DETERMINATION OF TOTAL ACID NUMBER IN THE OPTIMIZATION OF OLEATE PRODUCTION BY USING FOURIER TRANSFORM INFRARED SPECTROSCOPY AND MULTIVARIATE CALIBRATION

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Chemistry

by
Nihan TOYGAR TÜRKÜN

July 2019
İZMİR

We approve the thesis of **Nihan TOYGAR TÜRKÜN**

Examining Committee Members:

Prof. Dr. Durmuş ÖZDEMİR
Department of Chemistry, İzmir Institute of Technology

Prof. Dr. Şerife YALÇIN
Department of Chemistry, İzmir Institute of Technology

Assoc. Prof. Dr. Levent PELİT
Department of Chemistry, Ege University

16 July 2019

Prof. Dr. Durmuş ÖZDEMİR
Supervisor, Department of Chemistry,
İzmir Institute of Technology

Prof. Dr. Ahmet Emin EROĞLU
Head of the Department of Chemistry

Prof. Dr. Aysun SOFUOĞLU
Head of the Graduate School of
Engineering and Sciences

ACKNOWLEDGMENTS

I would like to express my sincere gratefullness to a group of people who made this thesis possible. This thesis could not be written without;

My supervisor, Prof. Dr. Durmuş ÖZDEMİR, who mentored me with his great expertise. His great efforts and patience made this thesis possible to be created. His explanations to all of my questions in a clear and simple way made me encouraged to keep up with my hard work on a long way towards my carier. Additionaly, I would like to thank Prof. Dr. Durmuş Özdemir's research group of their efforts.

My thesis committee, Prof. Dr. Şerife YALÇIN and Assoc. Prof. Dr. Levent PELİT for their insightful comments and constructive questioning.

The company that I work for, Farper Kimya and its research team for their proffesional help and warm friendship.

And finally to my great family and friends for always being with me.

ABSTRACT

DETERMINATION OF TOTAL ACID NUMBER IN THE OPTIMIZATION OF OLEATE PRODUCTION BY USING FOURIER TRANSFORM INFRARED SPECTROSCOPY AND MULTIVARIATE CALIBRATION

Polyethylene glycol oleate (PEG-Oleate) is a non-ionic surfactant, and is an important emulsifier for water-oil systems. It is produced by reacting oleic acid and polyethylene glycol (PEG) under vacuum for around 4 hours and at 160 °C, in the presence of acid catalyst which is para toluene sulfonic acid (PTSA). The quality and process control of this production is determination of total acid number (TAN) by the standard method ASTM D974 which is a color indicator titration. Although titration is a simple method, it is relatively time consuming and prone to human error. Besides, the solvents used in titration method, are significantly unhealthy for humans.

The aim of this study is to develop fast and simple procedure for the determination of total acid number based on Fourier Transform Infrared Spectroscopy (FTIR) combined with multivariate calibration methods namely Genetic Inverse Least Squares (GILS) and Partial Least Squares (PLS). The reference total acid number of the samples collected during the esterification reaction, had been carried out by the ASTM D974 standart method and the Fourier Transform Infrared (FTIR) spectra of the same samples were also collected simultaneously with single reflection diamond Attenuated Total Reflectance (ATR) accessory.

Univariate calibration was applied on a specific wavenumber corresponding to the ester peak around 1739 cm^{-1}. Although the changes in the ester peak was showing an inrease associated to the esterification of the reactants, the results of the univariate calibration was unsucsesful. The best regression coefficient was found to be 0.997 by GILS method along with SECV and SEP as 2.295 and 2.694 mg KOH/g, respectively. The results of GILS showed that it is possible to monitor esterification process of PEG oleate.

ÖZET

FOURİER DÖNÜŞÜMLÜ KIZILÖTESİ SPEKTROSKOPİSİ VE ÇOK DEĞİŞKENLİ KALİBRASYON METHODU KULLANILARAK OLEATE ÜRETİMİNİN OPTİMİZASYONUNDA TOPLAM ASİT SAYISININ BELİRLENMESİ

Polietilen glikol oleat ürünü su içerisinde yağ emülsiyon sistemlerinde yaygın olarak kullanılan noniyonik emülgatördür. Polietilen glikol oleat ürünü oleik asit ile polietilen glikolün 4 saat boyunca 160ºC de vakum altında esterleşme reaksiyonu sonucunda elde edilmektedir. Esterleşme reaksiyonunda para toluen sülfonik asit katalizör kullanılmaktadır. Ürünün kalite ve proses kontrolü titrasyon yöntemi ile ASTM D974 standardına uygun olarak toplam asit numarası (TAN) belirlenir. Titrasyon uygulaması kolay bir yöntem olmasına rağmen kişi kaynaklı hatalara açıktır ve uygulama sırasında insan sağlığı için zararlı solventler kullanılmaktadır.

Bu tez çalışmasında, Fourier Dönüşümlü Kızılötesi Spektroskopisi (FTIR) kullanılarak geliştirilen genetik algoritmalara dayallı ters en küçük kareler (GILS) ve kısmi en küçük kareler (PLS) çok değişkenli kalibrasyon modelleri sayesinde çok daha hızlı ve kolay yeni bir toplam asit sayısı ölçme yöntemi geliştirilmesi hedeflenmiştir. Bu amaçla, polietilen glikol ve oleik asit esterleşme reaksiyonu boyunca numuneler alınmış ve alınan bu numunelerin ASTM D974 standardına göre toplam asit sayıları ölçüldü. Eş zamanlı, aynı örneklerin Fourier dönüşümlü kızılötesi (FTIR) spektrumları tek yansımalı elmas kristalli zayıflatılmış toplam yansıma (ATR) ekipmanı ile kaydedilmiştir.

Ester pikinin oluştuğu dalga sayısında (1739 cm^{-1}) tek değişkenli kalibrasyon methotlarıyla analizlenmiştir. Reaksiyon boyunca artmasına rağmen,tek değişkenli kalibrasyon methodunda toplam asit sayısı ile doğrusal olmayan bir ilişki ortaya koymuştur. GILS metodu ile oluşturulan modelin regresyon katsayı (R^2) değeri 0.997 olarak belirlenirken SECV ve SEP değerleri sırasıyla 2.295 ve 2.694 (mg KOH/g) olarak hesaplanmıştır. Bu çalışma sonucunda, polietilen glikol oleat'ın esterleşme sürecinin optimizasyonunun ve takibinin FTIR spektraskopisine dayılı olarak yapılabilmesine olanak tanıyan çok daha hızlı ve kolay kemometrik yeni bir metot geliştirilmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

Figure **Page**

LIST OF TABLES

Table **Page**

CHAPTER 1

INTRODUCTION

1.1. PEG Oleate Production

PEG Ester surfactants are known to be excellent emulsifiers, and synthesized via esterification of polyethylene glycol and a fatty acid. The hydrophilic part of the PEG Ester constitutes to the polyethylene glycol, and the lipophilic part is created by the fatty acids. Polyethylene glycol and fatty acids with various molecular weights can be reacted to form emulsifiers with a wide range of HLB values. For a HLB value of greater than 13, it is reported that the emulsifier is water soluble and oil dispersible, while it is oil soluble for a HLB value lower than 13[1] .

One of the common PEG Ester surfactants, PEG (mono or di) oleates are produced from the reaction of oleic acid with polyethylene glycols, such as PEG-400 or PEG-600, in the presence of an acid catalyst. The catalysts can be either homogenous or heterogeneous. However, homogenous catalysts are difficult to recover from the product, can cause corrosion on the pipeline and in the reactor, and have more energy intensive reaction conditions. On the other hand, heterogeneous catalysts are more advantageous owing to the milder reaction conditions, easy separation from the reaction effluent and possibility to reuse, and being less energy intensive [2]. However, in this research, PTSA is used as a catalyst, which is a soluble (homogenous) one, since the type of catalyst has minor effect on the main goal of the project.

The reaction is optimally carried out between 140-180°C's, with around 4% catalyst loading, under vacuum and the time varies between 1 and 24 hours. As a result, a mixture of monoester and diester products is produced. As represented on the Figure 1.1, PEG monoester is formed in the first step of the reaction. The two hydroxyl groups of monoesters can further react with carboxylic acids to form the diesters. The reaction conditions determine the ratio of mono or diesters produced [2].

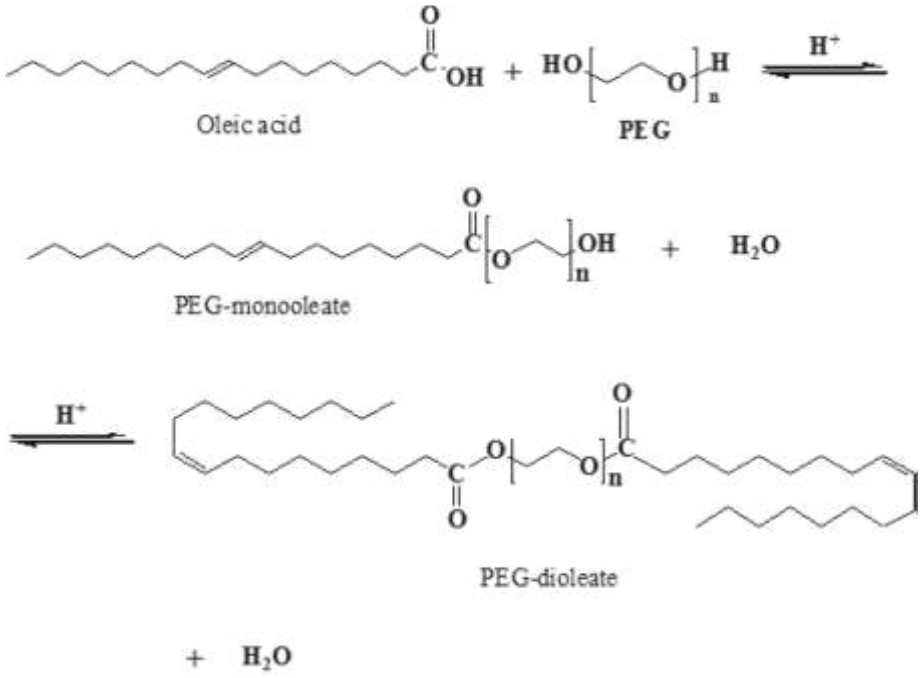

Figure 1.1. Reaction scheme [2]

The reaction scheme on the Figure 1.1 shows that the both reactions are at equilibrium and to achieve high conversions of oleic acid, excess of PEG can be used. Water can also be removed by vacum or by a current of inert gas to increase the yield of products.

1.2. PEG Oleate as an Emulsifier

PEG oleates are excellent non-ionic emulsifying agents, mainly for water/oil mixtures. They form a monolayer at the water/oil interface and stabilize the emulsion by increasing kinetic stability. They are often better emulsifiers than alcohol ethoxylates and nonyl phenol ethoxylates (NPE), and have similar solubility characteristics. Due to the environmental impacts of NPE's, their replacement by PEG Esters, which are biodegradable, has low toxicity, and low hazardous properties, is promoted. PEG esters are mainly utilized in the formulation of cosmetics or medical products, such as emulsifiers for cosmetic creams and toiletry emulsions, and emulsifiers for polymer latex production. It is also widely used in the lubricants in textile processing, or solvent cleaners and emulsifiable degreasers. As a good surfactant, PEG oleates can also be used as a wetting/dispersing agent.

1.3. Quality Control Methods

Various quality control methods are present to track PEG-Oleate production process. Cloud point, hydroxyl value, and total acid number are the most important methods. Among these methods, total acid number that is based on potentiometric or colorimetric titration is the most widely used and simple one. However, it is non-automated and there is a risk of human caused errors. Also the use of solvents during the measurement creates unhealthy and hazardous conditions. This thesis focuses on measuring the acid number in a more accurate and healthy way.

1.3.1. Cloud Point

The cloud point is the temperature, where the non-ionic surfactant or a glycol solution starts to form a cloudy or turbid appearance. At this temperature, phase separation starts to take place causing the cloudy appearance. The cloud point of a solution can be measured manually or automatically. Manual testing can be done in accordance with *ASTM D2500*, where the sample is placed in a constant temperature cooling bath, and at every 1 °C, it is taken ant inspected for the cloud. In the automatic method that is *ASTM D5773*, a continuous illumination is provided to the sample, and an array of optical detectors is used to monitor the sample until the first cloudy appearance is detected. The automatic method utilizes an automatic apparatus equivalent to the manual one, and needs shorter operation time. Also by D5773 a temperature range of –60 °C to 49 °C can be tested for cloud point with a resolution of 0.1 °C [3].

1.3.2. Hydroxyl (OH^-) Value

Hydroxyl value (HV) is measured to determine the content of free hydroxyl groups of the chemical substance, often as the amount of KOH (mg equivalent) to the OH^- content of one gram of chemical specie. In the conventional method, acetylation of the hydroxyl group of the substance takes place with acetic anhydride in pyridine solvent. When the reaction is completed, the remaining unreacted acetic anhydride is converted to acetic acid by the addition of water and measured by titration using KOH. The main

standardized methods to determine the hydroxyl value are *ASTM D1957* and *ASTM E222-10*. [4] The calculation of HV is expressed in Equation 1.1, where

V_A= KOH required for titration of the acetylated sample, mL,

V_B= KOH required for titration of the blank, mL,

N= normality of the titrant,

W= weight of sample used, g, and

AV= separately determined acid value of chemical substance.

$$HV = \frac{56.1\ N\ (V_B - V_A)}{W} + AV \tag{1.1}$$

1.3.3. Total Acid Number (TAN)

Total acid number (TAN) is used to determine the oxidation that the specie has undergone, and species having more amount of oxidation has higher TAN. It is measured by the amount of potassium hydroxide (KOH) spent in milligrams that is needed to neutralize the acids in one gram of a specie [5].

One of the standard test method of potentiometric titration is *ASTM D664*, where the titration curve is used to determine TAN. Another standard method, which is *ASTM D974,* is the color indicating titration, and it is consulted in this thesis. In this method, the sample is dissolved in ethanol or diethyl ether, and few drops of indicator (in our case phenolphthalein) is inserted. The titrant KOH with normality (N) of 0.5 is added by the help of a burette until a sharp color change in the sample to purple is observed [6]. The weight in grams of the sample (W), and the volume of 0.5N KOH consumed is recorded and the TAN is determined from the formula below:

$$TAN = \frac{A \times N \times 56.1}{W} \tag{1.2}$$

If the acid number of the product is lower than 10, the product passes the quality control test.

TAN is the most well-established and crucial quality control method for this type of emulsifiers. However, there are some drawbacks that make the pre-mentioned methods problematic in practice. First of all, it is still non-automated and prone to human error. Operator should be well experienced and very careful during the measurements to prevent any human-caused errors. In industry, these products are produced in tons every day, and in the case of an incorrect quality control, large amounts of improper product will reach costumer, which may cause costumer dissatisfaction and loss of profit. Also, the labor intensive nature of the test makes it time-consuming. Another drawback is the use of significant amount of organic solvents and corrosive reagents, which are hazardous, unhealthy and can be troublesome in disposing.

The main motivation of the thesis is to eliminate aforementioned drawbacks and find a fast, simple, and non-hazardous procedure in the analysis of acid number. Therefore, the aim of this study is to develop an automated analysis giving accurate results for TAN, by benefitting from the FTIR spectroscopy along with multivariate calibration techniques. All the rest of this thesis will focus on the experimental and modeling work of developing a routine analysis of PEG oleate emulsifiers. Several similar chemometric studies are present in the literature De Lira et al [7] used near infrared (NIR) and middle infrared (MIR) spectroscopy and multivariate calibration to monitor the stability of biodiesel. They have employed partial least squares (PLS) and multiple linear regression (MLR) models as their calibration method. Koczon, Gruczynska and Kowalski [8] have developed a chemometric method to determine the acid value of the butter, which indicates its freshness and quality, being stored at different temperatures. In their work, PLS model using FTIR spectroscopy with spectral range from 9000 to 400^{-1} cm is employed. Moreover, Carolei and Gutz [9] developed a simultaneous quality control method for three types of surfactants by FTIR spectroscopy in the middle infrared region, and compared two types of multivariate calibration methods that are classical least square (CLS) and inverse least square (ILS).

CHAPTER 2

INFRARED (IR) SPECTROSCOPY

2.1. IR Spectra and IR Region

The category of electromagnetic radiation having wavenumbers between 14000 and 20 cm^{-1} corresponds to the infrared radiation, and its application in analytical chemistry is named infrared spectroscopy. IR spectroscopy relies on the IR absorption theory and is a well-known technique for material characterization. Translational, rotational, and vibrational motion in molecules can be produced by the energy supplied from IR radiation. Each functional group has a characteristic IR radiation frequency, and this corresponds to transition results in a spectrum. This helps to identify the chemical structure of a compound. It is a simple technique, has lots of application areas, and almost any solid, liquid, or gas sample can be analyzed [10].

In an IR Spectrum, the detector response is visualized as percent transmittance (%T) (or absorbance) on the y-axis, and the IR radiation frequency as wavenumber (cm^{-}) on the x-axis. For application and instrumentation reasons, the infrared region is divided into three sub-groups, which are near (NIR), middle (MIR), and far (FIR) infrared regions, and they are given in Table 2.1.

Table 2.1. Infrared spectral regions [11]

Region	**Wavelength Range, μm**	**Wavenumber, cm^{-1}**
NIR	0.78 – 2.5	12800 – 4000
MIR	2.5 – 50	4000 – 200
FIR	50 – 1000	200 – 10

The spectra of the NIR involves overtones and combinations of the fundamental vibrational modes of MIR region. For qualitative work, this region is not recommended since all organic species absorb in the NIR and produce overlapping bonds. Yet, this holds true only for inspection with eye, and it is possible to investigate NIR region with chemometric methods. However, it is suitable for quantitative work, such as in situ monitoring of reactions. MIR is the most commonly used region, and subdivided into two

regions as group frequency, and fingerprint region. Group frequency region is in between wavenumbers 4000 and 1300 cm^{-1}, and useful to detect functional groups. The fingerprint region (1300-500 cm^{-1}) is useful to identify a sample when it is matched with a reference spectrum in the library. At last, FIR spectrum consists of low-frequency bending and torsional motions, and utilized in the identification and differentiation of isomers, many minerals, and colorants [11].

In Table 2.2, the most important types of bonds together with corresponding wavenumber and intensity in the IR region are represented.

Table 2.2. Important IR stretching frequencies [12]

Type of Bond	**Wavenumber (cm^{-1})**	**Intensity**
C≡N	2260 – 2220	Medium
C≡C	2260 – 2100	Medium to weak
C=C	1680 – 1600	Medium
C=N	1650 – 1550	Medium
C=O	1780 – 1650	Strong
C-O	1250 – 1050	Strong
C-N	1230 – 1020	Medium
O-H (alcohol)	3650 – 3200	Strong, broad
O-H (carboxylic acid)	3300 – 2500	Strong, very broad
N-H	3500 – 3300	Medium, broad
C-H	3300 – 2700	Medium

2.2. FTIR

2.2.1. Working Principle of FTIR

Rotational and vibrational motion of atoms and molecules emit or absorb light in the wavelength range of infrared radiation (IR), which is 700 nanometers (nm) to 1 millimeter (mm). Hence, it is a useful frequency range to analyse the energy states of molecules of proper symmetry. After the interferometer is invented and coupled with IR, analysis became much quicker. Interferometer makes spectral encoding and obtains signals in the time domain, and the Fourier Transformation converts the time domain into the frequency domain, where absorption is expressed as a function of frequency. In brief, Fourier-transform infrared spectroscopy (FTIR) is the analytical method to get an infrared spectrum of absorption or emission of a sample that can be solid, liquid or gas is called.

A high-spectral resolution data can be simultaneously collected over a broad spectral range by this type of spectrometer. The technique benefits from the mathematical process "Fourier Transform" in order to convert the raw data into the actual spectrum[13]. FTIR instruments involves an interferometer, a system to collect the spectrum, involving a light source, beam splitter, two mirrors, a laser, and a detector. An interferometer system is illustrated on Figure 2.1.

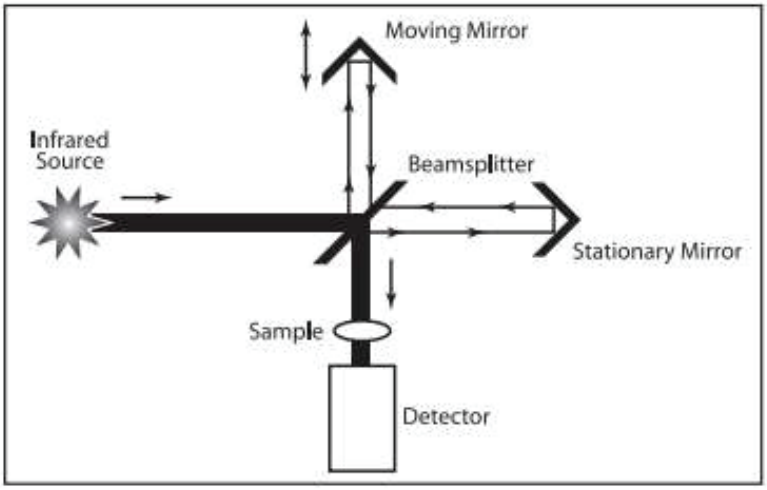

Figure 2.1 Interferometer system [14]

In the interferometer, the light beam generated from the infrared source is split into two by beamsplitter. Almost half of the beam is reflected to the stationary mirror, while the other half transmits into the moving mirror that moves back and forth at a fixed velocity. The beams are reflected by the mirrors to the beamsplitter and combined again. The light from the moving mirror travels a different distance. Recombination of the beams is either constructive or deconstructive, and this creates an interference pattern, namely an interferogram. The interferogram then meets the sample, while some of the light is absorbed by the sample and some are transmitted to the detector. Here, information about every wavelength in the infrared range is read simultaneously. The detector signal is delivered to the computer, where Fourier Transform is applied to convert the interferogram into single beam spectrum [14].

2.2.2. Advantages of FTIR over Dispersive Infrared Instruments

Dispersive or scanning monochromator method is an alternative method for obtaining spectra, but less practical for infrared, and more commonly used in UV-Vis spectroscopy. FTIR has various performance advantages over dispersive infrared instruments. Firstly, for the measurement of infrared spectrum, the interferometer of a dispersive infrared instrument separates the energy into individual frequencies. However, this is not the case in FTIR. The so called "multiplex advantage" leads to the simultaneous collection of information at all frequencies, so less time is needed for analysis, and better signal to noise ratio is obtained. The second advantage is called the "throughput advantage", referring to the fact that more energy is reached to the sample and the detector of the FTIR detector than in a dispersive instrument. The reason is FTIR does not include a slit and has fewer mirrors. Higher signal-to-noise ratio achieved in the FTIR then leads to a higher quality spectrum, where even the small peaks are very clearly represented. Finally, a FTIR instrument has a "precision advantage", which is due to the presence of a laser controlling the velocity of moving mirror. The laser also provides wavelength calibration within the instrument, and no external calibration method is required as in the dispersive instrument. This leads to a high precision spectrum in a FTIR and more reproducible data, meaning a measurement done today can be easily compared with a measurement done five years apart from today. In brief, FTIR is a more modern and highly preferred method than dispersive infrared instrument, due to the all aforementioned advantages [14].

2.2.3. Attenuated Total Reflectance (ATR)-FTIR

The most reliable and recognized method to characterize, identify, and quantify the substances is the mid-infrared (IR) spectroscopy. Wide range of solids, liquids, and gases can be analyzed with IR spectroscopy. Nevertheless, difficulty of the method is associated with the sample preparation. It is often time consuming to prepare the sample due to the melting, squishing or diluting processes that are required so that appropriate amount of light can be transmitted through the sample. Moreover, the spectral features are related to the sample thickness. If a sample has a thickness less than 1 micron, the

absorption will be small thus it is undetectable. In contrast, samples thicker than 20 microns will absorb too much IR, and a proper spectrum cannot be obtained.

Recently, ATR has become a more favorable analysis technique since it overcomes the main bottlenecks of the IR analysis, being sample preparation and sample reproducibility. This system also has high signal to noise ratio, thus giving reliable results even in nanogram quantities. The ATR-FTIR spectroscopy has a crystal accessory made of a material with high refractive index and mirrors focusing the IR radiation on the crystal face. Various crystal types can be used such as Diamond, Thallium Iodide/Bromide, Germanium, and Zinc Selenide. The sample is brought into contact with the crystal material. Since the material is IR transparent, IR beam undergoes multiple internal reflections and creates evanescent wave projecting orthogonally into the sample. Part of the evanescent wave is absorbed by the sample, and the other part is reflected and goes to the detector. This is schematized in Figure 2.2. [15,16].

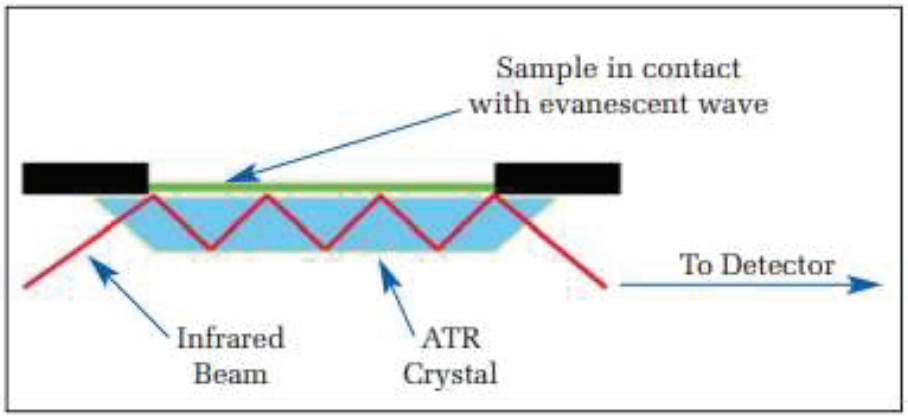

Figure 2.2. A multiple reflection ATR system [16]

Briefly, sample to sample reproducibility is enhanced, user dependence of the analysis is minimized, and the sampling became much faster on an ATR-FTIR system. The improved spectral accuracy and reproducibility, thus, lead to excellent quality data, and ATR became a reliable and widely used method for quantitative studies of various samples.

2.2.4. Quantitative Analysis by FTIR

In the last few decades, quantitative analysis (QA) of materials by FTIR has evolved from matching peaks at few wavenumbers to matching the entire spectra.

Thereby, FTIR QA is incorporated with the chemometrics to analyze many samples quickly and accurately. The quantitative analysis of an absorption spectroscopy starts with Beer-Lambert Law, and for a single homogenous compound, the absorbance (A) at a given frequency is expressed as:

$$A = \varepsilon bc \tag{2.1}$$

Where, ε is the molar absorptivity at the frequency, b is the path length of the sample, and c is the concentration of compound in solution. The law essentially represents that absorption intensity of a compound is linearly proportional to its concentration in the homogenous mixture.

At frequency n, the total absorbance of a mixture containing m number of species is given by the symbol A_n, and expressed as below:

$$A_n = \sum_1^m A_j = \varepsilon_1 bc_1 + \varepsilon_2 bc_2 + \ldots + \varepsilon_m bc_m \tag{2.2}$$

Where, j represents the component indices at frequency n (j=1,2,…,m) [17].

CHAPTER 3

MULTIVARIATE CALIBRATION

3.1. Overview

Chemometrics is a beneficial method for extracting chemically relevant information, either qualitative or quantitative, from the data produced by a chemical analysis or an experiment. Statistics, numerical analysis, operation analysis, and applied mathematics are often used in this method to structure the chemical problem into a mathematical expression [18].

The spectrometric data can be evaluated by calibration method, where the known concentration of specie is related to its spectral information, such as absorbance. The mathematical model constructed by calibration is then used for the prediction of unknown sample's concentration by using its spectral information [19]. When only single wavelength is used to determine the concentration of a single compound, it is called univariate calibration. However, in multivariate calibration, several wavelengths or full spectrum can be used to detect the concentration of a multi or single component mixture.

3.2. Univariate Calibration

In univariate calibration, concentration of a sample is determined by using the response of a single detector (i.e. chromatographic peak area) or a single spectroscopic wavelength. For the quantitative analysis, Beer-Lambert law is used for the model construction, where the absorbance at a wavelength is directly proportional to the absorptivity coefficient, light path length and concentration, as in described in Equation 2.1. In the case of a linear model, there are two options, namely, classical univariate calibration, and inverse univariate calibration.

3.2.1. Classical Univariate Calibration

In this type, the concentration of a compound is related to the absorbance value at one wavelength with the given model:

$$\mathbf{x} = \boldsymbol{c}.s + \boldsymbol{e} \tag{3.1}$$

where, x is the vector of absorbance values at one wavelength for *l* number of samples, $\boldsymbol{c}$ is the corresponding concentrations, and s; which is actually the multiplication of molar absorptivity and path length in Beer's law, is a scalar relating those vectors and determined by the experiments. Both the vectors $\boldsymbol{x}$ and $\boldsymbol{c}$ has length *l*.

The following mathematical operation is done to determine the scalar s:

$$\boldsymbol{c}'.\boldsymbol{x} = \boldsymbol{c}'.\boldsymbol{c}.s \tag{3.2}$$

$$(\boldsymbol{c}'.\boldsymbol{c})^{-1}.\boldsymbol{c}'.\boldsymbol{x} = (\boldsymbol{c}'.\boldsymbol{c})^{-1}.(\boldsymbol{c}'.\boldsymbol{c}).s \tag{3.3}$$

$$s = (\boldsymbol{c}'.\boldsymbol{c})^{-1}.\boldsymbol{c}'.\boldsymbol{x} = \frac{\sum_{i=1}^{l} x_i c_i}{\sum_{i=1}^{l} c_i^2} \tag{3.4}$$

Once s is determined, unknown concentration can be calculated as below, where the hat symbol represents the predicted values.

$$\hat{\boldsymbol{c}} = \frac{\hat{x}}{s} \tag{3.5}$$

The quality of a prediction is proportional to the error, which is the difference between the actual and predicted value, i.e. $\boldsymbol{e} = \boldsymbol{c} - \hat{\boldsymbol{c}}$ [20]. The error term that is also included in the Equation 3.1 indicates the error is due to the spectrum in classical univariate calibration.

3.2.2. Inverse Univariate Calibration

The error can be due to the predicted concentration obtained from instrument response, or error distribution due to the instrumental performance. The classical calibration assumes the error on predicted concentration is associated with the instrumental response. However, the instruments are becoming more sensitive and reproducible every day and the instrumental errors are being minimized. On the other

hand, there is almost no enhancement on the quality of basic lab equipment such as volumetric flasks, syringes, etc. This makes the process of concentration measurement involving human interaction (i.e diluting, weighing) a great source of error as well. Hence, inverse univariate calibration is developed to minimize the errors due to concentration. The classical and inverse calibration is well schematized in Figure 3.1.

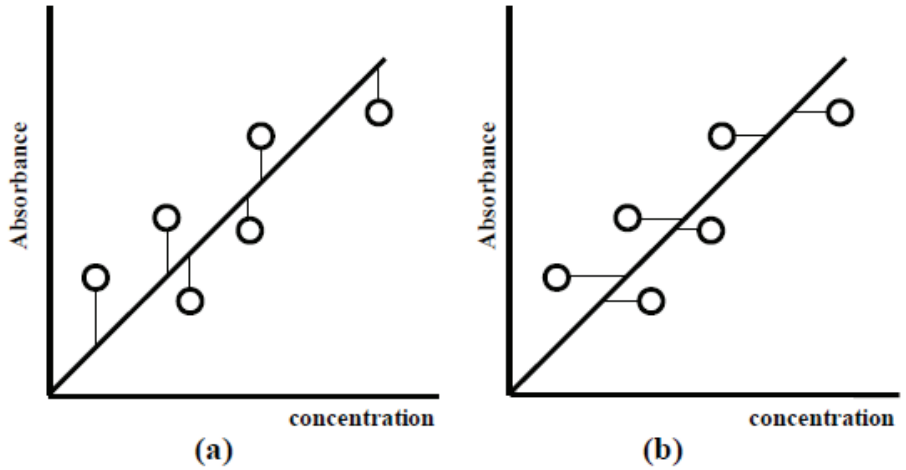

Figure 3.1. Errors in (a) Classical and (b) Inverse calibration [20]

The inverse calibration can be performed by the given model:

$$\boldsymbol{c} = \boldsymbol{x}.b \tag{3.6}$$

The different assumptions on the error distribution make the scalar b only approximately inverse of s. It is also evident on the equation 3.7 that is slightly different than the equation 3.4.

$$b = (\boldsymbol{x}'.\boldsymbol{x})^{-1}.\boldsymbol{x}'.\boldsymbol{c} = \frac{\sum_{i=1}^{l} x_i c_i}{\sum_{i=1}^{l} x_i^2} \tag{3.7}$$

For a good set of data, the predictions of both the models should be in fair agreement. If this is not the case, there might be other factors such as non-linearities, outliners, and unexpected noise distributions [20].

3.3. Multivariate Calibration

The necessity of finding molar absorptivity for each pure component separately makes the univariate calibration a cumbersome method. Multivariate calibration enables the analysis of multiple components in a sample simultaneously. It is a more selective and reliable tool, has fault detection ability, and can also analyze non-homogenous or contaminated analyte [19].

The equations developed with this type relate concentration to absorbance values, vice versa. In comparison to univariate calibration, absorbance and concentration values are no more represented as vectors, but as matrices. This is because multiple responses are related to the properties of a sample, and concentration values can be given for the sample having more than one component [21].

In the following, different types of multivariate calibration are given from simple to more complex.

3.3.1. Classical Least Squares (CLS)

CLS is an extension of Beer Lambert Law, in which the absorbance values are expressed as a function of concentration. Similar to classical univariate, errors are assumed to be based on instrument responses. **X** CLS model for *m* calibration samples, *l* chemical compounds, and *n* wavelengths, is expressed in matrix notation as:

$$\mathbf{X} = \mathbf{CK} + \mathbf{E_X} \tag{3.8}$$

Here, $\mathbf{X}$ represents a $m \times n$ matrix of calibration spectra, $\mathbf{C}$ $m \times l$ matrix of component concentrations, $\mathbf{K}$ $l \times n$ matrix of absorptivity pathlength constants, and $\mathbf{E_X}$ $m \times n$ matrix of spectral errors or residuals that are not fit by the model.

The estimation of **K** matrix is done by CLS with the following equation:

$$\widehat{\mathbf{K}} = (\mathbf{C}'.\mathbf{C})^{-1}.\mathbf{C}'.\mathbf{X} \tag{3.9}$$

Once the estimated values of absorptivity path length constants are known, the concentrations of unknown samples can be predicted from their spectrum by using equation 3.10. Here, **x** represents the spectrum of unknown sample.

$$\hat{\mathbf{c}} = (\hat{\mathbf{K}}.\hat{\mathbf{K}}')^{-1}.\hat{\mathbf{K}}.\mathbf{x} \quad (3.10)$$

The disadvantage of CLS is that the concentrations of interfering species must be known apriori and included in the model [22].

3.3.2. Inverse Least Squares (ILS)

ILS is a reliable method especially when the pure spectra of all the species in the unknown samples cannot be obtained. The inverse of Beer Lambert Law applies, thus concentration is expressed as proportional to the absorbance values, as in the equation 3.11. The model errors of ILS rely on the errors in the measurements of component concentrations.

$$\mathbf{C} = \mathbf{XP} + \mathbf{E_c} \quad (3.11)$$

where, **C** is the $m \times l$ concentration matrix, **X** $m \times n$ absorbance matrix, $\mathbf{E_c}$ $m \times l$ error matrix of concentrations that are not fit by the model, and **P** is the calibration coefficients matrix with the size $n \times l$, relating component concentrations to the spectral intensities. If the elements in the $\mathbf{E_c}$ are assumed to be independent, identical analysis for each individual analyte can be done from the equation 3.12, where a single component is modeled at a time.

$$\mathbf{c} = \mathbf{Xp} + \mathbf{e_c} \quad (3.12)$$

Here **p** is a *nx1*, $\mathbf{e_c}$ is a *mx1*, and thus **c** is a *mx1* matrix. When making calibration, the least square solution of **p** in equation 3.12 yields:

$$\hat{\mathbf{p}} = (\mathbf{X^t}.\mathbf{X})^{-1}.\mathbf{X^t} \times \mathbf{c} \quad (3.13)$$

Finally, the concentration of the component in the unknown sample can be predicted as follows:

$$\hat{c} = \mathrm{x}'.\hat{\mathrm{p}} \quad (3.14)$$

where $\hat{c}$ is the scalar estimated concentration of the analyzed component, and **x** is the spectrum of unknown sample.[23] The strength of this method is that one does not need to know all the components in the sample. Also, one can select as many variables, i.e. wavelengths, instead of using the full spectra. The main drawbacks of ILS is not being

able to detect all the outliers, and not being very effective in selecting the optimal wavelength for predicted models. Moreover, adding more wavelengths to the model can lead to overfitting [19].

3.3.3. Partial Least Squares (PLS)

PLS is considered as the major regression technique for multivariate data. It has two algorithmic approaches for complex chemical mixtures, which are PLS1 and PLS2. PLS1 models one component at a time for the building step of model, and hence the predictions are more accurate. PLS2 is more recommended for qualitative analysis or when the components to be modelled are expected to be correlated. The algorithm of PLS1 will be presented here.

In PLS1 algorithm, each component can be modeled one at a time, and only the concentration of the component of interest is involved in the calibration, and the other concentrations are not needed. It is assumed that errors can be based due to spectrum or concentrations. The most significant distinction of PLS from ILS is being less prone to overfitting, and being more robust models.

In PLS1, 2 model sets are constructed as follows:

$$\mathbf{X} = \mathbf{TP} + \mathbf{E} \tag{3.15}$$

$$c = \mathbf{Tq} + \mathbf{f} \tag{3.16}$$

The absorbance and concentration matrices are represented as $\mathbf{X}$ and c, respectively. A crucial feature of PLS is that scores matrix, $\mathbf{T}$, is common both for concentration and measurement. Here, P and q are analogous to loadings vector, and the multiplications $\mathbf{TP}$ and $\mathbf{Tq}$ are used to approximate spectral data and true concentration, respectively. The matrix operation is schematized in Figure 3.2. The sum of the squares of the scores of each component is called eigenvalue. A more significant component has a greater eigenvalue.

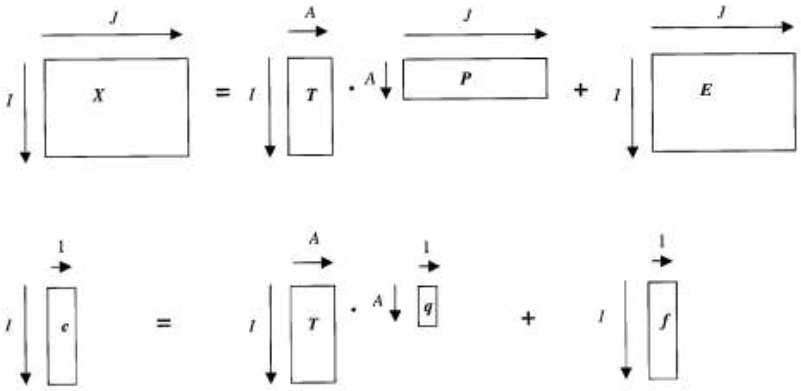

Figure 3.2. A schematized matrix view of PLS1 algorithm

3.3.4. Genetic Inverse Least Squares (GILS)

The selection of wavelength and construction of a multivariate calibration data set is made possible by GILS. It is a method in which the ILS is combined with Genetic Algorithm (GA). Genetic algorithm is a global search and optimization technique that is based on Darwin's theory of evolution and natural selection [24]. According to Darwin's theory, the species that do better adapt to the environment has a better chance of survival and breeding, and the ones who don't are not likely to survive. As this process continues, over a long period of time, through generations, the subsequent offspring will appear as a better fit to the environment and has a more chance to survive than the previous generations. The simulation of genetic algorithms in scientific researches started in 1960's by biologists doing genetic systems experiments on a computer. The father of the field is Holland, who developed a GA in his research on adaptive systems in early 1960s [25]. Until today, many global optimization problems are solved by making use of GA tools, and many applications exist in calibration, more specifically on wavelength selection [22, 26, 27].

The genetic algorithm has 5 main steps: initialization, evaluation, breeding and mating, crossover and mutation, and replacing the parents with their offspring. The algorithm is schematized in the Figure 3.3.

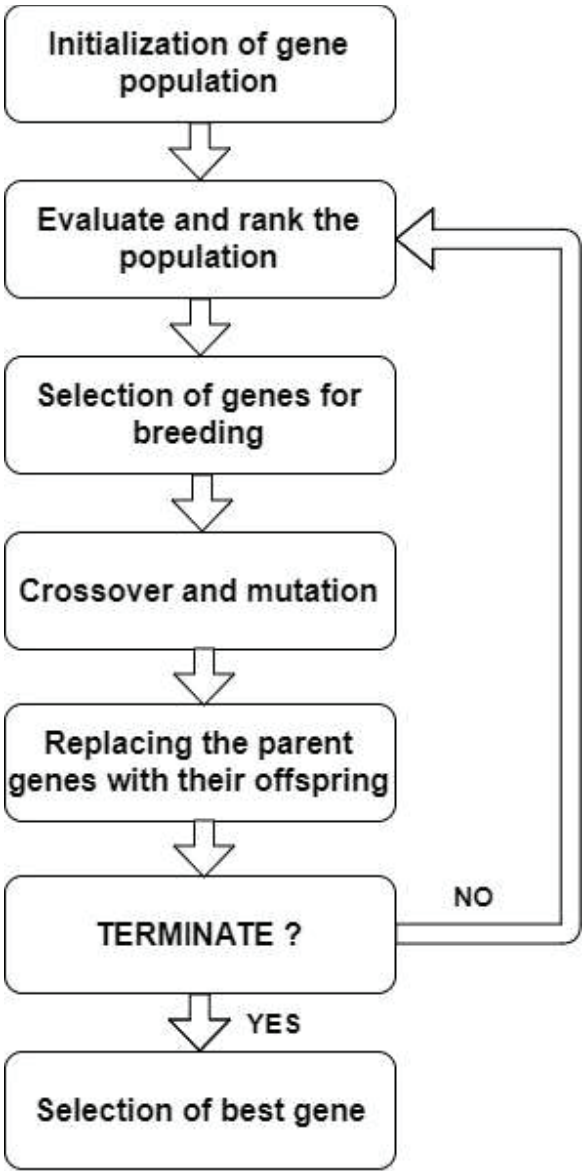

Figure 3.3. Flow diagram of GILS method.

Initialization

A gene is formed by gathering the randomly selected instrumental responses. It can be represented by the following expression, where S symbolizes the gene and A the absorbance measured at the indicated wavelength in the subscript:

$$S = [A_{2678}A_{256}A_{1478}A_{560}]$$

A population is the collection of individual genes. In the initialization step, the first generation of genes is randomly generated with a fixed population size. The random selection of responses enables the minimization of bias and maximization of number of

recombination. The population size is an important matter since it determines the time to complete an individual run of the algorithm, i.e. a larger size needs more time. The number of genes in the population needs to be an even number to allow breeding of the genes.

Moreover, there is a constraint in choosing the number of wavelength points in a gene, that is, it must be obtained randomly between a specified high and low limit. The higher limit is chosen to prevent overfitting problems and to reduce the computation time, and lower limit is set to 2 in order to allow single-point crossover.

Evaluation of the Population

In the second step, evaluation and ranking of the genes is done using a fitness function, which is defined as the inverse of standard error of calibration (SECV) with cross validation, and shown in equation 3.17.

$$Fitness = \frac{1}{SECV} \tag{3.17}$$

SEC indicates the success of each gene, and calculated from equation 3.18.

$$SECV = \sqrt{\frac{\sum_{i=1}^{m}(c_i - \hat{c}_i)^2}{m-2}} \tag{3.18}$$

Here, c_i is the reference concentration, $\hat{c}_i$ is the predicted concentration, $m-2$ is the degrees of freedom, while m is the number of sample and 2 indicates the extracted parameters, which are the intercept and the slope between the reference and the predicted concentrations.

The fitness values of the evaluation genes are measured in each evaluation step, and iteratively the fitness is enhanced [22].

Selection of Genes for Breeding

This step relies on the natural evolution principle and the members with the highest fitness values are selected, who are to be replaced with parent genes. The highest fitness means a better suited gene that can survive and transfer information to the next generation.

There are several methods for parent selection. One of them is the top down method, in which the genes are ranked in the pool, and they mate consequently, i.e. first

gene (S1) mate with second gene (S2), third (S3) with fourth (S4) and so on. This process gives all the genes a chance to breed. Another method is called roulette wheel selection, and it is illustrated in Figure 3.4.

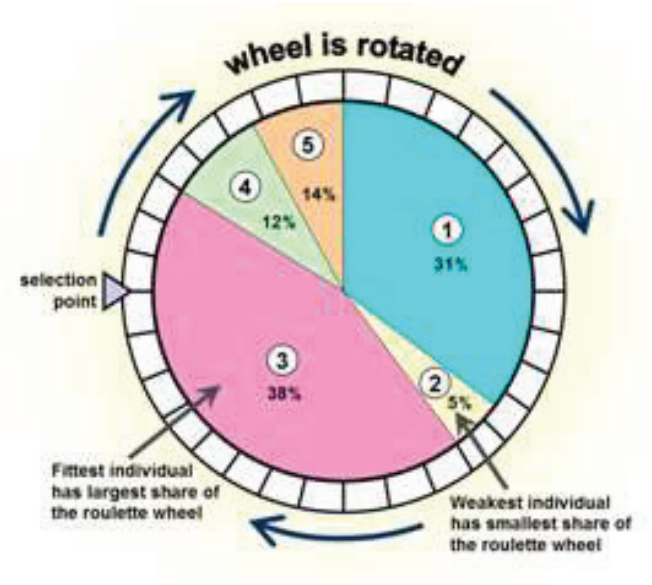

Figure 3.4. Roulette wheel selection [28]

Each part of the wheel represents a gene, and each area in the wheel is proportional to the fitness of a gene. A gene occupying a higher area in the wheel has more chance of being selected. In this method, some genes can be selected multiple times, while others might not be even selected and thrown out of the pool. After the selection, parents that are selected are mated top-down. Since there is no ranking, the number of possible recombination increases.

Crossover and Mutation

This step involves the process of breaking of genes at random points and their offspring by cross-coupling to create new genes. Most of the work of GA is on this step. In the following, the process is illustrated where the parents are S1 and S2 genes, and their offspring are S3 and S4.

Parents:

$S1 = [A_{456}A_{3745}A_{8212} \vdots A_{8123}A_{344}]$

$S2 = [A_{125}A_{7786} \vdots A_{569}A_{4064}]$

Offspring:

$S3 = [A_{456}A_{3745}A_{8212}A_{569}A_{4064}]$

$S4 = [A_{125}A_{7786}A_{8123}A_{344}]$

The parent genes are randomly cut from the parts indicated as ⋮ where separation and crossover of the genes takes place. This is an example of single point crossover, which is the most common type for GILS.

Also, sometimes mutation can take place in this step by introducing random deviations into the population. One can realize this into the algorithm during mating at a rate of 1%, which is a typical rate. Generally, one of the wavelengths is replaced in an existing gene with another wavelength. Though, the GILS studies today usually don't apply mutation.

Replacing the Parent Genes by Their Off-Springs

After the cross-over, the parent genes S1 and S2 are replaced by their offspring S3 and S4, which are evaluated and ranked. Now, the selection for breeding starts all over again with the new genes, and repeated until the predefined iterations or the minimum tolerance value is reached.

Finally, the gene with the lowest SEC is selected for model building, which will then be used for the prediction of concentration of the sample being analyzed in the validation set. The success of the model in the prediction of validation set is determined by the standard error of prediction (SEP), as given by equation 3.19. Here, m represents the number of validation samples.

$$SEP = \sqrt{\frac{\sum_{i=1}^{m}(c_i - \hat{c}_i)^2}{m}} \tag{3.19}$$

Once the predefined iteration number is reached, termination takes place. It can also be optimized by extensive statistical tests. Often, the decision of best run is given when the lowest SEC for the calibration is quite close to SEP value.

The GILS approach has various advantages over univariate and other multivariate calibration methods. It doesn't require any complex mathematical operation in the construction of model and during the prediction process. Also, GILS uses data set with

reduced number of columns thus the algorithm operates in a memory efficient manner [25]. In this study, GILS is used considering the advantages and accuracy of the method.

CHAPTER 4

EXPERIMENTATION AND INSTRUMENTATION

4.1. Experimental Procedure

This PEG-Oleate esterification reaction was carried out using Polyethylene glycol with molecular weight 400 g/mole (PEG 400) , Oleic acid and acid catalyst which is Para toluene sulfonic acid (PTSA). Precent formulation of initial reactants composition of PEG-Oleate production is shown in Table 4.1.

Table 4.1 Precent formulation of initial reactants composition of PEG-Oleate production

Component	(w/w %)
Oleic Acid	58
PEG 400	41.6
PTSA	0.4

The production of PEG-Oleate was done under vacuum at 160 °C, and the total reaction time was around 4 hours. Pressures of system was maintained at -0.4 bar (0.6 bar absolute) with portable vacuum pump for removing water from system and preventing oxidation of PEG-Oleate product. There were three stages in the process. The first stage is heating in which reactants are mixed and the temperature of the mixture is raised from room teemperature to 160 ^{o}C in a 24 minute period. This is followed by reaction period where another 186 minutes is waited at constant temperature (160 ^{o}C) and the system is allowed cool down to 40 ^{o}C in remaining 40 minutees. In order to monitor the process, samples were taken in every 10 minutes for referance total acid number and FTIR analysis resulting in a total of 26 samples from reactor as shown in Table 4.2. As it is impossible to take sample under vacuum conditions, the vacum generator was stopped without interrupting the stiring process for sample collection. Immediately after the sampling, total acid number of these samples was determined and FTIR-ATR analysis was conducted, simultaneously.

As can be seen from Table 4.2, heating period took place about 24 minutes, and the first 4 samples represent that stage. A linear temeperature program with 5.35 °C/min

average temperature raise was observed in heating process. Heater was turned off at 157 °C and reaction stage was contuniued about 3 hours which was the longest part of the process. During the cooling stage, the last 3 samples were taken in around 25 minutes where rate of cooling was 3 °C/min in the process.

The total acid number of the samples was determined at triplicate by the *ASTM D974* standart method and the average TAN values were used in further calibration modelling studies. Similarly, triplica FTIR-ATR spectra of the samples were simultinasly collected against air background by using Perkin Elmer Spectrum 100 FTIR Spectrometer (PerkinElmer Inc., MA, USA), which is equipped with diamond crystal ATR accessories and single reflection ATR plate.

Table 4.2. PEG-Oleate esterification reaction monitoring table

Sample No	Time (min)	Temperature (ºC)	Total Acid Number (mg KOH/ g)	Standart Deviation of Total Acid Number (mg KOH/ g)	Stages Of Esterification Process
S1	0	28.7	157.80	1.15	Heating Process
S2	8	77.0	114.70	0.46	
S3	15	107.0	105.20	0.96	
S4	24	157.0	79.40	0.85	
S5	32	160.0	41.70	1.25	Reaction Stage
S6	44	160.5	38.80	1.25	
S7	52	159.9	38.50	0.53	
S8	60	160.3	33.20	0.56	
S9	70	160.2	30.10	0.26	
S10	80	160.1	27.30	0.50	
S11	90	160.0	17.30	0.44	
S12	100	159.9	15.80	0.66	
S13	110	159.7	13.10	0.26	
S14	120	159.9	12.40	0.46	
S15	130	157.7	11.10	0.44	
S16	140	159.8	9.62	0.68	
S17	150	160.3	9.45	0.42	
S18	160	159.9	8.40	0.14	
S19	170	160.2	7.39	0.14	
S20	180	160.3	6.91	0.12	
S21	190	159.8	6.72	0.12	
S22	200	159.9	5.81	0.31	
S23	210	160.0	5.74	0.26	
S24	225	100.0	5.72	0.04	Cooling Process
S25	240	70.0	4.46	0.05	
S26	250	40.0	4.41	0.05	

Among the 26 samples, 21 samples were assigned to calibration set, and the remaining 5 selected for validation set in order to buid calibration models not only for multivaretive calibration methods which are Genetic Inverse least squares (GILS) and Partial Least Squares (PLS) but also simple univariate calibration.

4.2. Instrumentation

The infrared spectra were collected with a Perkin Elmer Spectrum 100 FT-IR Spectrometer (PerkinElmer Inc., MA, USA), which is equipped with diamond crystal ATR accessories and single reflection ATR plate. The spectral data were recorded on absorbance mode aginst air background with wavenumber having a range from 4000 to 650 cm^{-1}. Delivering of the samples onto the dimand crystal was done by Pasteur pipettes allowing enough time to cool samples to room temperature before collecting the spectrum. The spectral resolution was fixed to 16 cm^{-1}, and 4 scans were done.

4.3. Data Analysis

The collected spectra were transferred in ASCII file format to another computer where the data processing has been carried out. Then, data files for for calibration and independent validation sets were prepared for both univariate and multivariate models. Genetic algorithm based calibration method GILS and PLS algorithms were written in MATLAB programming language Version 17.0 (MathWorks Inc., Natick, MA,USA).

CHAPTER 5

RESULTS AND DISCUSSION

Three different calibration methods were applied, which are univariate calibration, Partial Least Squares (PLS) and Genetic Inverse Least Squares (GILS). The reference TAN analysis of samples was carried out by *ASTM D974* standart method which is based on tiration of the sample with 0.5 N KOH in the presence of phenolphthalein indicator. While the titration of samples had been carried out, the FTIR-ATR spectra of each sample was recorded immediately in order to avoid changes on the sepectral featues associated to the reaction.

5.1. FTIR-ATR Results

FTIR-ATR spectrum of Oleic acid, Polyetylene Glycol (PEG) and PEG-Oleate product are given in Figures 5.1–5.3, respectively. The spectra of all process samples are shown in Figure 5.4.

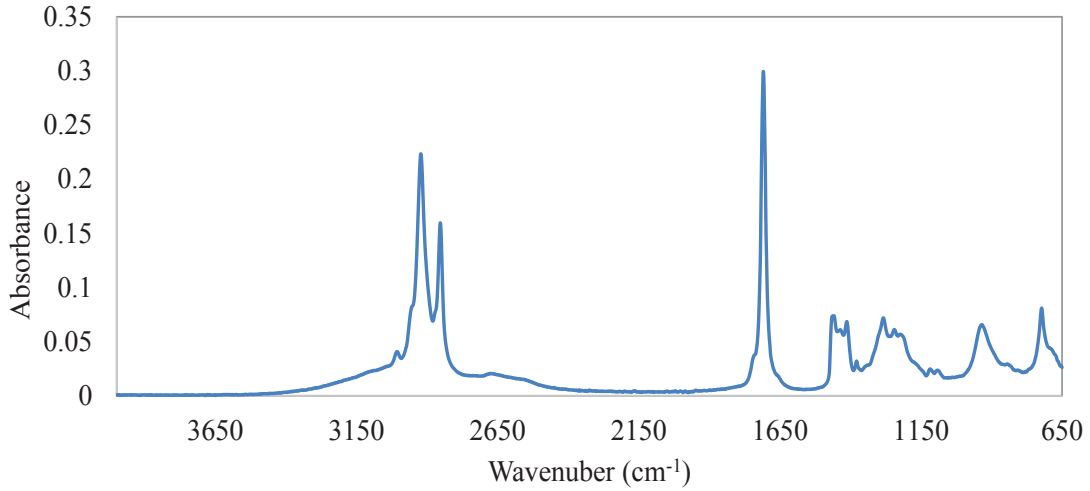

Figure 5.1. FTIR-ATR spectrum of Oleic acid

In oleic acid FTIR-ATR spectrum, C-H stretching peak was found in range of 2920-2850 cm^{-1} while C=O streching was found at 1710 cm^{-1}. Additionaly, peaks with

weak absorbances were also found for O-H bending at 1415 cm^{-1} and C-O streching at 1285 cm^{-1}.

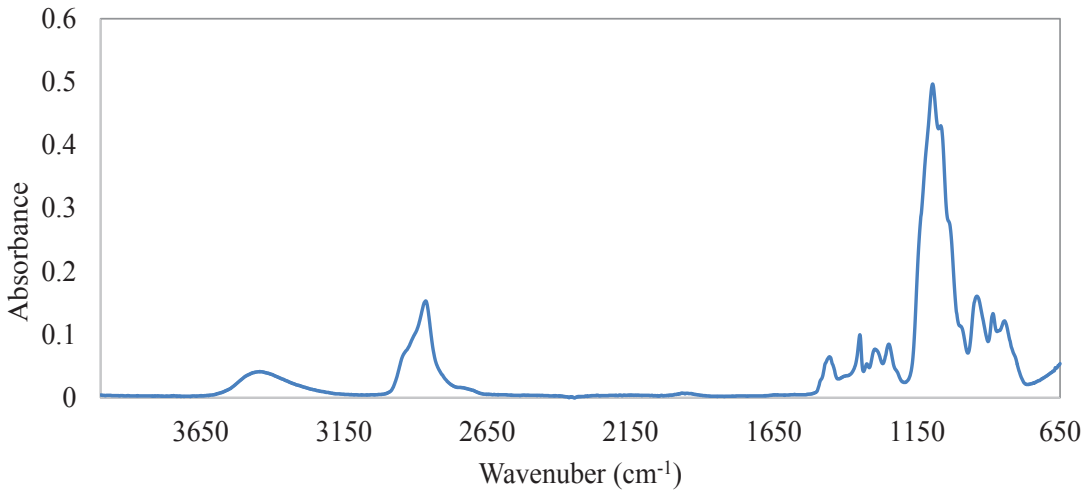

Figure 5.2. FTIR-ATR spectrum of Polyethylene glycol (PEG)

In polyethylene glycol (PEG) FTIR-ATR spectrum, C-H stretching band was found at 2880 cm^{-1} while C-O-C was found in range of 1100 - 1060 cm^{-1}. Additionaly, peaks with weak absorbances were also found for O-H bending at 3470 cm^{-1} and C-H scissor and bending at 1440 - 1265 cm^{-1}.

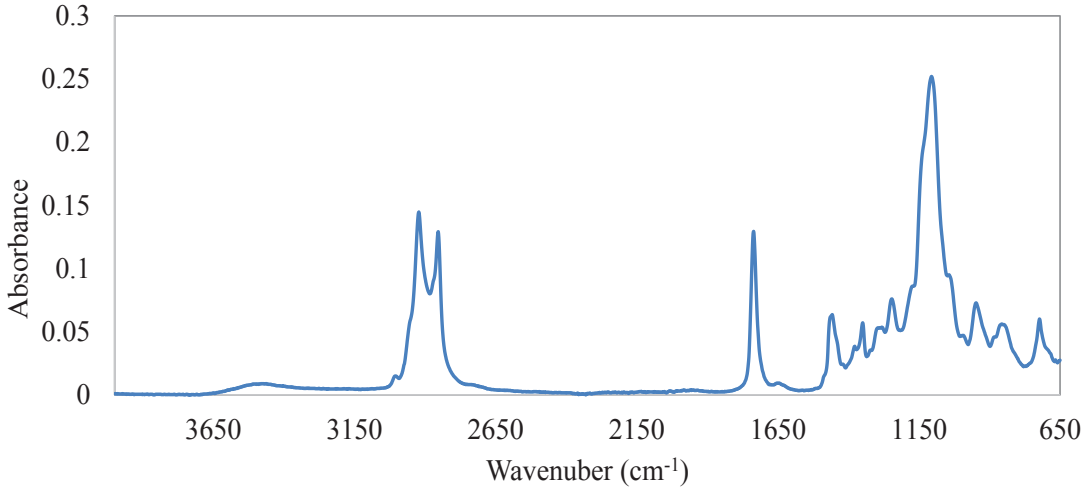

Figure 5.3. FTIR-ATR spectrum of PEG-Oleate

In PEG-Oleate FTIR-ATR spectrum, C-H stretching band was found in range of 2920 - 2850 cm^{-1} same as oleic acid. Also, C=O streching was found at 1739 cm^{-1} while

C-O-C was found in range of 1100 – 950 cm^{-1} similir as polyethylene glycol. Additionaly, a waek peak with lower absorbance was also found for C-O streching at 1111 cm^{-1}.

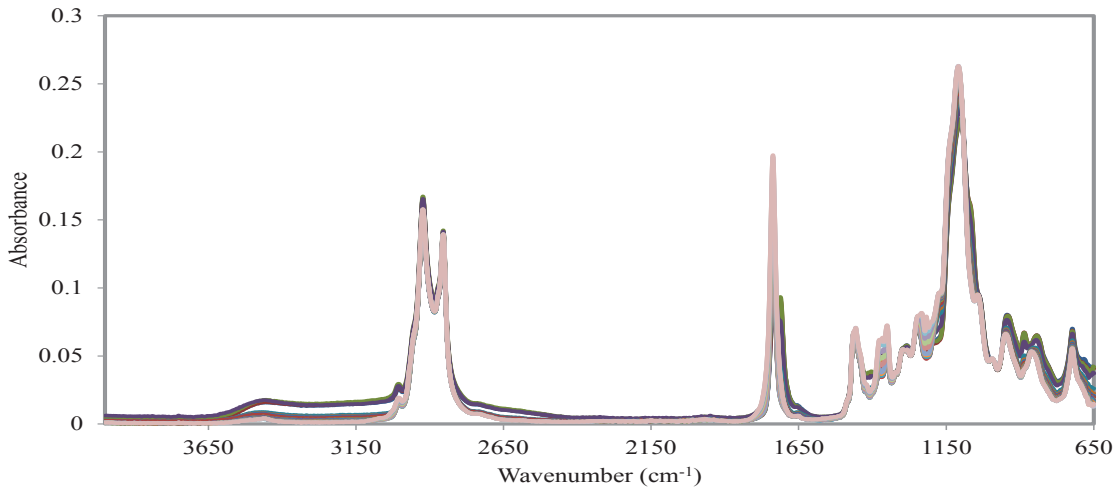

Figure 5.4. FTIR-ATR spectrum of all samples

Both carboxylic acids and esters are contain single and double bond carbon and oxygen atoms. Therefore, the carbon-oxygen double bond (C=O) is found in the range of 1735 - 1750cm^{-1} for esters and 1700 -1725 cm^{-1} for carboxylic acids. The carbon-oxygen single bond (C-O) is found in the range of 1000 - 1300 cm^{-1} for esters and 1210-1320 cm^{-1} for carboxylic acids. The spectral differences between oleic acid and oleate product are illustrated in Figure 5.5.

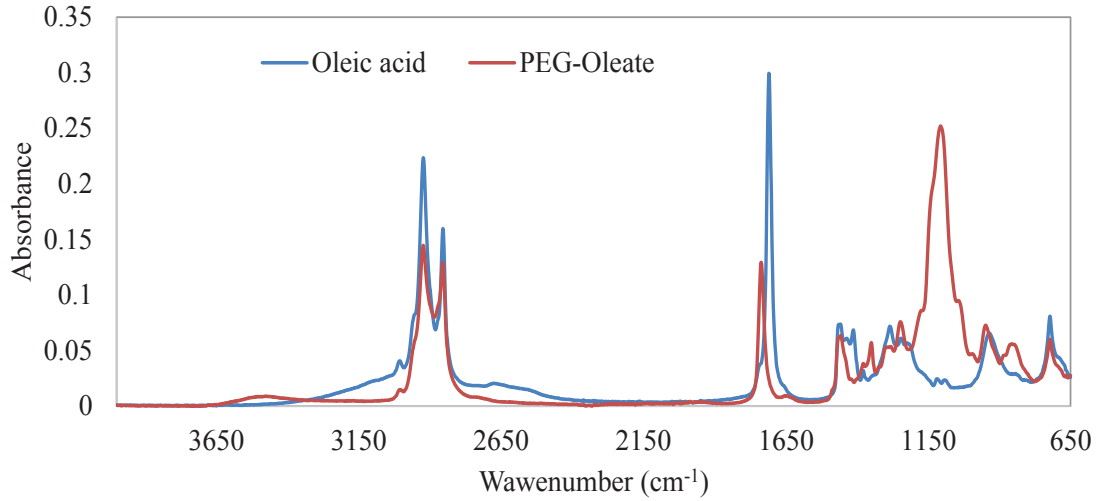

Figure 5.5. FTIR-ATR spectra of oleic acid and PEG-Oleate

Oleic acid carbon-oxygen double bond (C=O) peak is found at 1710cm^{-1} whereas Peg-Oleate double bond peak is found at 1739 cm^{-1}. As seen from the figure, the Carbon-oxygen double bond peak of the oleate is observed to be increasing during the esterfication reaction at 1739cm^{-1} while the carbon-oxygen double bond of oleic acid peak at 1710 cm^{-1} is decreasing. Eventually, the peak at 1739 cm^{-1} reaches its maximum while the peak at 1710 cm^{-1} totally disappears at the end of the reaction. Figure 5.6 shows enlarged wiew of the rangion from 1800-1650 cm^{-1} where C=O ester peak raising up while C=O acid peak start to decreasing.

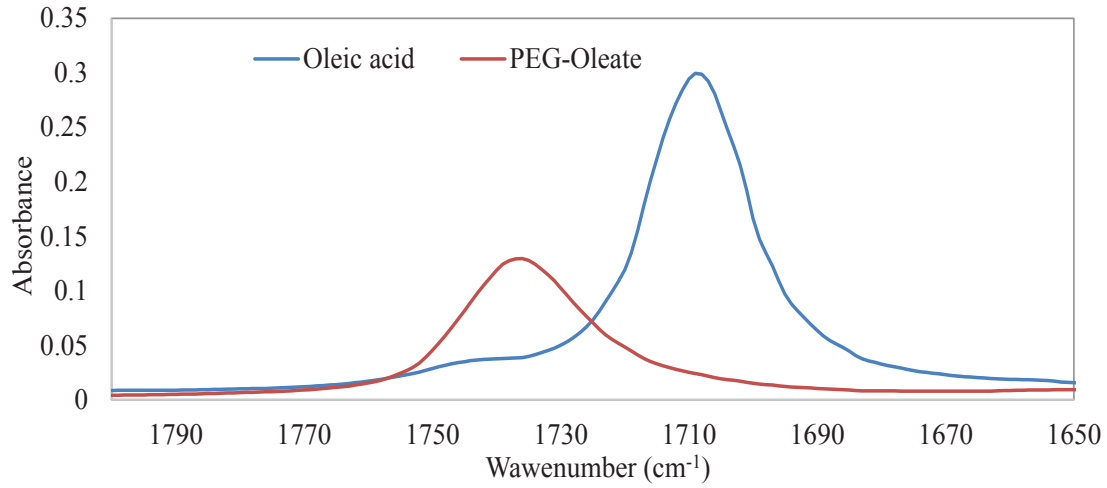

Figure 5.6. FTIR-ATR spectra of oleic acid and PEG-Oleate in the range of 1800 -1650 cm $^{-1}$

5.2. Univeriate Calibration

Regarding the changes of peak absorbances at 1739 cm^{-1}, univariate calibration is applied to this C=O streching peak for PEG-Oleate product. Reference total acid number of samples and absorbance values at 1739 cm^{-1} are given in Table 5.1 and 5.2 for calibration and independent validation sets. Enlarged FTIR-ATR spectra of samples are given in Figure 5.7 where the wavenumber range from 1800-1650 cm^{-1} is displayed.

As seen from Figure 5.7 , FTIR-ATR spectra of samples shift the C=O stretching peak to a higher wavenumber, which not only gradually change the vibrational frequency of this band but also cause deviation from linearity at the peak maxima.

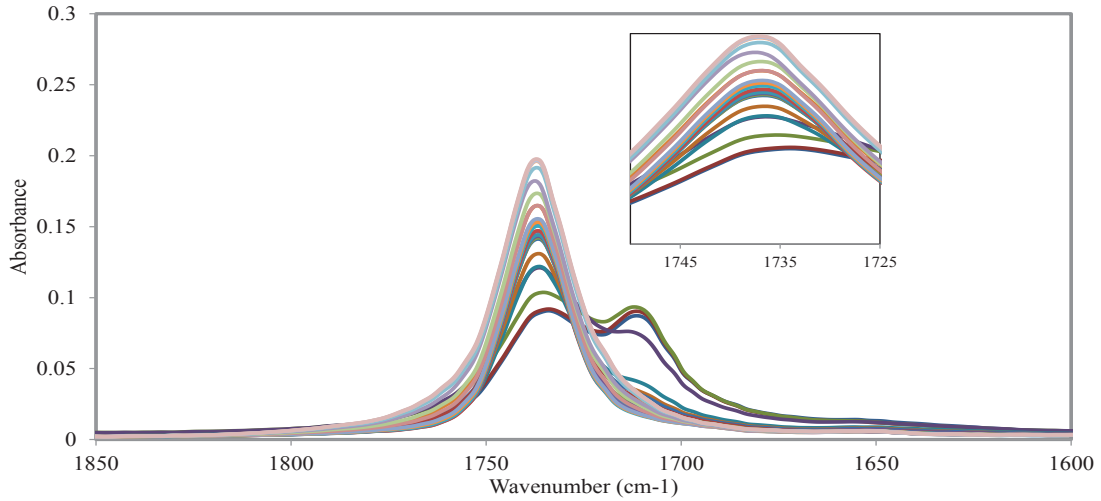

Figure 5.7. FTIR-ATR spectra of samples from 1800 - 1650 cm^{-1}

Table 5.1. Reference total acid number of calibration samples and coressponding absorbance values at 1739 cm^{-1}

Sample No	Refence TAN (mg KOH/g)	Absorbance values at 1739cm^{-1}
S1	157.80	0.0846
S2	114.70	0.0859
S3	105.20	0.0990
S4	79.40	0.1169
S5	41.70	0.1174
S7	38.50	0.1366
S8	33.20	0.1391
S10	27.30	0.1383
S11	17.30	0.1390
S12	15.80	0.1599
S14	12.40	0.1419
S15	11.10	0.1456
S16	9.62	0.1478
S18	8.40	0.1480
S19	7.39	0.1504
S20	6.91	0.1598
S21	6.72	0.1686
S22	5.81	0.1795
S24	5.72	0.1913
S25	4.46	0.1919
S26	4.41	0.1929

Table 5.2. Reference total acid number of independent validation samples and coressponding absorbance values at 1739 cm^{-1}

Sample No	Refence TAN (mg KOH/g)	Absorbance values at 1739cm^{-1}
S6	38.80	0.1264
S9	30.10	0.1372
S13	13.10	0.1418
S17	9.45	0.1453
S23	5.74	0.1867

In the heating process of the esterification reaction during which the first four measurements are made, the Carbon-oxygen double bond peak of the PEG-Oleate is observed to start raising up around 1739 cm^{-1} while the carbon-oxygen double bond of oleic acid peak is observed to start disappearing around 1710 cm^{-1}. During reaction stage and cooling process, carbon-oxygen double bond peak of the PEG-Oleate is significantly incerasing. Therefore, this peak was selected to build simple univariate calibration medel. The plot of reference total acid numbers versus absorbances at 1739 cm^{-1} is given in Figure 5.8.

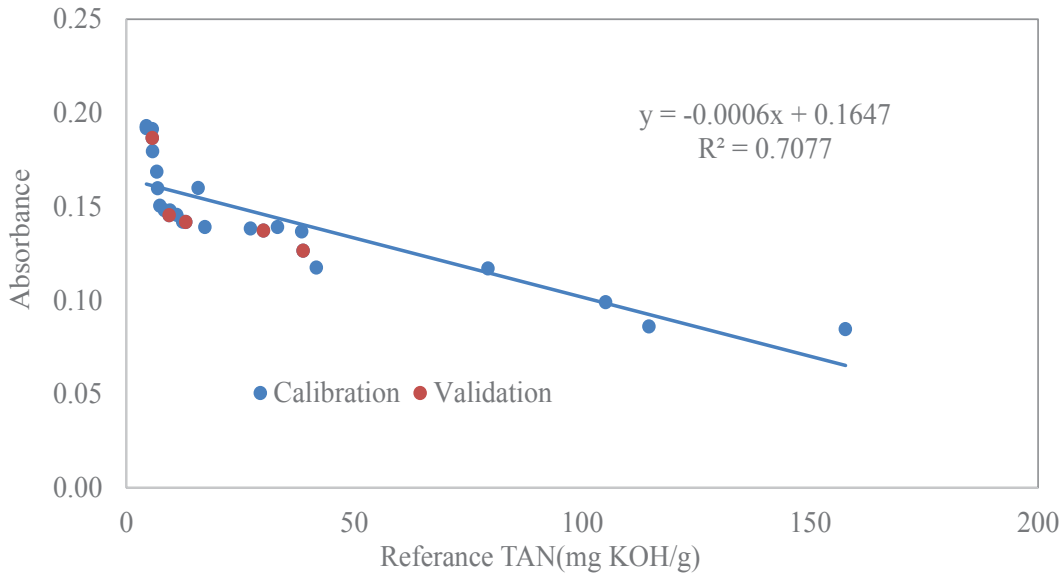

Figure 5.8. Simple univariate calibration plot of PEG-Oleate samples at 1739 cm^{-1}

As seen from Figure 5.8, absorbance values at 1739 cm^{-1} shows a nonlineer trend as a fuction of total acid number of the samples. Resulting regression coefficient shows that the univariate calibration method is insufficient for reaction traceability as the

esterification reaction kinetics is not homogenous enough. Regression coefficient is found to be 0.707. Standart Error of Cross Validation (SECV) and Standart Error of Prediction (SEP) values are calculated for the model as 26.021 and 25.715 mg KOH/g respectively.

The graph showing the relationship between the acid number of PEG-Oleate and the reaction time can be observed in Figure 5.9. The graph shows that the acid number decreased drastically in the heating pocess, which is approximately the initial 24 minutes of reaction time, then gradually slowed down towards the end of the reaction. Finally, it was almost reached to a minimal constant value in the cooling stage.

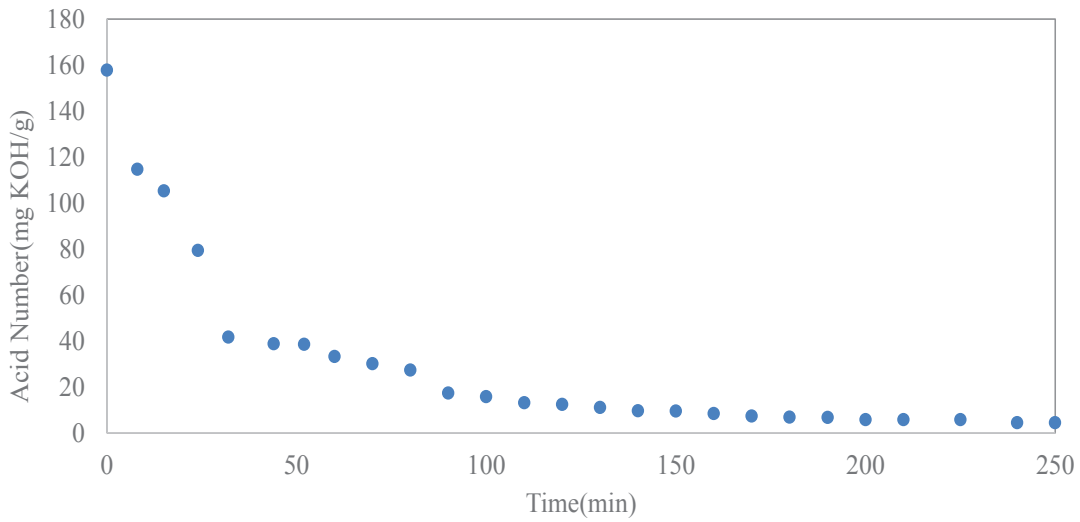

Figure 5.9. Total acid number of samples vs reaction time

As a result, a single wavenumber based calibration determining acid number of polyethylene glycol and oleic acid esterification reaction is not suitable by univariate calibration. Therefore, multivariate calibration methods are applied for the full spectral data which are Genetic Inverse Least Squares (GILS) and Partial Least Squares (PLS).

5.3. Multivariate Calibration Results

5.3.1. Partial Least Squares (PLS) Results

Quantitative analysis of a spectral data can be made by using the Partial Least Squares (PLS) which is a very common and useful multivariate calibration method. In

PLS method, full-spectrum information was used to build a regression model in order to explain the relationship between the TAN values and FTIR spectra.

A total of 26 samples were taken to create PLS calibration model during the esterification reaction of PEG and Oleic Acid. The same calibration and independent validation sets (Table 5.1 and 5.2) in which 21 samples is used as calibration set and 5 samples is used as validation set for the model. These samples were randomly selected except the samples with very small and large TAN value are reserved for the calibration set in order to cover the dynamic range.

Reference and Predicted TAN values by the PLS method for each set can be seen in Table 5.3 and 5.4. The graphical representation of actual TAN vs. predicted TAN values based on FTIR-ATR spectra using PLS method is shown in Figure 5.10.

Table 5.3. Referance versus Partial Least Squares (PLS) predicted total acid number, residual and standart residual values for calibration samples

Sample No	Reference TAN (mg KOH/g)	Predicted TAN (mg KOH/g)	Residuals	Standard Residuals
S1	157.80	133.96	23.84	2.92
S2	114.70	128.86	-14.16	-1.73
S3	105.20	114.19	-8.99	-1.10
S4	79.40	75.59	3.81	0.47
S5	41.70	59.19	-17.49	-2.14
S7	38.50	30.51	7.99	0.98
S8	33.20	24.60	8.60	1.05
S10	27.30	24.91	2.39	0.29
S11	17.30	20.33	-3.03	-0.37
S12	15.80	18.09	-2.29	-0.28
S14	12.40	12.16	0.24	0.03
S15	11.10	11.15	-0.05	-0.01
S16	9.62	9.39	0.23	0.03
S18	8.40	8.80	-0.40	-0.05
S19	7.39	7.88	-0.49	-0.06
S20	6.91	7.68	-0.77	-0.09
S21	6.72	6.64	0.08	0.01
S22	5.81	4.10	1.71	0.21
S24	5.72	5.30	0.42	0.05
S25	4.46	5.23	-0.77	-0.09
S26	4.41	5.30	-0.89	-0.11

Table 5.4. Referance versus Partial Least Squares (PLS) predicted total acid number , residual and standart residual values for validation samples

Sample No	Reference TAN (mg KOH/g)	Predicted TAN (mg KOH/g)	Residuals	Standard Residuals
S6	38.80	42.00	-3.20	-0.39
S9	30.10	28.56	1.54	0.19
S13	13.10	14.27	-1.17	-0.14
S17	9.45	9.96	-0.51	-0.06
S23	5.74	5.31	0.43	0.05

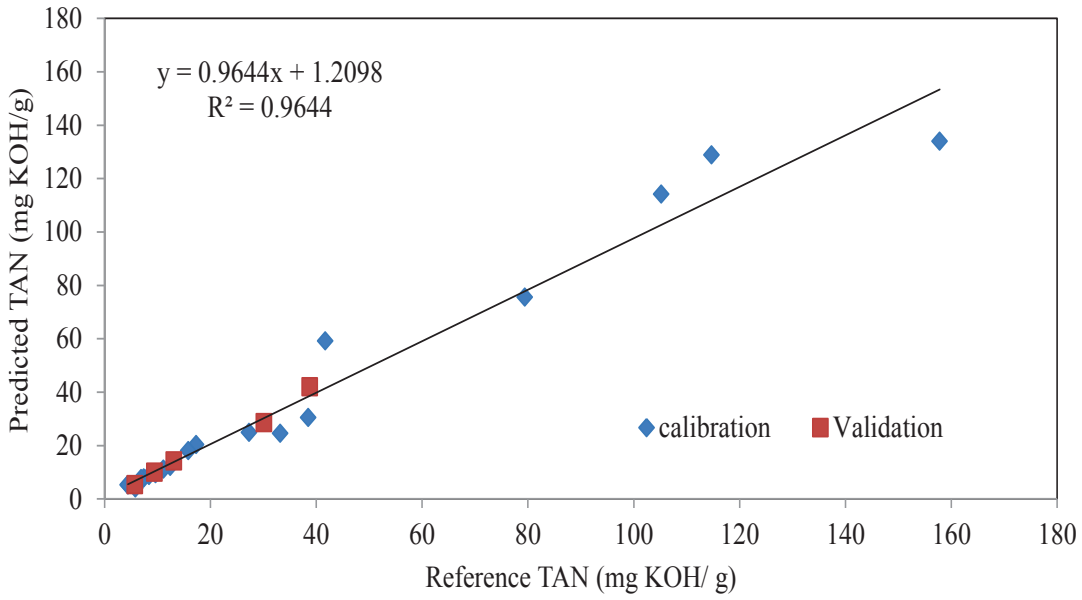

Figure 5.10. Referance versus Partial Least Squares (PLS) predicted acid number

As cen be seen from Figure 5.10, the value of regression coefficient is found to be 0.964. Standart Error of Cross Validation (SECV) and Standart Error of Prediction (SEP) values are calculated for the model as 8.378 and 1.900 mg KOH/g respectively. The number of principle components is 3 according to the PLS method.

SECV values were found much higher than SEP values for the independent validation set, which was caused by deviation in spectral responses from linearity at high TAN values. Acid number of PEG-Oleate has wide range during esterification reaction which is ranged between 4.4 and 157.80 mg KOH/g whereas in validation data set, the TAN values of the selected samples was between 6.20 and 40.74 KOH/g. Residal and standart residual plot for PLS model are shown in Figure 5.11 and Figure 5.12

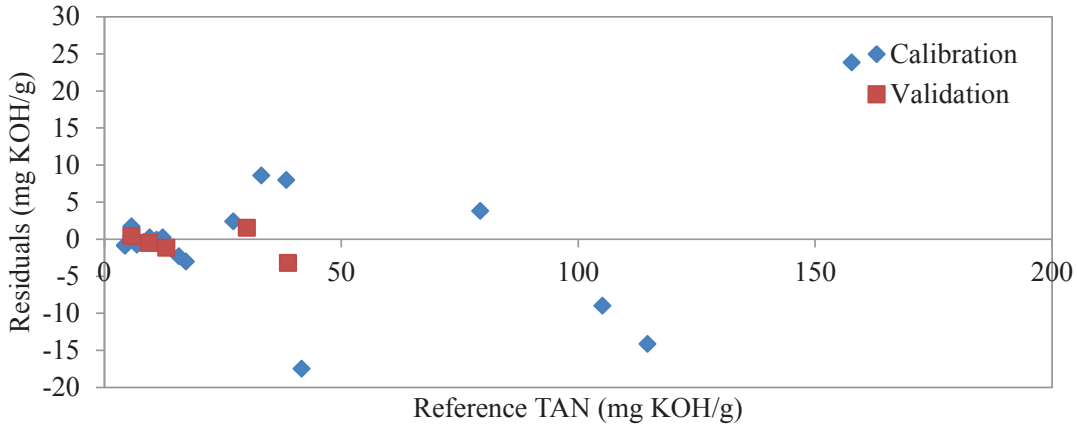

Figure 5.11. Residual plot for PLS model

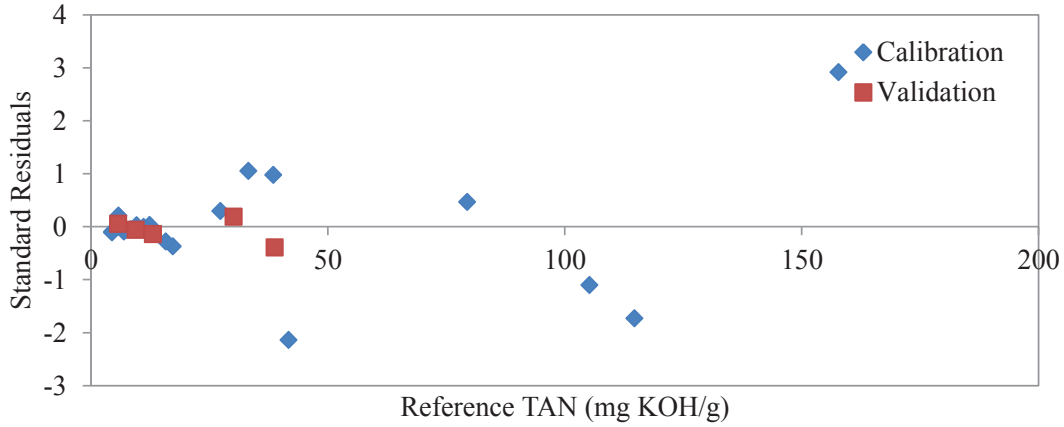

Figure 5.12. Standart residuals plot for PLS model

As the residual plot of PLS model for calibration set found to be scattered in a wide range and validation set is found to have a lower range of residuals which shows the SECV value found much higher than SEP value.

5.3.2 Genetic Inverse Least Squares (GILS) Results

The same 26 samples used in both simple univariate and PLS methods were taken to create GILS calibration model during the esterification reaction of PEG and Oleic Acid. A set of 21 samples is used as calibration set and a set of 5 samples is used as validation set for the model. Experimental and predicted total acid numbers of each set can be seen in Table 5.5 and 5.6.

Table 5.5. Referance versus Genetic Inverse Least Squares (GILS) predicted acid number, residual and standart residual values for calibration samples

Sample No	Reference TAN (mg KOH/g)	Predicted TAN (mg KOH/g)	Residuals	Standard Residuals
S1	157.80	157.52	0.28	0.13
S2	114.70	113.21	1.49	0.66
S3	105.20	106.75	-1.55	-0.69
S4	79.40	79.42	-0.02	-0.01
S5	41.70	44.80	-3.10	-1.39
S7	38.50	37.38	1.12	0.50
S8	33.20	30.92	2.28	1.02
S10	30.10	28.37	1.73	0.77
S11	27.30	26.43	0.87	0.39
S12	17.30	18.55	-1.25	-0.56
S14	15.80	13.95	1.85	0.83
S15	13.10	15.67	-2.57	-1.15
S16	12.40	8.54	3.86	1.72
S18	11.10	10.70	0.40	0.18
S19	9.62	11.18	-1.56	-0.70
S20	8.40	10.38	-1.98	-0.89
S21	7.39	11.19	-3.80	-1.70
S22	6.72	1.72	5.00	2.23
S24	5.81	6.75	-0.94	-0.42
S25	5.72	6.53	-0.81	-0.36
S26	4.41	5.69	-1.28	-0.57

Table 5.6. Referance versus Genetic Inverse Least Squares (GILS) predicted acid number, residual and standart residual values for validation samples

Sample No	Reference TAN (mg KOH/g)	Predicted TAN (mg KOH/g)	Residuals	Standard Residuals
S6	38.80	37.60	1.20	0.54
S9	9.45	5.58	3.87	1.73
S13	6.91	7.50	-0.59	-0.27
S17	5.74	4.34	1.40	0.62
S23	4.46	0.27	4.19	1.87

The graphical representation of referance total acid number vs. predicted values based on - FTIR-ATR spectra using GILS method is shown in Figure 5.13. When compared with the R^2 value from PLS model, the value of regression coefficient is found to be 0.997. Standart Error of Cross Validation (SECV) and Standart Error of Prediction (SEP) values are calculated for the model as 2.295 and 2.694 mg KOH/g respectively.Examination of the calculated and observed values of SECV, SEP and R^2

indicated that the results are aggreable among each other. SECV and SEP values are very closed to each other in GILS method which shows that GILS method eliminates the variables that causes the nonlinearity problems.

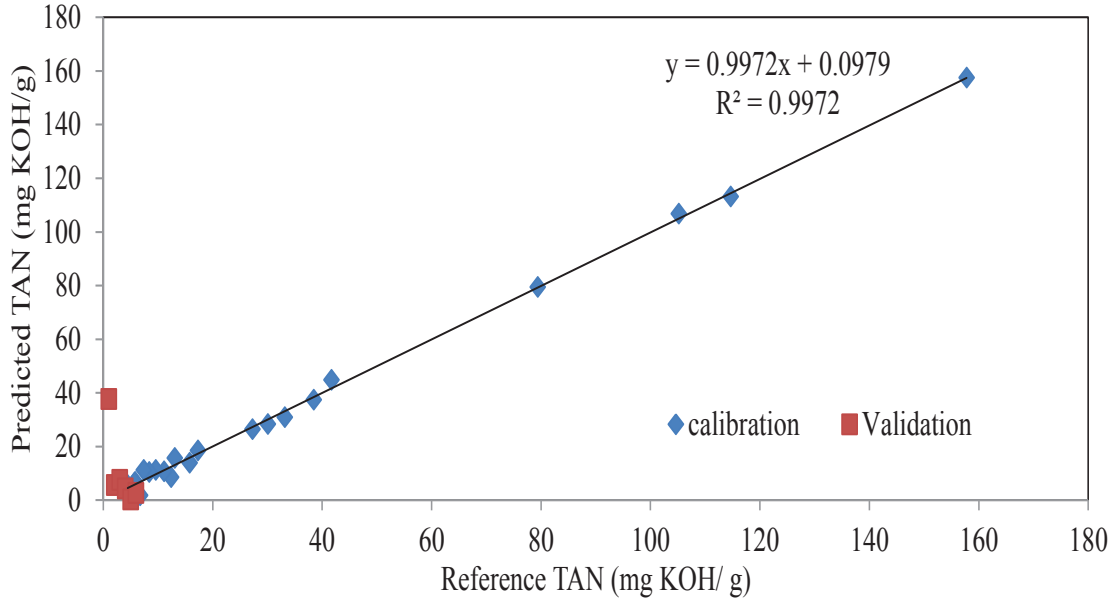

Figure 5.13. Referance versus Genetic Inverse Least Squares (GILS) predicted acid number.

Residal and standart residual plot for GILS model are shown in Figure 5.14 and Figure 5.15. As the residuals and standart residual of the model found to be scattered randomly, it can be concluded that the GILS model is acceptable and indicating a good fit.

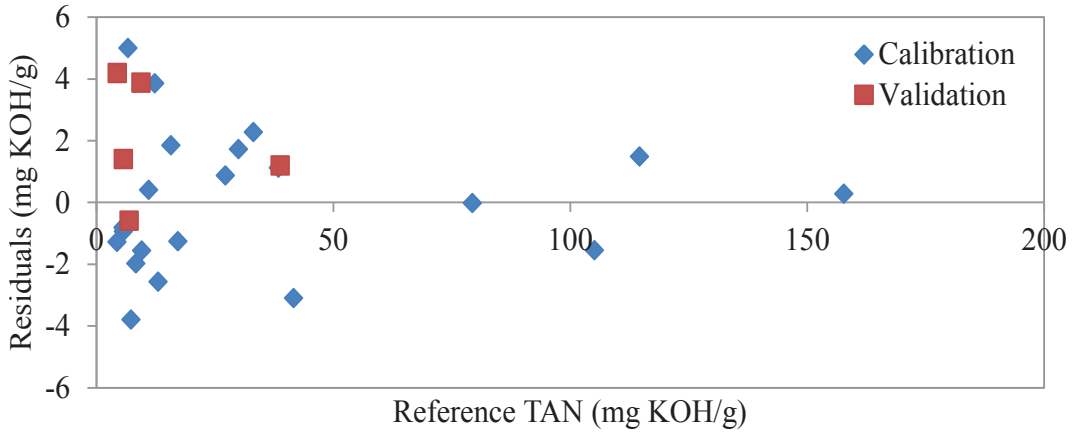

Figure 5.14. Residual plot for GILS model

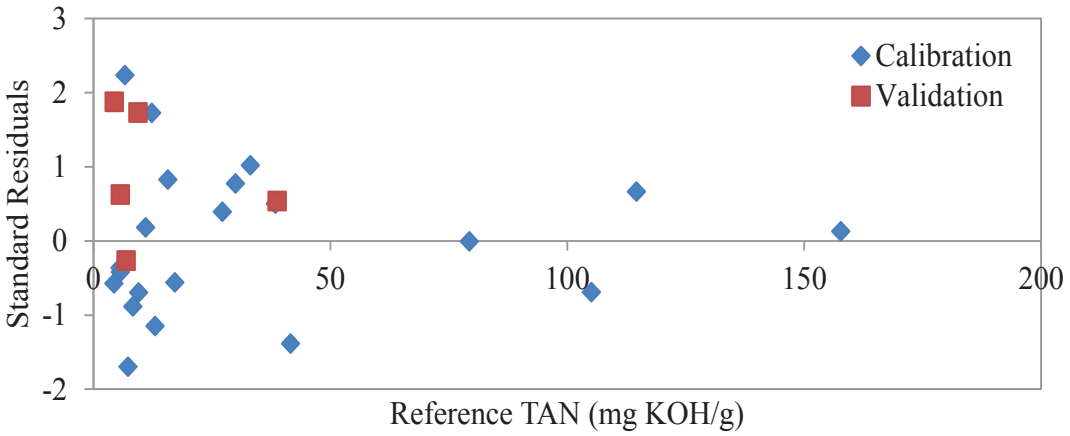

Figure 5.15. Standart residuals plot for GILS model

GILS is a method which depends on variable selection, the distribution of selected wavelengths or wavenumbers in multiple runs over the entire full spectral region would be useful to observe selectivity of GILS over the full spectral range. The plot which can be seen in Figure 5.16 is created by collecting the data in 250 runs with 30 genes and 50 iterations for the given wavelength range of peg-oleate product. As seen in Figure 5.16, a wide range of wavenumbers are selected by GILS method where the highest section frequency is seen around ester peak.This indicetes that GILS has a tendency for finding variables with high correlation to TAN values.

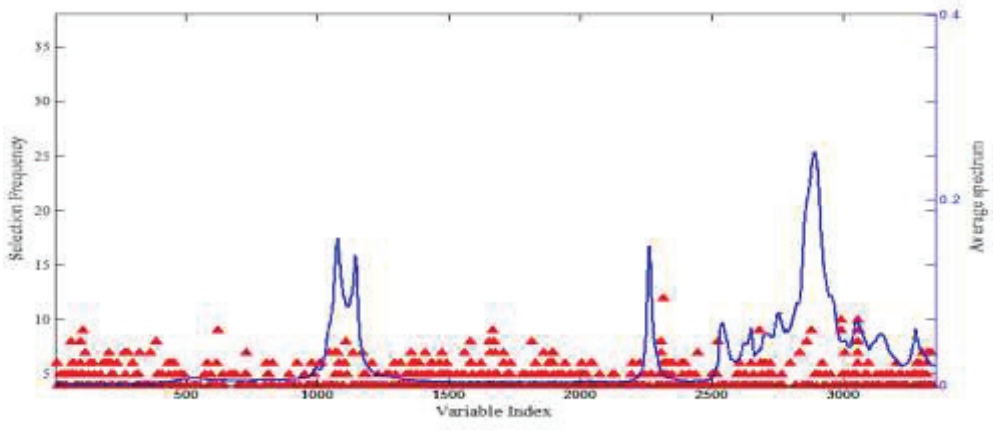

Figure 5.16. Frequency distribution of GILS selected FTIR-ATR wavelengths product

5.4. Comparing Calibration Models

Comparing the univariate and multivariate calibration methods, the multivariate methods found to be more successfull at determining TAN values besed on FTIR spectra.

The strength of the multivariate methods are clearly observed when the regression coefficient values are compared. For a comparison between univariate and multivariate calibration method which are PLS and GILS, the SECV, SEP and regression coefficient values along with the numbeer of PLS components are given in the Table 5.7

Table 5.7. Calibration methods summary

Calibration Model	R^2	SECV (mg KOH/g)	SEP (mg KOH/g)
Genetic Inverse Least Squares (GILS)	0.997	2.295	2.694
Partial Least Squares (PLS)	0.964	8.378	1.900

When results of GILS and PLS methods are compared to each other, GILS method is found to have much higher regression coefficient and much lower SECV values which indicates that the GILS method is more reliable method for the determination of TAN based on FTIR spectroscopy.

5.5. Response Surface Model

For the purpose of creating a mathmatical expression that explains how temperature, time and their interaction affects the total acid number of polyetheylene glycol and oleic acid esterififcation reaction, two factor response surface methodology which include not only linear terms but also quadratic interactions has been applied.

Here, the response surface model was applied to investigate statistical meaning of measurements that had been made. In order to find significance of the factor effects for the total acid number of the product, regression analysis was performed at 95% confidence (alfa value 0.05) level. The second order quadratic model equation is given by Equation 5.1 where acid number is represented by *y* whereas time and temperature are represented by *t* and *T*, respectively. Reference acid number of samples collected from PEG-Oleate esterification along with Time (Min) and Temperature(°C) can be seen in Table 5.8. Regression results are given in Table 5.9.

$$y = b0 + b1 * t + b2 * T + b11 * t * t + b22 * T * T + b12 * t * T + e \quad (5.1)$$

Table 5.8. Total acid number of samples PEG-Oleate esterification along with Time (Min) and Temperature (°C)

Sample No	Time (min.)	Temperature (ºC)	Referance TAN (mg KOH/g)
S1	0	28.7	157.80
S2	8	77.0	114.70
S3	15	107.0	105.20
S4	24	157.0	79.40
S5	32	160.0	41.70
S6	44	160.5	38.80
S7	52	159.9	38.50
S8	60	160.3	33.20
S9	70	160.2	30.10
S10	80	160.1	27.30
S11	90	160.0	17.30
S12	100	159.9	15.80
S13	110	159.7	13.10
S14	120	159.9	12.40
S15	130	157.7	11.10
S16	140	159.8	9.62
S17	150	160.3	9.45
S18	160	159.9	8.40
S19	170	160.2	7.39
S20	180	160.3	6.91
S21	190	159.8	6.72
S22	200	159.9	5.81
S23	210	160.0	5.74
S24	225	100.0	5.72
S25	240	70.0	4.46
S26	250	40	4.41

Table 5.9. Regression results

Term	Coeffient	Standart Error of Coeffient	*T*- value	*P*- value
Constant	25.425	6.517	3.901	0.001
Time (min.)	-53.771	2.121	-25.348	0.000
Temperature (ºC)	-14.058	3.935	-3.573	0.002
Time (min.) x Time (min.)	40.247	7.463	5.393	0.000
Temperature (ºC) x Temperature (ºC)	-1.357	4.707	-0.288	0.776
Time (min.) x Temperature (ºC)	25.134	2.667	9.423	0.000

Table 5.9 shows the significance of the coefficients with the associated *p*-values of each term in the analysis. The *p*-value less than 0.05 indicated that the model terms were significant. According to this information, it can be observed from the table that the yield of the total acid number has a strong relationship with the terms "Time", "Temperature", "Time*Time" and "Time*Temperature" while the term "Temperature*Temperature" is not significant.

The relatiopship between the predicted and the referance values of the total acid numbers can be seen in Figure 5.17. As it can be observed from the Figure 5.17, experimental and predicted values are in an acceptable aggreement with a R^2 value of 0.981. Figure 5.18 represents the result of the residual analysis of the model which shows the values of the residuals versus the predicted response of the model.

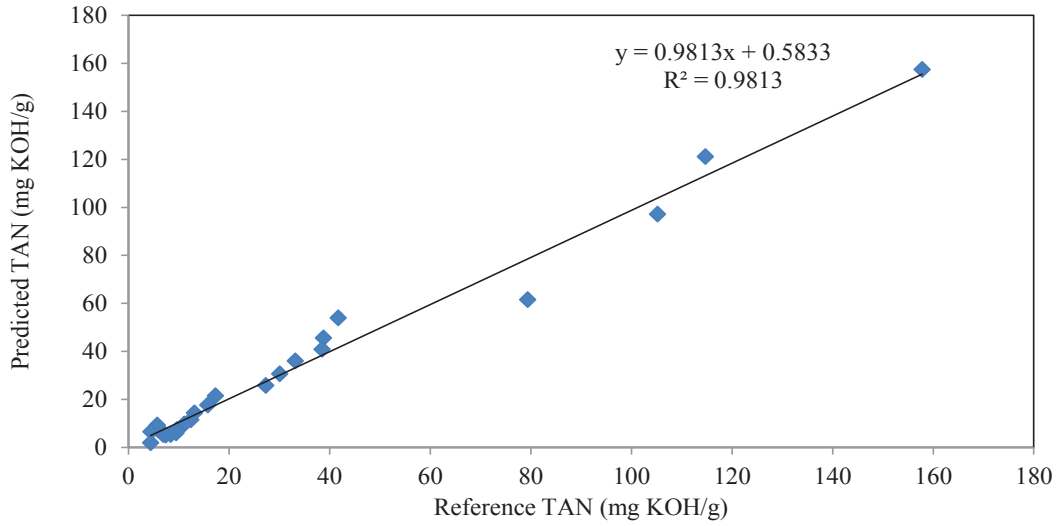

Figure 5.17. Referance acid number vs predicted acid number for the full quadratik model from response surface modelling

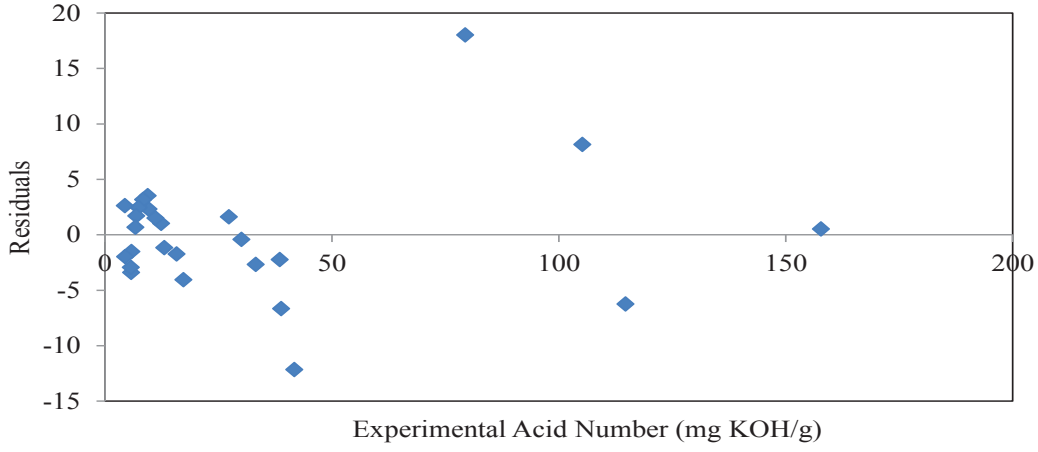

Figure 5.18. Residuals plot for the response surface model

As the residuals of the model found to be scattered randomly, it can be concluded that the proposed mathematical model is acceptable for being an explanatory expression for the experimental result.

The potential relevance between the factors, time, temperature and the response total acid number can be determined by using 3-D response surface plot. The 3-D response surface which is the graphical representation of the regression equation as a fuction of time and temperature. The 3-D surface plot and two dimensional contour plot of the total acid number can be found in Figure 5.19 and Figure 5.20.

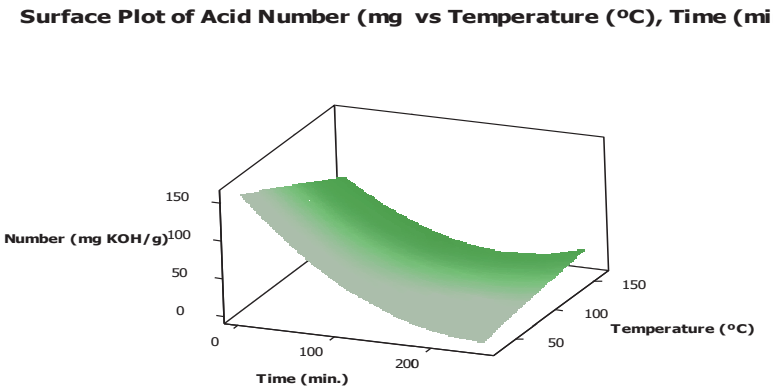

Figure 5.19. Surface Plot of Acid Number (mg KOH/g) vs Temperature (°C), Time (min)

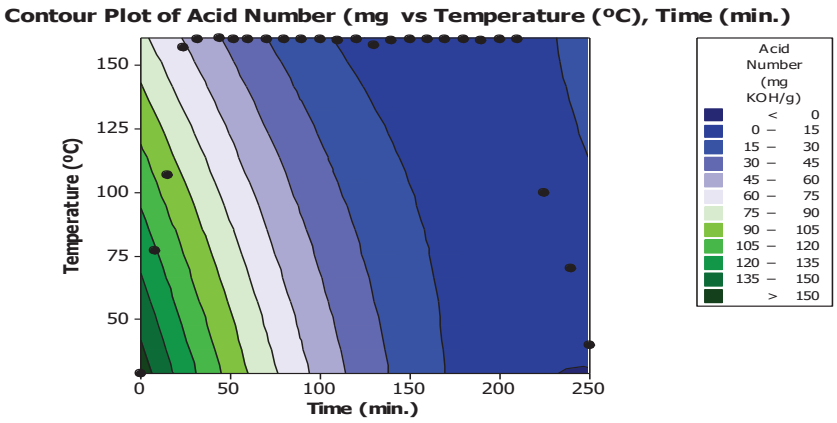

Figure 5.20. Contour Plot of Acid Number (mg KOH/g) vs Temperature (°C), Time (min)

As it can be seen in Figure 5.19, the total acid number of PEG-Oleate esterification proces was found to be minimized where time (min) is held at its highest value and temperature is held at its lowest. Among the two factors, time is found to have a significant effect on acid number. There is a dramatic decrease in the acid number with an increase of time (min.) where any change of temperature is not as effective as the time variable to the acid number.

The contour plot in Figure 5.20 shows the effect of temperature and time on acid number. When the time increases, the tendency of the acid number decreases. After approximately 170 minutes, the acid number is observed to be minimized to range of 0-15 mg KOH/g which is represented by the dark blue area on contour plot. Additionaly, the contour plot of the fitted model magnified the insignificance of the temperature variable over the resulting Acid Number for the time values higher than about 170 minutes.

If the model is going to be aimed to optimize the efficiency of the reaction, the actual effect of the temperature over the total acid number must be discussed in more detail. Theoretically, most efficient reaction temperature of the PEG-Oleate esterification is known to be between 140-180º C according to Section 1.1 PEG-Oleate Production. Given that, a different statistical design of experiment should be created including new data sets in which acid number values of the samples are recorded at different constant temperature values over the time.

CHAPTER 6

CONCLUSION

In this study, a new method for the determination of total acid number was developed for PEG-Oleate esterification reaction based on use the FTIR-ATR spectroscopy along with univariate calibration and chemometric multivariate calibration which are Genetic Inverse Least Squares(GILS) and Partial Least Squares (PLS).

The spectral data and the reference acid number values of the samples were then used in multivariate calibration step using GILS and PLS methods along with univariate calibration on a specific wavenumber corresponding to the ester peak around 1739 cm^{-1}. Altough the changes in the ester peak around 1739 cm^{-1} was showing an inrease associated to the esterification of the reactants, the results of the univariate calibration was unsucsesful and indicated partially nonlinear response as a fuction of TAN values. Compared to the univariate calibration, the PLS results were much better but for the high TAN values there was still some deviation from linearity therefore standard error of cross validation (SECV) values were much higher than standard error of prediction (SEP) values for the independent validation set. Accuracy of the multivariate calibration models were determined by SECV and SEP values as well as with the R^2 values from the reference vs. predicted plots. In addition, the perfomance of PLS and GILS were compared where GILS method is found to be more accurate and reliable. As a result, GILS method which has the advantage of eliminating the variables that causes the nonlinearity problems, was also used for model building. The best regression coefficient was found to be 0.997 along with SECV and SEP as 2.295 and 2.694 mg KOH/g, respectively.

With the help of this new FTIR spectroscopy based GILS method, determination of the acid number of a product turned out to be possible to be calculated in an easier, healtier and independent from human errors.

REFERENCES

1. Echem-group.com. (2019). *PEG esters | Echem*. [online] Available at: http://www.echem-group.com/product/peg-esters

2. Abdullah, F., Ma'amor, A., Daud, N. and Abd. Hamid, S. (2017). Selective synthesis of PEG-monoester using cesium heteropoly acid as heterogeneous catalyst. *Química Nova*.

3. 17e1, A. (2019). *ASTM D5773 - 17e1 Standard Test Method for Cloud Point of Petroleum Products and Liquid Fuels (Constant Cooling Rate Method)*. [online] Astm.org. Available at: https://www.astm.org/Standards/D5773.htm

4. En.wikipedia.org. (2019). Hydroxyl value. [online] Available at: https://en.wikipedia.org/wiki/Hydroxyl_value

5. Cpieng.com. (2019). *CPI Engineering Services*. [online] Available at: http://www.cpieng.com/totalAcidNumber.htm

6. 14e2, A. (2019). *ASTM D974 - 14e2 Standard Test Method for Acid and Base Number by Color-Indicator Titration*. [online] Astm.org. Available at: https://www.astm.org/Standards/D974.htm

7. de Lira, L., de Albuquerque, M., Pacheco, J., Fonseca, T., Cavalcanti, E., Stragevitch, L. and Pimentel, M. (2010). Infrared spectroscopy and multivariate calibration to monitor stability quality parameters of biodiesel. *Microchemical Journal*, 96(1), pp.126-131.

8. Koczon, P., Gruczynska, E. and Kowalski, B. (2008). Changes in the Acid Value of Butter During Storage at Different Temperatures as Assessed by Standard Methods or by FT-IR Spectroscopy. *American Journal of Food Technology*, 3(3), pp.154-163.

9. CAROLEI, L. and GUTZ, I. (2005). Simultaneous determination of three surfactants and water in shampoo and liquid soap by ATR-FTIR. *Talanta*, 66(1), pp.118-124

10. Handbook of Instrumental Techniques for Analytical Chemistry. (1998). *Journal of Liquid Chromatography & Related Technologies*, 21(19), pp.3072-3076.

11. Bischoff, J., Derrick, M., Stulik, D. and Landry, J. (2001). Infrared Spectroscopy in Conservation. *Journal of the American Institute for Conservation*, 40(3), p.268.

12. Wps.prenhall.com. (2019). *Media Portfolio*. [online] Available at: https://wps.prenhall.com/wps/media/objects/724/741576/chapter_13.html

13. Mendelsohn, R. (2007). Fourier Transform Infrared Spectrometry, 2nd ed By Peter R. Griffiths (University of Idaho, Moscow) and James A. De Haseth (University of Georgia, Athens). J. Wiley & Sons, Inc.: Hoboken, NJ. 2007. xviii + 530 pp. Journal of the American Chemical Society, 129(43), pp.13358-13358.

14. Thermo.com.cn. (2019). [online] Available at: http://www.thermo.com.cn/Resources/200802/productPDF_21615.pdf

15. Anon, (2019). [online] Available at: https://shop.perkinelmer.com/content/TechnicalInfo/TCH_FTIRATR.pdf

16. Zeegers-Huyskens, T. and Bator, G. (1996). Fourier transform infrared and Fourier transform Raman investigation of alkylammonium hexachloroantimonates. *Vibrational Spectroscopy*, 13(1), pp.41-49.

17. Mackie, D., Jahnke, J., Benyamin, M. and Sumner, J. (2016). Simple, fast, and accurate methodology for quantitative analysis using Fourier transform infrared spectroscopy, with bio-hybrid fuel cell examples. MethodsX, 3, pp.128-138.

18. Wold, S. (1995). Chemometrics; what do we mean with it, and what do we want from it?. Chemometrics and Intelligent Laboratory Systems, 30(1), pp.109-115.

19. Tobias, R., Beebe, K., Pell, R. and Seasholtz, M. (1999). Chemometrics: A Practical Guide. Technometrics, 41(4), p.375.

20. Brereton, R. (2000). Introduction to multivariate calibration in analytical chemistry. The Analyst, 125(11), pp.2125-2154.

21. Üner, B., Karaman, İ., Tanrıverdi, H. and Özdemir, D. (2010). Determination of lignin and extractive content of Turkish Pine (Pinus brutia Ten.) trees using near infrared spectroscopy and multivariate calibration. Wood Science and Technology, 45(1), pp.121-134.

22. Öztürk, B., Yalçin, A. and Özdemir, D. (2010). Determination of Olive Oil Adulteration with Vegetable Oils by near Infrared Spectroscopy Coupled with Multivariate Calibration. *Journal of Near Infrared Spectroscopy*, 18(3), pp.191-201

23. Haaland, D. and Thomas, E. (1988). Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Analytical Chemistry, 60(11), pp.1193-1202.

24. Wang, Y., Veltkamp, D. and Kowalski, B. (1991). Multivariate instrument standardization. Analytical Chemistry, 63(23), pp.2750-2756.

25. Gilbert, R., Goodacre, R., Woodward, A. and Kell, D. (1997). Genetic Programming: A Novel Method for the Quantitative Analysis of Pyrolysis Mass Spectral Data. Analytical Chemistry, 69(21), pp.4381-4389.

26. Li, T., Lucasius, C. and Kateman, G. (1992). Optimization of calibration data with the dynamic genetic algorithm. Analytica Chimica Acta, 268(1), pp.123-134

27. Lucasius, C. and Kateman, G. (1991). Genetic algorithms for large-scale optimization in chemometrics: An application. TrAC Trends in Analytical Chemistry, 10(8), pp.254-261.

28. JAD, 2. (2019). *January 2007: Genetic Algorithms (GAs)*. [online] Edc.ncl.ac.uk. Available at: http://www.edc.ncl.ac.uk/highlight/rhjanuary2007.php