6 (0.002633)	NA
3rd	Sample 5 (0.006163)	Sample 5 (0.006163)	Sample 3 (0.002670)	NA
4th	Sample 3 (0.007004)	Sample 3 (0.007004)	Sample 7 (0.003488)	NA
5th	Sample 4 (0.01018)	Sample 4 (0.01018)	Sample 4 (0.008144)	NA

Figure 6.6(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.6(b) shows the graph for 25 Point 1st Derivative on Centered readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

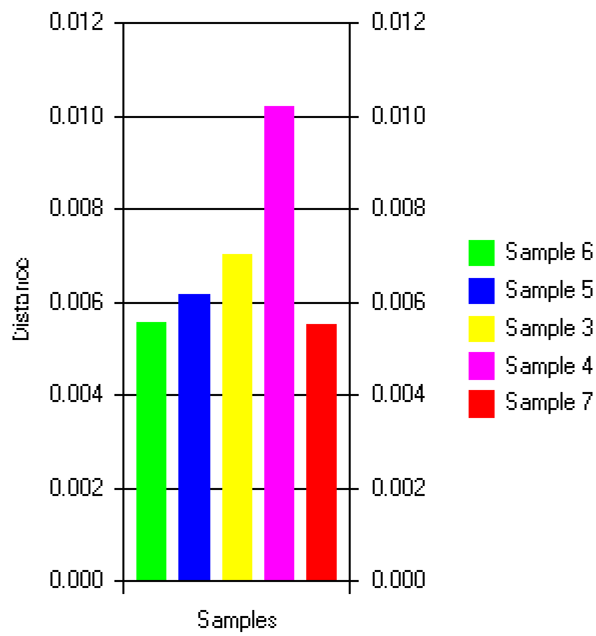

Fig 6.6(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using 25 Point 1st Derivation on Centered data.

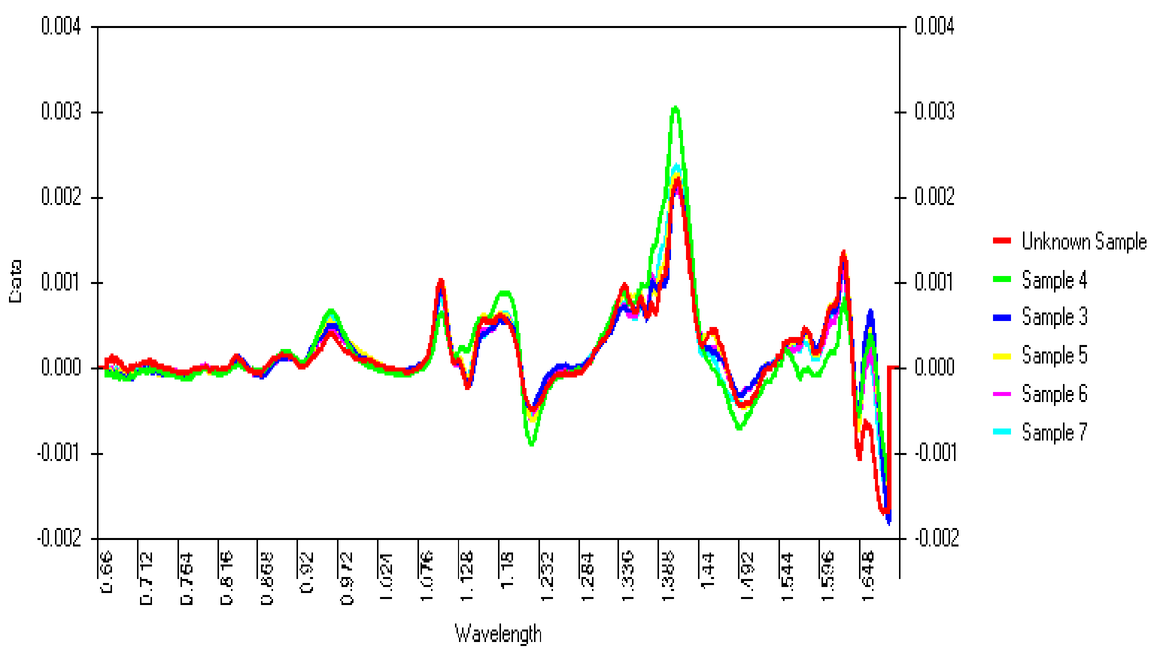

Fig 6.6(b) Graph of 25 Point 1st Derivative method using Centered data readings for the first five samples and unknown sample at 660 to 1699 nm wavelength range.

6.2.7 First Derivation Using Centered and Scaled Data

Table 6.7 gives the comparison of samples using 25 Point 1st Derivative method on Centered and Scaled data.

Table 6.7 Comparison of samples using 25 Point 1st Derivative on Centered and Scaled data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 7 (0.072587)	Sample 7 (0.072587)	Sample 5 (0.023584)	Sample 5
2nd	Sample 6 (0.073322)	Sample 6 (0.073322)	Sample 6 (0.036572)	Sample 6
3rd	Sample 5 (0.081603)	Sample 5 (0.081603)	Sample 3 (0.037921)	Sample 3
4th	Sample 3 (0.093041)	Sample 3 (0.093041)	Sample 7 (0.047815)	Sample 7
5th	Sample 4 (0.119779)	Sample 4 (0.119779)	Sample 4 (0.088067)	Sample 4

Figure 6.7(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.7 (b) shows the graph for 25 Point 1st Derivative on Centered and Scaled readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

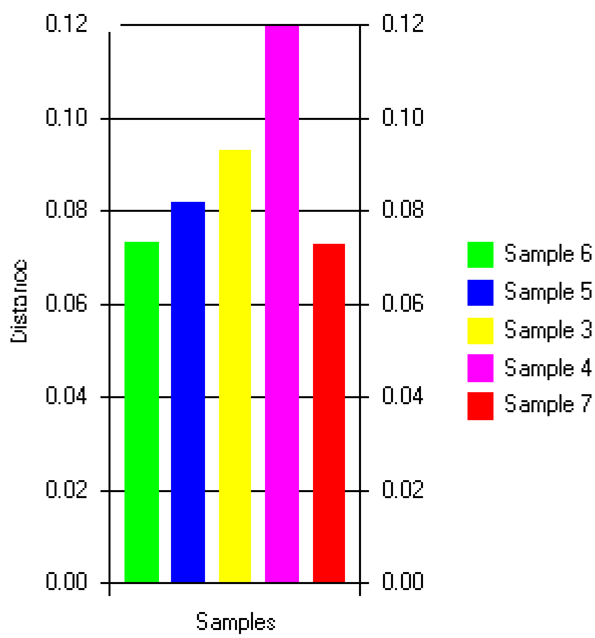

Fig 6.7(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using 25 Point 1st Derivation on Centered and Scaled data.

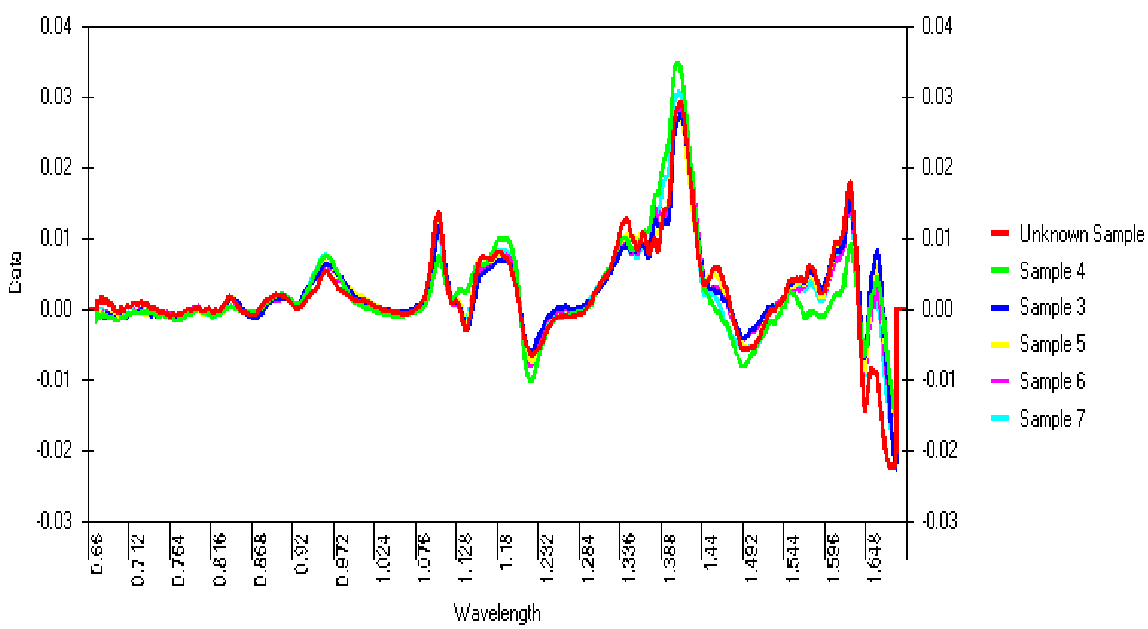

Fig 6.7(b) Graph of 25 Point 1st Derivative method using Centered and Scaled data readings for the first five samples and unknown sample at 660 to 1699 nm wavelength range.

6.2.8 First Derivation Using Min-Max Scaled Data

Table 6.8 gives the comparison of samples using 25 Point 1st Derivative method on data obtained from Min-Max Scaling.

Table 6.8 Comparison of samples using 25 Point 1st Derivative on Min-Max Scaled data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 6 (0.023137)	Sample 6 (0.023137)	Sample 5 (0.007245)	NA
2nd	Sample 7 (0.023697)	Sample 7 (0.023697)	Sample 6 (0.011269)	NA
3rd	Sample 5 (0.025960)	Sample 5 (0.025960)	Sample 3 (0.011678)	NA
4th	Sample 3 (0.029702)	Sample 3 (0.029702)	Sample 7 (0.015699)	NA
5th	Sample 4 (0.041935)	Sample 4 (0.041935)	Sample 4 (0.032522)	NA

Figure 6.8(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.8 (b) shows the graph for 25 Point 1st Derivative on Min-Max Scaled readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

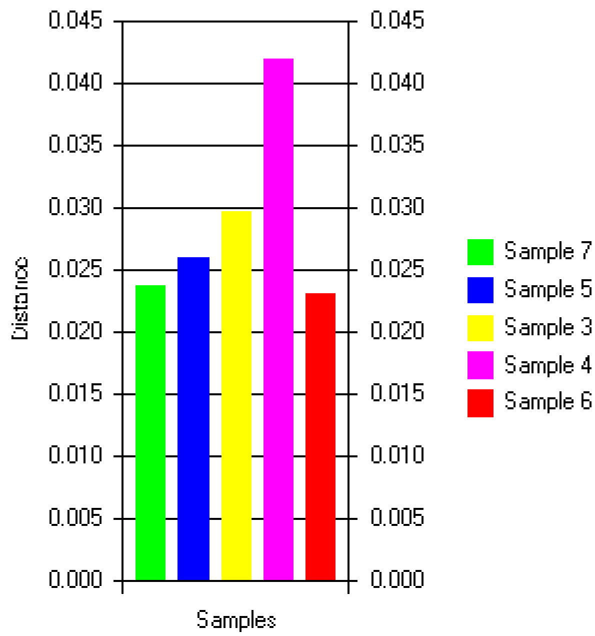

Fig 6.8(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using 25 Point 1st Derivation on Min-Max Scaled data.

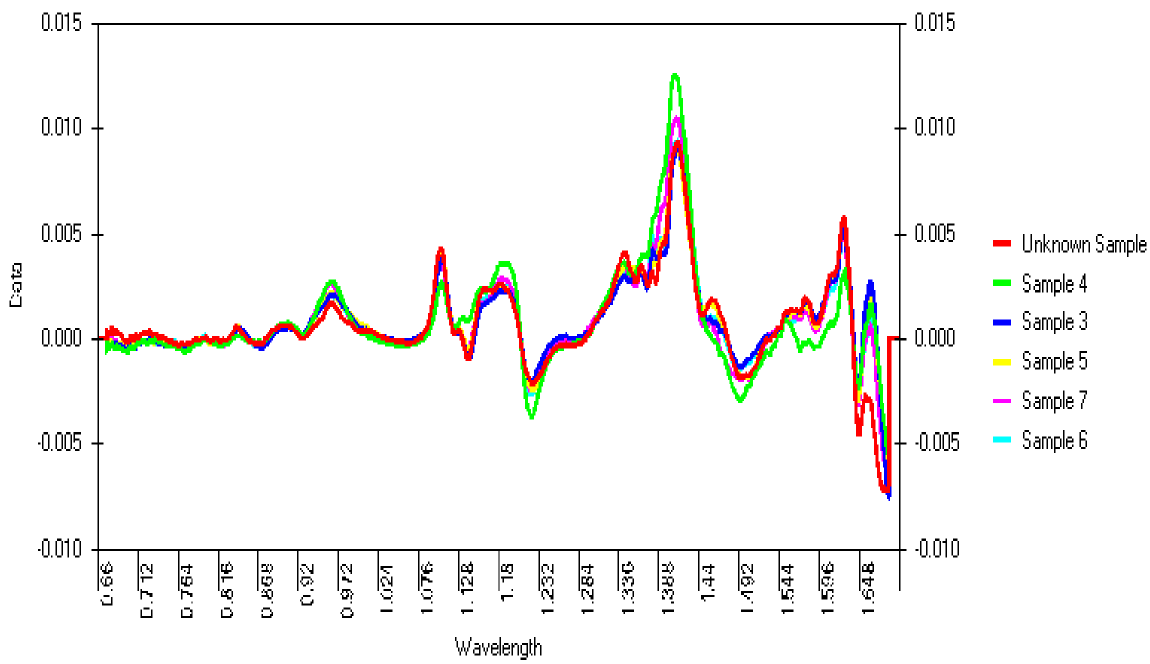

Fig 6.8(b) Graph of 25 Point 1st Derivative method using Min-Max Scaled data readings for the first five samples and unknown sample at 660 to 1699 nm wavelength range.

6.2.9 Second Derivation Using Absorption Data

Table 6.9 gives the comparison of samples using 25 Point 2^{nd} Derivative on Absorption data.

Table 6.9 Comparison of samples using 25 Point 2^{nd} Derivative on Absorption data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1^{st}	Sample 7 (0.000427)	Sample 7 (0.000427)	Sample 5 (0.000261)	NA
2^{nd}	Sample 5 (0.000497)	Sample 5 (0.000497)	Sample 7 (0.000305)	NA
3^{rd}	Sample 6 (0.000525)	Sample 6 (0.000525)	Sample 6 (0.000377)	NA
4^{th}	Sample 3 (0.000650)	Sample 3 (0.000650)	Sample 3 (0.000403)	NA
5^{th}	Sample 4 (0.000680)	Sample 4 (0.000680)	Sample 4 (0.000519)	NA

Figure 6.9(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.9 (b) shows the graph for 25 Point 2^{nd} Derivative on Absorption readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

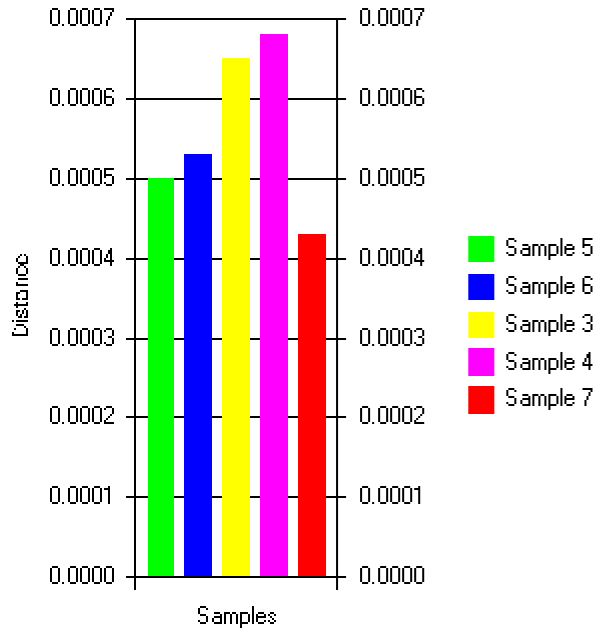

Fig 6.9(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using 25 Point 2nd Derivation on Absorption data.

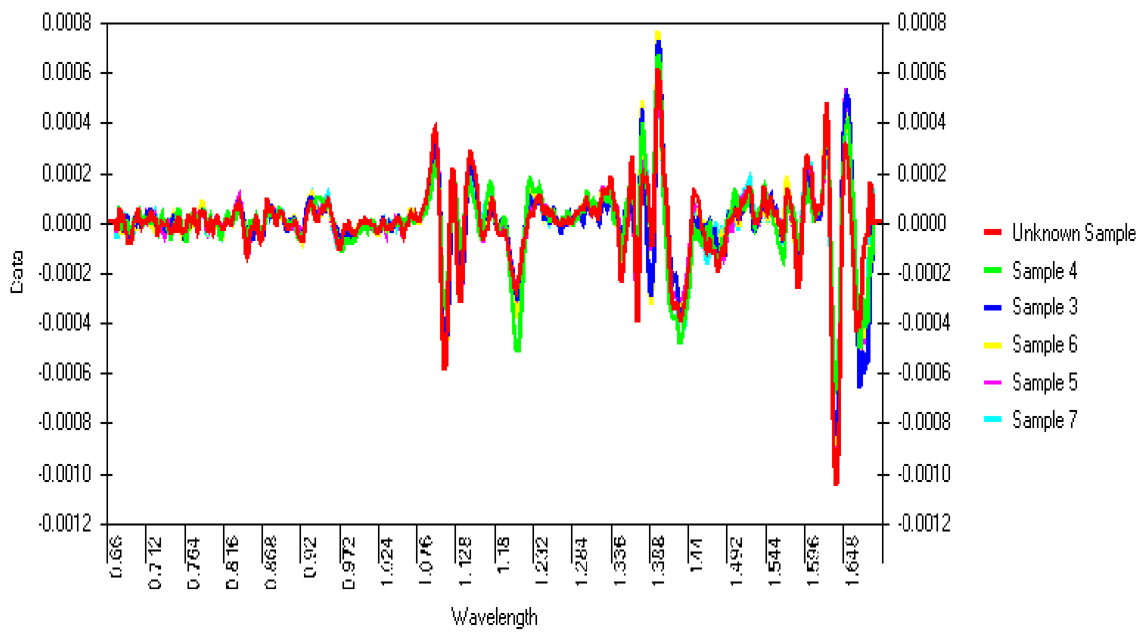

Fig 6.9(b) Graph of 25 Point 2nd Derivative method using Absorption data readings for the first five samples and unknown sample at 660 to 1699 nm wavelength range.

6.2.10 Second Derivation Using Centered Data

Table 6.10 gives the comparison of samples using 25 Point 2nd Derivative on Centered data.

Table 6.10 Comparison of samples using 25 Point 2nd Derivative on Centered data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 7 (0.000427)	Sample 7 (0.000427)	Sample 5 (0.000261)	NA
2nd	Sample 5 (0.000497)	Sample 5 (0.000497)	Sample 7 (0.000305)	NA
3rd	Sample 6 (0.000525)	Sample 6 (0.000525)	Sample 6 (0.000377)	NA
4th	Sample 3 (0.000650)	Sample 3 (0.000650)	Sample 3 (0.000403)	NA
5th	Sample 4 (0.000680)	Sample 4 (0.000680)	Sample 4 (0.000519)	NA

Figure 6.10(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.10 (b) shows the graph for 25 Point 2nd Derivative on Centered readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

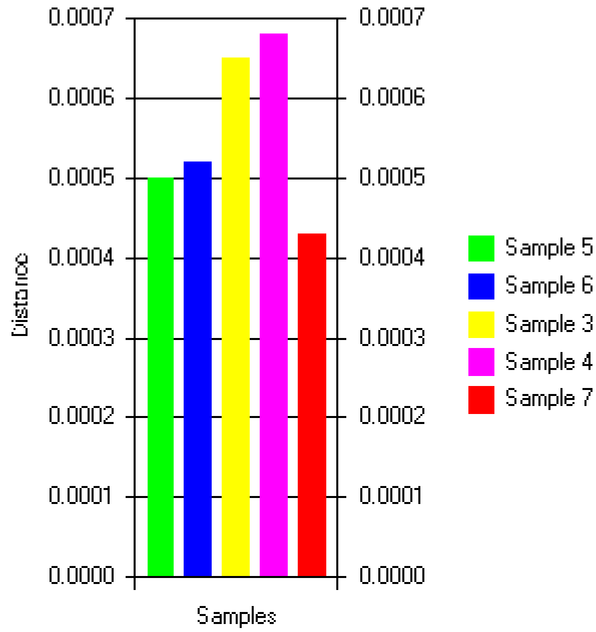

Fig 6.10(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using 25 Point 2nd Derivation on Centered data.

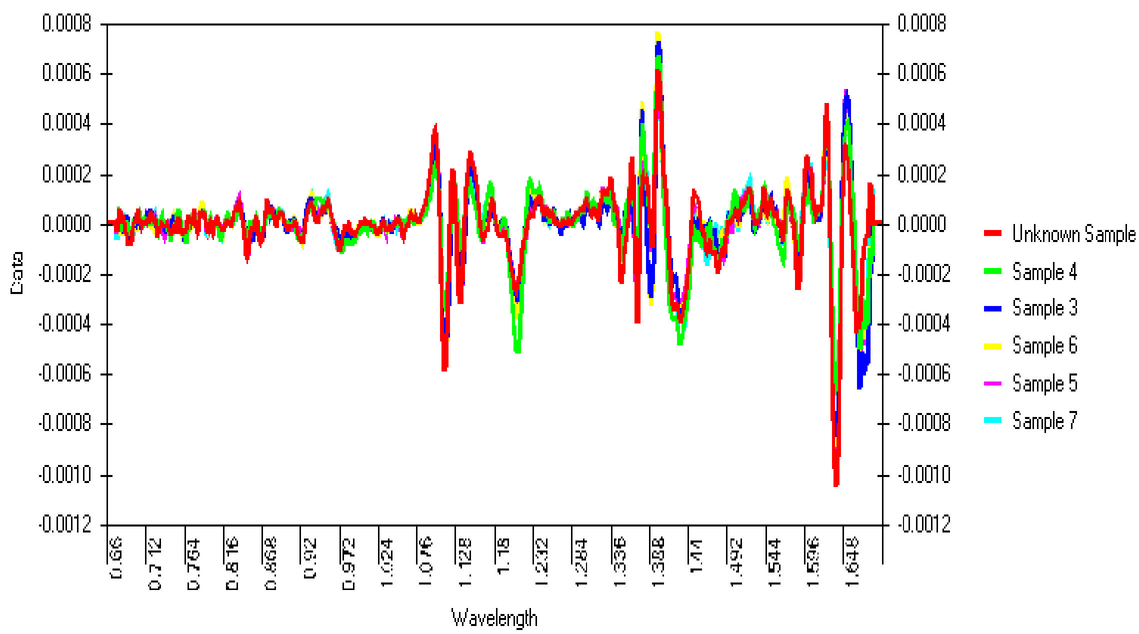

Fig 6.10(b) Graph of 25 Point 2nd Derivative method using Centered data readings for the first five samples and unknown sample at 660 to 1699 nm wavelength range.

6.2.11 Second Derivation Using Centered and Scaled Data

Table 6.11 shows the comparison of various samples using 25 Point 2nd Derivative techniques on Centered and Scaled data. Some of the observations are

Table 6.11 Comparison of samples using 25 Point 2nd Derivative on Centered and Scaled data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 7 (0.005662)	Sample 7 (0.005662)	Sample 5 (0.0003714)	Sample 7
2nd	Sample 5 (0.006320)	Sample 5 (0.006320)	Sample 7 (0.004262)	Sample 5
3rd	Sample 6 (0.006931)	Sample 6 (0.006931)	Sample 6 (0.005244)	Sample 6
4th	Sample 3 (0.008379)	Sample 3 (0.008379)	Sample 3 (0.005600)	Sample 3
5th	Sample 4 (0.008863)	Sample 4 (0.008863)	Sample 4 (0.006914)	Sample 4

Figure 6.11(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.11(b) shows the graph for 25 Point 2nd Derivative on Centered and Scaled readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

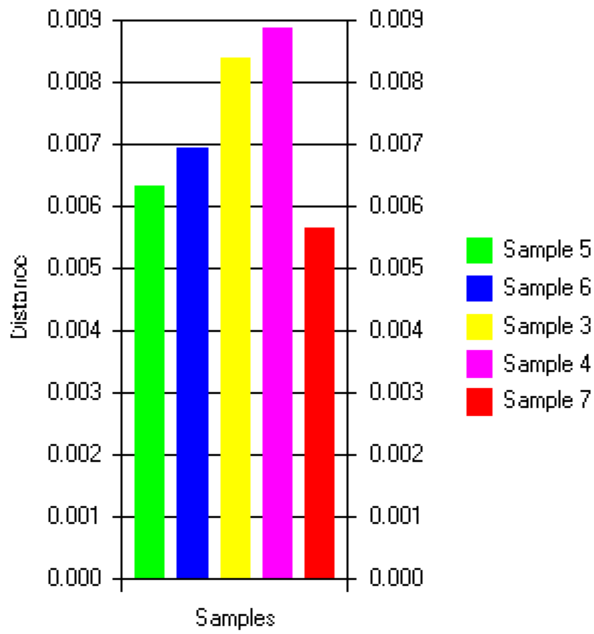

Fig 6.11(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using 25 Point 2nd Derivation on Centered and Scaled data.

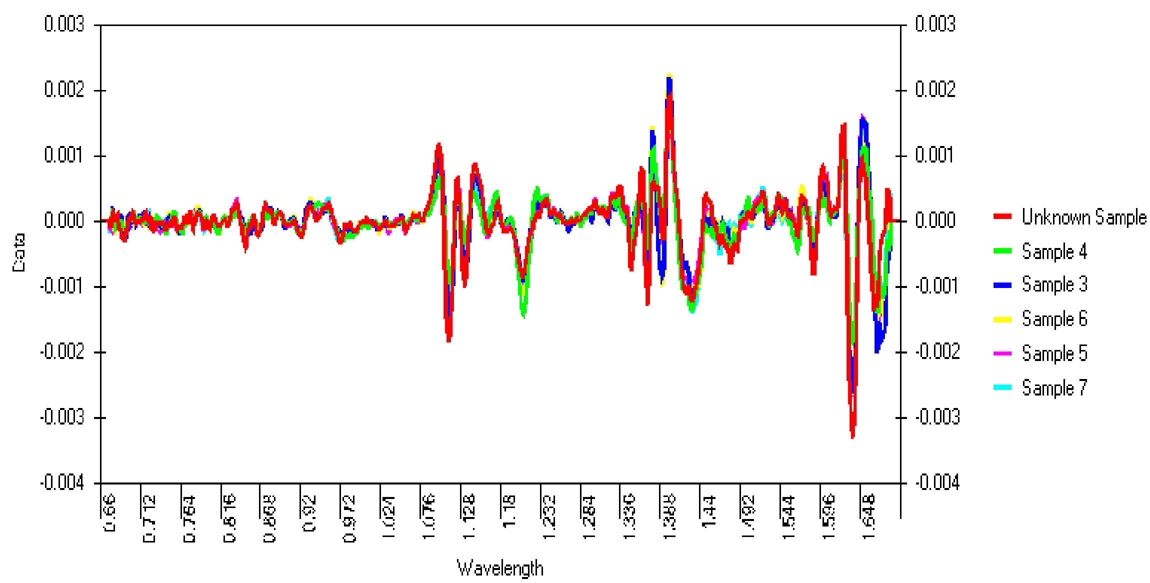

Fig 6.11(b) Graph of 25 Point 2nd Derivative method using Centered and Scaled data readings for the first five samples and unknown sample at 660 to 1699 nm wavelength range.

6.2.12 Second Derivation Using Min-Max Scaled Data

Table 6.12 compares the samples using 25 Point 2nd Derivative technique on Min-Max Scaled data.

Table 6.12 Comparison of samples using 25 Point 2nd Derivative on Min-Max Scaled data.

Sample Position	660-1699 nm Excel Calculations	660-1699 nm Software Calculations	781-1640 nm Software Calculations	660-1699 nm OPUS Calculations
1st	Sample 7 (0.000427)	Sample 7 (0.000427)	Sample 5 (0.000261)	NA
2nd	Sample 5 (0.000497)	Sample 5 (0.000497)	Sample 7 (0.000305)	NA
3rd	Sample 6 (0.000525)	Sample 6 (0.000525)	Sample 6 (0.000377)	NA
4th	Sample 3 (0.000650)	Sample 3 (0.000650)	Sample 3 (0.000403)	NA
5th	Sample 4 (0.000680)	Sample 4 (0.000680)	Sample 4 (0.000519)	NA

Figure 6.12(a) shows the distance bar graph of first five closest samples to unknown sample at 660 to 1699 nm wavelength range and Figure 6.12(b) shows the graph for 25 Point 2nd Derivative on Min-Max Scaled readings of first five closest samples and the unknown sample at 660 to 1699 nm wavelength range.

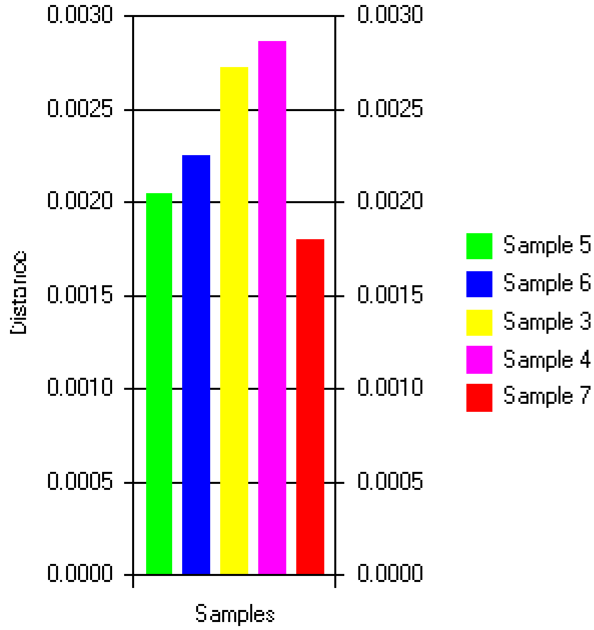

Fig 6.12(a) Distance graph of first five closest samples at 660 to 1699 nm wavelength range using 25 Point 2nd Derivation on Min-Max Scaled data.

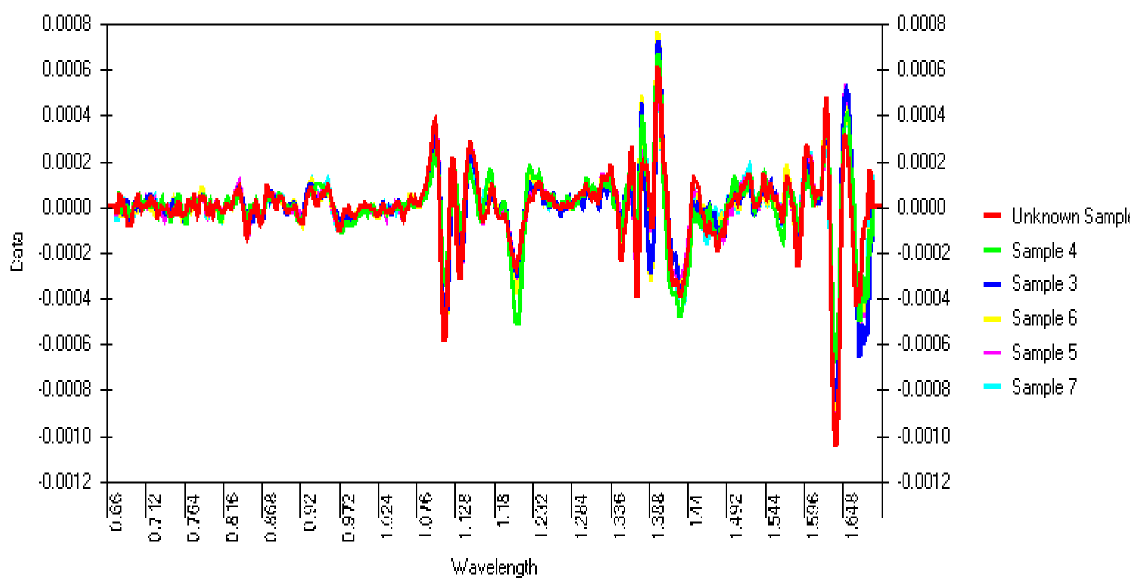

Fig 6.12(b) Graph of 25 Point 2nd Derivative method using Min-Max Scaled data readings for the first five samples and unknown sample at 660 to 1699 nm wavelength range.

.

6.3 Discussions

Various trends are seen in the arrangement of samples in Tables 6.1 to 6.12. For all the tables, the second column of each table matches exactly with the third column of that table. Column two gives the results of the comparison of samples using Microsoft Excel whereas the third column gives the results of comparison using our software tool both at 660 to 1699 nm wavelength range. It can also be noticed that not only the hierarchy of samples but even the distance of unknown sample from the known samples matches each other. This validates the software developed has all the methods implemented correctly. Table 6.1 compares the samples using the raw data (Absorption data). The reflection data taken from the spectrometer is simply converted to absorption data. No pretreatment of data is done at this level. Since no pretreatment of samples is done samples like “Sample 1” which is nowhere near the unknown sample is in the list of top five closest samples. Also other samples that appear in the list are not arranged in order of closeness of their concentrations to the concentration of unknown sample. Even changing the wavelength range to 781 to 1640 nm does not have any effect in this case. The only effect of changing the wavelength range is in the distance. The distance has decreased. This is due to the fact that new wavelength range has lesser points of comparison then the original range. Table 6.2 to Table 6.4 shows the comparison results of pretreated data. It is noticed that in Table 6.2 (using Centered data), “Sample 1” has moved down in the list. Furthermore in Table 6.3 (using Center and Scaled data) and Table 6.4 (using Min-Max Scaled data), “Sample 1” has completely disappeared from the list of closest samples. All other samples like “Sample 3”, “Sample 4”, “Sample 5”, “Sample 6” and “Sample 7” which are made of similar components as the unknown sample now appear in the list of top five closest samples. But still these samples are not arranged according to their closest concentration match to the unknown sample in Table 6.2 and Table 6.4. But in Table 6.3 using Centered and Scaled data the samples are arranged in accordance to their concentration match to the unknown sample. In all the three tables, results obtained at 660 to 1699 nm wavelength range are better than the results at 781 to 1699 nm wavelength range. This is due to the fact that noise has not made a significant effect when we use these pretreatment methods. OPUS gives results using Centered and Scaled data at the 660 to 1699 nm wavelength range. So in Table 6.3 OPUS shows “Sample 3” at the top of the list. This sample is not similar to the unknown sample in terms on its concentration. If OPUS results are compared with the results obtained from our software tool at 660 to 1699 nm wavelength range then it is seen that our results are better than OPUS results because in our case all samples with same concentration as unknown sample are in the top three spots of the list.

Tables 6.5, 6.6, 6.6 and 6.8 use 25 Point 1st Derivative technique for sample comparison on Absorption data, Centered data, Centered Scaled data and Min-Max Scaled data respectively. In these tables at 660 to 1699 nm wavelength range, samples are arranged in accordance to their closeness to the unknown sample in terms of their concentrations. But “Sample 5” which is the exactly same as the unknown sample, appears on the third place in the list. Whereas if the wavelength range is changed to 781 to 1640 nm then “Sample 5” appears on the top spot. But at this range “Sample 7” appears after “Sample 3” which is not true because “Sample 3 concentration is different than the concentration of the unknown sample but “Sample 7” concentration is the same as the unknown sample. OPUS results are compared in Table 6.7 where 25 Point 1st Derivative is done on Centered and Scaled data at 660 to 1699 nm wavelength range. OPUS shows "Sample 5" at top of the list which is desirable but it shows "Sample 3" before "Sample 7" which is not desired. Our results are more consistent at this level if compared to results got from OPUS software.

Tables 6.9, 6.10, 6.11 and 6.12 use 25 Point 2nd Derivative technique for sample comparison on Absorption data, Centered data, Centered and Scaled data and Min-Max Scaled data respectively. At 660 to 1699 nm wavelength range in all comparisons, samples are arranged according to closeness in concentration to the unknown sample. But still "Sample 5" is not at the top spot. OPUS gives exactly the same results as given by our software tool at 660 to 1699 nm wavelength range in Table 6.11. It means that "Sample 5" is not at the top spot even with OPUS. Now if the wavelength range is changed to 781 to 1640 nm in all the comparisons of 25 Point 2nd Derivative, "Sample 5" comes at the top spot followed by "Sample 7" and "Sample 6". This is the ideal condition.

Figure 6.1(a) to Figure 6.12(a) shows the distance of the closest five samples to the unknown sample in form of bar graphs given by the software tool at 660 to 1699 nm range of wavelength. Figure 6.1(b) to Figure 6.12(b) shows the graph of readings for the first five closest samples depending upon the method of comparison selected. Figure 6.1(b) shows the absorption reading of closest five samples and unknown sample. Since the data is not treated it is noticed that a large offset exists between different samples. Figure 6.2(b) shows centered data of closest samples. It is seen that the offset has been compensated to an extent as the data was centered and the mean of the experimental distribution was brought to zero [21]. While centering corrects the model offset, scaling is a way to fit models with a weighted least squares function [22]. So in Figure 6.3(b) the offset is better compensated.

Figure 6.5(b) to Figure 6.8(b) shows the graph of readings for first five closest samples and unknown sample using 25 Point 1^{st} Derivation method on pretreated data. It is seen in all the graphs that the first derivative of pretreated data are centered along the zero line and the central moments are conserved [24]. Figure 6.9(b) to Figure 6.12(b) shows the graph of readings of first five samples along with the unknown sample using 25 Point 2^{nd} Derivative technique on pretreated data.

So it is seen that good results are not achieved by simply using raw data. The data has to be pretreated and then compared. Out of the various techniques that have been developed and used, the 25 Point 2^{nd} Derivative has given the best results. OPUS software results can be compared with any technique that uses Centered and Scaled data but still it is not able to yield as accurate results as given by our software tool. Moreover we cannot use OPUS to remove noise which is present in the beginning and the end of spectral range because it does not have any utility by which we can ignore data in the beginning or the end of original wavelength range.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

The work accomplished in this thesis presents a novel tool that has been developed in particular to analyze the data obtained from a NIR-MOEMS spectrometer. Since spectroscopy is an important technique for analyzing various stages of production, a matching tool has been developed to meet the requirements of data analysis. Raw data available from spectroscope cannot be used directly but it has to be pretreated using various mathematical operations before it is fit for comparison. In this thesis we have first introduced NIR spectroscopy and MOEMS spectrometer. Then various mathematical functions that have been used to develop the software tool are discussed in detail. The functioning of the software tool utilities are explained with the help of flow charts. Some of the salient features of the software tool are also discussed. The results have been compared at different wavelength ranges and also with the results available from OPUS software. Our software tool provides an inexpensive and more accurate method of analyzing data available after NIR spectroscopy. The software tool is easy to understand and relatively easy to work with.

7.2 Future Work

The software tool that has been designed is tested using data available from different drug samples. It can be extended to samples available from food and beverages, polymers, petroleum products etc. The software tool can be made more dynamic if the need of making libraries at different wavelength ranges is eliminated and instead we have one library which contains only the data for different samples obtained from the spectrometer at original wavelength range. Now this library containing the data at original wavelength range is used each time to compare samples using different comparison techniques at different wavelength ranges depending on user choice. This means that the system will process the original data only after it gets the required command. However this would increase the processing time of the software but it would help in saving a lot of memory space on the computer. The results of transmission mode can also be compared with diffuse reflection mode using the tool. Principal Component Analysis technique can also be extended to further data processing. Implementation of regression analysis may provide with more accurate results. Since it is impossible to measure anything exactly, so the chances of occurrence of errors are high. These errors should be taken into consideration. So error analysis would also help in making the software tool more sensitive and accurate.

REFERENCES

[1] Otto,T.; Saupe,R.; Fritzsch,U.; Stock,V.; Bruch,R.; Gessner,T: *A Novel Dual-Detector Micro Spectrometer*, Photonics West 2005: Procedings of SPIE, Volume 5719, pp. 76-82, 2005.

[2] Smith, B.C.; *Fundamentals of Fourier Infrared Spectroscopy*, Boca Raton : CRC Press, pp. 55-86, 1996

[3] D. Burns, E. Ciurczak, *Handbook of Near Infrared Analysis,* Marcel-Dekker, Inc. New York, 2nd Edition, 2001.

[4] Otto,T.; Saupe,R.; Weiss, A.; Stock,V.; Bruch,R.; Gessner,T: *Principle and Applications of a new MOEMS-Spectrometer*, Proceedings of SPIE, San Jose 2006, Volume 6114, pp. 77 – 86, 2006.

[5] Herschel, William: *Experiments on the Refrangibility of the Invisible Rays of the Sun*, Philosophical Transactions of the Royal Society of London, Volume 1, pp. 22-23, 1800.

[6] W. Kauzmann, *Quatum Chemistry*, Academic press New York, pp. 667, 1957.

[7] Franklin E. Barton II, *Theory and principles of near infrared spectroscopy*, Spectroscopy Europe, Volume14/1, pp. 12-18, 2002.

[8] Jerry Workman and Lois Weye: *Practical Guide to Interpretive Near-Infrared Spectroscopy*, Published by CRC Press, 2007

[9] "Beer Lambert Law"; <http://en.wikipedia.org/wiki/Beer-Lambert_law>,Wikipedia, 2008, 19th January, 2009

[10] “Near Infrared Spectroscopy”; <http://en.wikipedia.org/wiki/Near_infrared_spectroscopy>, Wikipedia, 2008.

[11] “Introduction to NIR”, <http://www.asdi.com/ASD-600510_NIR-Introduction.pdf >, ASD, Document 600510 Rev. 2.

[12] Otto,T.; Saupe,R.; Weiss, A.; Fritzsch, M.; Stock,V.; Wiesner, K.; Lampe, U.; Gessner,T: *Dual-Detector Optical MEMS Spectrum Analyzer: advantages, applications and prospects,* Procedings of SPIE, Volume 6887, pp. 6887 0D, 2008.

[13] Gessner T.; Bonitz, J.; Kaufmann, C.; Kurth, S.; Specht, H.; *MEMS Based Micro Scanners: Components, Technologies and Applications*, 10th International Conference on New Actuators, pp. 193-198, 2006.

[14] Otto, T.; Saupe, R.; Bruch, R.; Gessner, T.; *Development of a Microspectrometer for the Infrared Range*, Conference Digest Twenty Seventh International Conference on Infrared and Millimeter Waves, Edited by R.J. Temkin., pp. 31-32, 2002.

[15] IrSys Infrared Spectrometer, User Documentation Version 1.04, Color Control, July 2007.

[16] Neumeister, V.; Henker, J.; Kaltenborn, G.; *Simultaneous Determination of Fecal Fat, Nitrogen and Water by Near-Infrared Reflectance Spectroscopy*, Journal of Pediatric Gastroenterology and Nutrition , Volume 25, 388-393, 1997.

[17] Osborne, B.; Fearn, T.; *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, John Wiley & Sons, Inc., New York, 2nd Edition, 1993.

[18] Valleur, M.; *Spectroscopic methods in refining and petrochemicals*, Petroleum Technology Quarterly, 1999.

[19] Feldhoff, R. et al., *Fast identification of packaging waste by near infrared spectroscopy with an InGaAs array spectrograph combined with neural networks*, J. Near Infrared Spectrosc. 3, pp. 3-9, 1995.

[20] Downey, G.: *Qualitative Analysis in the Near Infrared Region*, Analyst, Volume 119, pp. 2367-2375, November 1994.

[21] Howard Mark and Jerry Workman, *Chemometrics in Spectroscpy ,2007Amsterdam*; London: Elsevier/Academic Press.

[22] Rasmus Bro, Age K. Smilde: *Centering and Scaling in component analysis*; Journal of Chemometrics, Volume 17, pp. 16-33, January 2003.

[23] Udelhoven, T.; Emmerling, C.; Jarmer, T.: *Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study*, Journal of Plant and Soil, Volume 251, pp. 319-329, April 2003.

[24] BCH Turton: A *novel variant of Savitzky-Golay filter for spectroscopic applications*, Meas. Sci Technol., Volume 3, pp. 858-863, 1992

[25] Savitzky, A.; Golay, M.J.E; *Smoothing and Differentiations of Data by Simplified Least Square Procedures*, Anal. Chem., Volume 36 (8), pp. 1627-1639, 1964.

[26] Jianguo Sun: *Statistical analysis of NIR data: Data pretreatment*, Journal of Chemometrics, Volume 11, pp. 525-532, Dec 1998.

[27] Lindsay I Smith: *A tutorial on Principal Component Analysis*, February 2002.

[28] Dr. E. Garcia**:** *A tutorial on eigenvalues, eigenvectors and their properties. Includes step by step how-to calculations. An introduction to vector iteration, the Power Method and the Deflation Method is provided, July 2006.*