THE DEVELOPMENT OF CHEMOMETRIC METHODS BASED ON MOLECULAR SPECTROSCOPY FOR THE STANDARDIZATION OF PRODUCTION PROCESSES AND PRODUCT TRACEABILITY OF PERSONAL CARE AND CLEANING PRODUCTS

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
Berfu ÇİFTÇİ İLMEK

July 2019
İZMİR

We approve the thesis of **Berfu ÇİFTÇİ İLMEK**

Examining Committee Members:

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

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

Doc. 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
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGEMENTS

This thesis is dedicated to them who have given me their unconditional support, both financially and emotionally throughout my degree.

I wish to sincere thanks to my research advisor Dr. Durmuş ÖZDEMIR, for his support and guidance in the completion of my master research program. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Dr. Şerife Hanım YALÇIN and Doc. Dr. Levent PELİT for their encouragement, insightful comments, and hard questions.

This study would not be possible without the support of Dalan Kimya Endüstri A.Ş. and The Scientific and Technical Research Council of Turkey (TUBİTAK-Sanayi AR-GE Projeleri Destekleme Programı, Project number 3140461).

I would also like to thank the members of Chemometrics Research Group, Başak BAŞAR and Ekin Meşe for helping me during my graduate years.

I owe very special thanks to Nilüfer KIZILKAYA ÖKSÜZ who is my best friend from high school. She is always with me every time I needed her. My sincere thanks to Nazlı NOYAN TAŞTAN and Merve SAMUR for all the emotional support, friendship, entertainment, and caring they provided every time.

Lastly, I would like to thank my family, my husband Selim İLMEK, my father Ali ÇİFTÇİ, my mother Tülin ÇİFTÇİ and my lovely sister Alkım, for their motivation, patience, understanding and supporting me spiritually throughout my life.

ABSTRACT

THE DEVELOPMENT OF CHEMOMETRIC METHODS BASED ON MOLECULAR SPECTROSCOPY FOR THE STANDARDIZATION OF PRODUCTION PROCESSES AND PRODUCT TRACEABILITY OF PERSONAL CARE AND CLEANING PRODUCTS

Personal care and cleaning products are the main consumer goods. Changes in our heath caused by all of the chemicals that we exposed to everyday if these products are not produced according to the regulations and determined formulations. Because of this reason, quality control of the product formulation quantitatively is very important. There are some analytical methods for the determination of anion active matter, nonionic matter and total active matter in the product mixture. However, these techniques are expensive and do not give accurate results.

The purpose of this thesis principally based on development of rapid, accurate and practical infrared spectroscopic technique based on multivariate chemometrics data analysis methods for the standardization of production processes and product traceability of personal care and cleaning products.

In this thesis, two different products are studied which are namely liquid soap and shower gel. Fourier Transform Infrared spectroscopy coupled with Attenuated Total Reflectance accessory based chemometrics multivariate calibration models were developed for the quantitative determination of liquid soap and shower gel compounds. Genetic Inverse Least Squares was used as the chemometrics method for the development of multivariate calibration models in the quantitative determination of liquid soap and shower gel compositions.

Standard error of cross validation and standard error of prediction values for content of the liquid soap samples were found 0.26% and 0.21 % (w/w %), respectively. Standard error of cross validation and standard error of prediction values for content of the shower gel samples were found 0.27 % and 0.30 % (w/w %), respectively.

ÖZET

KİŞİSEL BAKIM VE TEMİZLİK ÜRÜNLERİ ÜRETİM SÜREÇLERİNİN STANDARDİZASYONU VE ÜRÜN İZLENEBİLİRLİĞİ İÇİN MOLEKÜLER SPEKTROSKOPİYE DAYALI KEMOMETRİK METOTLARIN GELİŞTİRİLMESİ

Kişisel bakım ve temizlik ürünleri ana tüketim maddelerindendir. Eğer bu ürünler regülasyonlara ve belirlenen formülasyonlara uygun üretilmezlerse her gün maruz kaldığımız bu kimyasallar sağlık problemlerine yol açabilir. Bu sebeple ürün formülasyonlarının kontrol edilebilmesi önemlidir. Ürün karışımındaki anyon aktif madde, katyonik aktif madde, noniyonik aktif madde ve toplam aktif madde içeriğini belirlemek için analitik metotlar bulunmaktadır. Ancak bu teknikler pahalıdır ve kesin sonuçlar verememektedir.

Bu projenin esas amacı kişisel bakım ve temizlik ürünleri üretim süreçlerinin standardizasyonu ve ürün izlenebilirliği için kızılötesi spektroskopiye dayalı çok değişkenli kemometrik veri analizi ile hızlı, kesin ve pratik metotlar geliştirmektir.

Bu projede, sıvı sabun ve duş jeli olarak 2 ayrı ürün çalışılmıştır. Sıvı sabun ve duş jeli bileşenlerinin miktarsal tayini için Fourier dönüşümlü kızılötesi spektroskopisi kullanılarak alınan spektral verilere kemometrik çok değişkenli kalibrasyon metotları uygulanarak yeni bir analitik metot geliştirilmesi hedeflenmiştir. Çok değişkenli kalibrasyon metotlerından biri olan genetik algoritma tabanlı ters en küçük kareler yöntemi kullanılarak çok değişkenli kalibrasyon modelleri oluşturulmuş ve bu modeller bağımsız kalibrasyon ve validasyon data seti ile test edilmiştir. Sıvı sabun bileşenine ait kalibrasyon ve validasyon setlerine ait hata değerleri %0,26 ve %0,21 aralığında tanımlanmıştır. Duş jeli bileşenine ait kalibrasyon ve validasyon setlerine ait hata değerleri %0,27 ve %0,30 aralığında tanımlanmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The cosmetic products are any substance or preparation that include mixtures from naturally or synthetically sourced chemical compounds (Manayi and Saeidnia 2014). The cosmetic products are employed to essentially or completely cleaning, perfuming, improving appearance, correcting odors, and protecting the human body or keeping it in the right conditions (European Parliament and of the Council 2009). The striking element in the definition of cosmetic ingredients one of such products is usually a mixture which includes major and minor components. Hence, when contained in these products are produced as mixtures containing more than one substance content components mentioned for the terms to be determined whether produced within the legal limits components that develop analytical methods for quantitative analysis of high priority and importance. A cosmetic product must be produced according to the formulation which are given by R&D Department, if the production is not done in that way the consequences are; consumer health might be in danger and there can have bad effects on the environment. Therefore, the most important thing is to control production.

1.1. Scope of the Thesis

The TS ISO 6842 standard defines the analytical methods to regulate anion active matter, nonionic active matter, and total active matter for the cosmetic product mixtures (TS ISO 6842 1989). However, these techniques have failed to give accurate results due to peak of some raw materials nearby. Accordingly, this research aims to develop a project that performs the rapid, accurate and practical molecular spectroscopic technique principals. In the methodology of this research, multivariate chemometric data analysis methods were used due to their advantages in analyzing and visualizing the complex and multi-dimensional data to extract information (Kowalik and Einax 2006). This research adapts multivariate chemometric data analysis methods to test the standardization of production processes and product traceability of personal care and cleaning products.

1.2. Literature Review

In recent years, there has been an increasing interest on the studies focusing surfactant percentages in some cosmetic products determination. These studies applied different techniques for to test and analyze the cosmetic products determination, such as spectrophotometric and potentiometric (with the use of ion_selective electrodes) techniques, including flow injection analysis versions, chromatographic techniques (primarily, high-performance liquid chromatography) attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy in the middle infrared region, the Immersion Refractometer, ion-pair potentiometric surfactant titrations, Fourier transform infrared spectrometry and multivariate analysis, RP-HPLC/ELSD and LC/MS.

To determine anionic surfactants, the majority of procedures are used for the analysis of environmental materials. Hence, the determination of anionic surfactants studies is the first in different water samples. In the last years, potentiometric (with the use of ion-selective electrodes) and spectrophotometric techniques, including the versions of flow injection analysis, have been the most common methods of anionic surfactants determinations. Ya. R. Bazel *et al.* (2014) studied about investigation anionic surfactants determination in water samples in Ukraine (Bazel et al. 2014). In their study, spectrophotometric determination methods were developed to decrease the toxic extracts, which occurs consequences of extraction. Currently, highly selective sensors have been developed which oriented the studies to potentiometry. In addition to these methods, Carolei and Gutz (2005) applied two multivariate qualification methods to determine three different surfactants in liquid soap/shampoo (Carolei and Gutz 2005). These methods are inverse least squares (ILS) and classical least squares (CLS). In the middle infrared region of the spectrum (800–1600 and 1900–3000 cm−1) absorbance all data was collected, which belongs to the undiluted samples and of the calibration standards. Sodium lauryl ether sulfate (SLES) and Cocamidopropyl betaine (CAPB) are common for both liquid soap and shampoo, alkylpolyglucoside (APG) is the third surfactant of the liquid soap and cocodiethanolamide (CDEA), the corresponding ingredient of the shampoo were selected for the determination by the attenuated total reflection Fourier transform infrared (ATR-FTIR) were 5% of the components.

Another method to analyze liquids is the immersion refractometer, which was developed to determine refractive index of solution concentrations, such as in the sugar

industry, pharmaceuticals, and milk (Niskanen, Hibino, and Räty 2016). Inspired by this information, Hoyt and Verweibe (1926) studied determination of the liquid soaps by the Immersion Refractometer (Hoyt and Verwiebe 1926). Mentioned liquid soap is not surfactant base, it contained vegetable oil, coconut oil, and distilled water. If direct measurement of total solid material and Immersion Refractometer measurement are compared, there are differences 0,3-1,8 percent for the nearly 20 % concentration of solid matter. In the studies conducted by Samardžić *et al.* (2011), potentiometric surfactant sensor which is based on a tetraphenylborate (TPB) antagonist and highly lipophilic 1,3-didecyl-2-methylimidazolium cation an ion were used as the end-point detectors in ion-pair potentiometric surfactant titrations using sodium TPB as a titrant (Samardžić, Sak-Bosnar, and Madunić-Čačić 2011). In their study, a sensor named the end-point detector was used to potentiometrically titration of two-component combinations of each cationic surfactant and ethoxylated nonionic surfactants. On the other hand, the potentiometrically titration of three commercial products that contain CSs as disinfectants and nonionic surfactants were performed. As a result, the comparative results obtained with two-phase titrations for CSs and a gravimetric method for nonionic surfactants. Besides, to determine one nonionic and two anionic surfactants, the studies of Kargosha *et al.* (2008) based on Fourier transform infrared spectrometry and multivariate analysis were used (Kargosha et al. 2008). As the anionic surfactant, sodium lauryl ether sulfate (SLES) and linear alkylbenzene sulfonate (LABS), as nonionic surfactant, coconut diethanol amide (CDEA) were identified. The data combined with PLS regression shows significant appropriations for the determinations of three surfactants. In comparison with conventional methods such as classical, potentiometric, and extraction methods, the PLS method is better than those. In addition, Im *et al*. (2008) performed the shampoo and hair conditioner anionic, amphoteric, nonionic and cationic surfactant mixture analysis by reversed-phase -HPLC/evaporative light scattering detection and LC/MS (Im, Jeong, and Ryoo 2008).

Overall, the studies reviewed in the literature have been done for only synthetic products, disregarding the effects of commercial products. Therefore, in this thesis study, the surfactants that reveal as the products sold to the consumer were analyzed through combined methodology of Fourier Transform Infrared Spectroscopy (FTIR) and multivariate calibration techniques. The aim this study is to standardize the production processes and product traceability of personal care and cleaning products.

CHAPTER 2

INSTRUMENTATION

2.1. Infrared Spectroscopy (IR)

Spectroscopy is a scientific measurement technique which deals with electromagnetic radiations and matter. The light that is adsorbed, emitted, and scattered by materials is measured, and the technique can be used to quantify and identify those materials. Spectroscopic techniques are consist of two basic concepts including atomic spectroscopy and molecular spectroscopy. The study of the electromagnetic radiation absorbed and emitted by atoms in the vapor state is atomic spectroscopy. The study of the electromagnetic radiation absorbed and emitted by molecules in the vapor or solid state is molecular spectroscopy.

Infrared (IR) spectroscopy is the most useful tool for molecular structural identification and quantitative analyses of materials. The advantage of Infrared spectroscopy is any solid, liquid, or gas sample to be analyzed. Therefore this is widely used in chemical, pharmaceutical, environmental, surface sciences, and food chemistry. Basically, it is the absorption measurement of various IR frequencies by a sample positioned in the path of an IR beam. Different functional groups in a sample absorb different frequences of infrared radiation, and in order to determine chemical structures over these characteristic frequencies.

Infrared radiation spans a section of the electromagnetic spectrum having wavenumbers from approximately 13,000 to 10 cm^{-1}, or wavelengths from 0.78 to 1000 μm. It is limited by the red end of the visible region at high frequencies and the microwave region at low frequencies. IR absorption positions are usually shown as not only wavenumber (ν) but also wavelengths (λ). Wavenumber are defined by the number of waves per unit length. Besides that, wavenumbers are directly proportional to frequency, as well as the energy of the IR absorption. Unit of the wavenumber (cm^{-1}, reciprocal centimeter) is commonly used in modern IR instruments which are linear in the cm^{-1} scale. The other way round, wavelengths are inversely proportional to frequencies and

their associated energy. At present, the recommended unit of wavelength is μm (micrometers), but μ (micron) is used in older literature. (Settle, 1998)

Infrared is generally consists of three spectral regions: near, mid and far-infrared. The higher-energy near-IR, roughly 12800–4000 cm^{-1} (0.78–2.5 μm wavelength) can excite overtone or harmonic vibrations. The mid-infrared, roughly 4000–200 cm^{-1} (2.5–50 μm) may be used to study the associated vibrational-rotational structure and fundamental vibrations. The far-infrared, roughly 200–10 cm^{-1} (50–1000 μm), nearest adjacent to the microwave region, has low energy and might be used for rotational spectroscopy. (Derrick et al. 1999) The wavelength ranges of these three infrared regions are shown in Table 2.1. (Skoog, Holler, and Crouch 1998)

Table 2.1. The corresponding wavelengths and wavenumbers of the Infrared regions.

Regions	**Wavelength Range (nm)**	**Wavenumber Range (cm^{-1})**
Near-Infrared (NIR)	780-2500	12.800-4000
Mid-Infrared (MIR)	2500-50.000	4000-200
Far-Infrared (FIR)	50.000-1 x 10^6	200-10

According to the past studies the IR region is divided into two segments by the chemists: The region from 4000 cm^{-1} to approximately 1500 cm^{-1} is named as 'Peak ID Region' since it is mainly used for correlating peak location with bonds. Additionally, the region from 1500 to 600 cm^{-1} is typically very engaged and is not as constructive for such correlation, but it remains very constructive as the molecular fingerprint. Because of this, the region is called as 'Fingerprint Region'. This means that the region can still be used for peak-for-peak matching with a known spectrum from a library of known spectra.

For infrared absorption measurement, commercially, there are three types of instruments which are dispersive instruments, multiplex instruments, and nondispersive instruments (Skoog, et al. 1998).

A dispersive instrument has a monochromator with a grating element to disperse the radiation flowing from the source into its wavelengths where could used as a wavelength selecting the instrument. It is generally designed double-beam, that is, incoming IR radiation is separated into two beams in order to pass through the reference and sample materials. Therefore, the amplified signal and interferences of air during the

analysis are prevented. A shown figure for a dispersive instrument is shown in Figure 2.1 (Smith 1996).

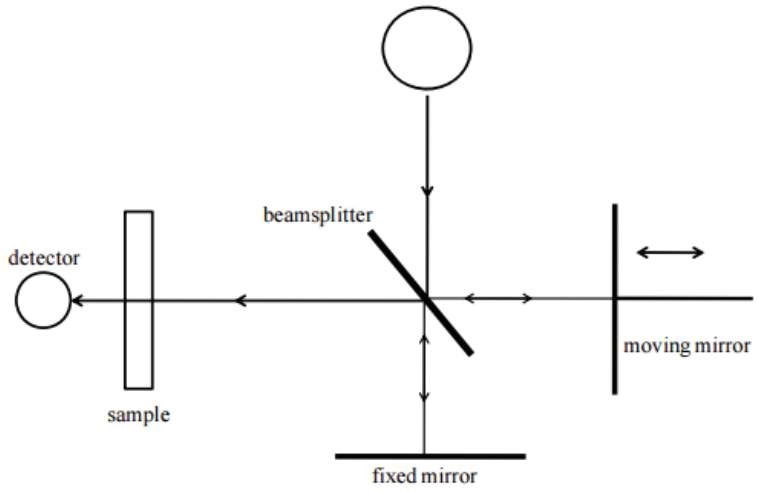

Figure 2.1. A schematic representation for an FT instrument (Skoog, *et al.* 1998).

Non-dispersive instruments have a filter or non-dispersive photometers which are designed for quantitative analysis. Mostly, they are not complicated. Moreover, they are user friendly and not expensive compared to the instruments discussed above (Skoog, et al. 1998).

2.1.1. Attenuated Total Reflectance Fourier Transform Infrared (ATR -FTIR) Spectroscopy

Some scientists used the popular and trustable fingerprinting method which is Mid-Infrared spectroscopy for the characterization and the quantification of the substances. Although it has many disadvantages. The transmission technique of sampling is mostly preferred the way of obtaining mid-infrared. If samples are thicker than 20 microns absorb too much infrared radiation, as a result it will be not probable to get a spectrum. Because of this, thickness problem could be counted as one of the disadvantages. Also, if the sample is thinner than 1 micron, its absorption will be so weak that it cannot be analyzed. Other disadvantage is wasting of time because of the sample preparation consist of the following processes as squishing, melting, or diluting requirements lead into transmits the appropriate amount of light. For this reason, in order to collect mid-infrared spectra, the reflectance techniques can be more preferred than transmission ones. There is no thickness problem for the reflectance techniques. Therefore, the thickness or the concentration of the sample is not concerned. In addition

to this, there is not much time-consumption for sample preparations. As it means that the sample preparation for reflectance samples is easier and faster than for transmittance samples. A final advantage of some reflectance techniques is that they are nondestructive. The sample is left same condition right after its spectrum is collected, where the sample can be analyzed for other studies (Stuart, 2004).

The attenuated total reflectance (ATR) technique is used to obtain the spectra of solids, liquids, semi-solids, and thin films that is the reflectance technique used with mid-infrared spectroscopy.

It can be clearly seen that the ATR-FTIR specroscopy is highly performed where an accessory is placed into sample holder of an FTIR instrument. The accessory has a infrared transparent crystal material with a high refractive index that mirrors the IR radiation to focus on the face of the crystal. Infrared radiation passes through the crystal and reaches to its top surface. At this case, if the crystal has the proper refractive index and the light has the proper angel of incidence, the infrared radiation and reflects off the crystal surface rather than leaving it. This is named as total internal reflection. The infrared beam reflects off the crystal surface three times earlier as leaving the crystal. This procedure is presented in the schematic diagram of an ATR accessory in Figure 2.4. (Stuart, 2004).

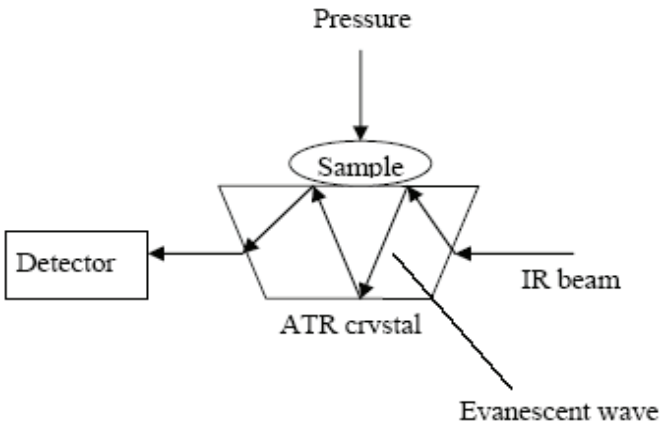

Figure 2.2. A schematic diagram of an ATR accessory (Stuart, 2004).

CHAPTER 3

MULTIVARIATE CALIBRATION METHODS

According to The International Chemometrics Society (ISC), chemometrics is the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods. Design of experiments, optimization of experimental parameters, signal processing, calibration is employed for principal component analysis, pattern recognition, collecting good data are used for obtaining information from these data in chemometrics.

3.1. Calibration Method

Calibration is a generated model which is used in order to determinate the relation between features of samples and instrumental response. Firstly, analyte's concentration levels and instrument responses are used to construct the model. Secondly, by using this model, the features are predicted according to the instrumental response of a sample. Terminally, the concentration of unknown samples is predicted by utilizing the model.

Generally, calibration methods are examined in two parts, which are multivariate and univariate calibration methods. One single wavelength is used to determine the a single compound's concentration of when the univariate calibration model is applied. In contrast to the univariate calibration, most of the wavelengths are used to determine the multi-component mixture's concentration when the multivariate calibration method is applied.

3.1.1. Univariate Calibration

When univariate calibration model is used, Lambert Beer's law can be applied to determine the correlation between instrument response and analyte concentration. If this correlation is linear, two options can be listed:

- Classical calibration
- Inverse calibration

3.1.1.1. Classical Univariate Calibration

This kind of calibration method uses a single wavelength or data point in a spectrum, and concentration is modeled by the absorbance. The general classical univariate calibration formula is:

$$a = c.\ s + e \tag{3.1}$$

a is the vector of absorbance at one wavelength for a number of samples and **c** is the vector of corresponding concentrations.

where the c′ shown as the transpose of the concentration vector and the scalar coefficients can be applied according to the formula as below:

$$s \approx (c' \cdot c)^{-1} \cdot c' \cdot a \tag{3.2}$$

After s is calculated by the formula on scalars a and c with hat mention as predictions, the concentration of an unknown can be predicted as:

$$\hat{c} \approx a / s \tag{3.3}$$

3.1.1.2 Inverse Univariate Calibration

Since unknown sample concentration is predicted by using instrumental response and instrumental performance can cause response error, classical calibration is not always preferred in analytical chemistry. Mostly errors are derived from the concentration, that is larger than instrumental error. Figure 3.1 represents the errors, where (a) is obtained from the instrument, and (b) is from concentration (Brereton 2003).

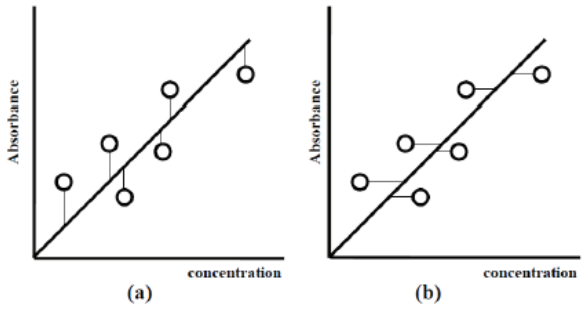

Figure 3.1. Difference between errors in (a) classical and (b) inverse calibration (Source: Brereton 2003).

The formula can express Inverse calibration method:

$$\mathbf{c} \approx \mathbf{a} \cdot b \tag{3.5}$$

b is approximately inverse of s, which is a scalar coefficient, and can be calculated as:

$$b \approx (\mathbf{a}' \cdot \mathbf{a})\text{-}1 \cdot \mathbf{a}' \cdot \mathbf{c} \tag{3.6}$$

the concentration of an unknown sample prediction is calculated by the following formula:

$$\hat{c} = \hat{a} \,.\, b \tag{3.7}$$

3.1.2. Multivariate Calibration

There is a need for improved quantitative information in science and technology. According to this purpose, multivariate calibration is a general selectivity and reliability enhancement tool. (Martens et al., 2001) This method is related to the multiple responses to properties of a sample process. The samples could be not only a single chemical component but also a mixture of chemicals components. Unlike univariate analysis, multivariate calibration makes possible fault-detection.

In this study, genetic inverse least squares (GILS) method is used. Before explaining this method, as an introduction to the multivariate calibration method, comparison between classical least squares (CLS) and inverse least squares (ILS) methods should be examined.

3.1.2.1 Classical Least Squares (CLS)

Classical least squares (CLS) method is also known as Beer's Law or K-matrix method. According to the model, the absorbance at every single wavelength is concentrations of an analyte's function. This method can be calculated by the equation below.

$$\mathbf{A} = \mathbf{C} \times \mathbf{K} + \mathbf{E} \tag{3.8}$$

where **A** is a m x n matrix that consists absorbance values of m calibration samples at n wavelengths of the samples at different wavelengths, **C** is the m x h matrix which contains of concentrations of each of the h components in the m calibration samples. **K** is the matrix of absorptivity coefficients multiplied by path length, and $\mathbf{E}_A$ is the matrix of residuals not fit by the model or spectral errors.

K matrix is evaluated by the least square with the equation.

$$K = (C' \cdot C)_{-1} \cdot C' \cdot A \tag{3.9}$$

The prediction of the concentrations of the unknown sample is calculated by the following equation:

$$\hat{c} = (K \cdot K') - 1 \cdot K' \cdot a \tag{3.10}$$

where **a** is the spectrum of unknown sample and **ĉ** is the vector of the predicted component concentrations.

The difference between the predicted and reference concentration values is called the residual and can be calculated by:

$$\mathbf{e = c - \hat{c}} \tag{3.11}$$

Classical Least Square method can provide significant enhancement in precision because it is a full spectrum method. Moreover, as a consequence of supplying spectral baselines that are randomly fitted and analyzed pure component spectra along with the residuals. In spite of all benefits, this method has one disadvantages that all meddling chemical components must be known and their concentrations incorporated into the model. Concentrations of all species are not possibly known in real life samples. There would be a large error since the instrument response of interfering species is not put in the calibration model. By using Inverse Least Square (ILS) method, this error can be reduced.

3.1.2.2. Inverse Least Squares (ILS)

In practice, all concentrations can not be analyzed at all. Therefore Classical Least Square method is not always applicable. In this case, the Inverse Least Squares (ILS) method that concentrations of an analyte are modeled as a function of absorbance is used. This method can be explained by the following equation.

$$\mathbf{C = AP + E_C} \tag{3.12}$$

where **C** and **A** are the same as in CLS (**C** is the concentration matrix, and **A** is the absorbance matrix). $\mathbf{E}_C$ is the *m x h* matrix of errors in the concentrations not fit by the model. **P** is the *n x h* matrix of the unknown calibration coefficients relating *h* component concentrations to the spectral intensities.

As you can see the below formulation, equation 3.12 can be reduced for the analysis of one component at a time. This reduction is the greatest advantage of ILS.

$$\mathbf{c} = \mathbf{Ap} + \mathbf{e_c} \tag{3.13}$$

where **c** is the *m x 1* vector of concentrations for the analyte that is being analyzed, **p** is *n x 1* vector of calibration coefficients, and **ec** is the *m x 1* vector of concentration residuals not fit by the model. During the calibration step, $\hat{\mathbf{p}}$ which is estimated **p** can be calculated with:

$$\hat{\mathbf{p}} = (\mathbf{A} \cdot \mathbf{A}')^{-1}\ \mathbf{A}' \cdot \mathbf{c} \tag{3.14}$$

After $\hat{\mathbf{p}}$ is calculated, the concentration of the analyte of the unknown can be predicted with the equation 3.15:

$$\hat{c} = \mathbf{a}' \cdot \hat{\mathbf{p}} \tag{3.15}$$

where $\hat{c}$ is the scalar estimated concentration and $\mathbf{a}'$ is the spectrum of the unknown sample.

An ILS demonstrate contains a major advantage in that it does not ought to know and incorporate all components within the calibration set. This implies that ILS accept that the power for each measured variable within the investigation all carry on perfectly independent. Furthermore, you're confined from utilizing all of the spectral channels in making the model. The number of channels of spectral data utilized cannot surpass the number of calibration benchmarks. Exactness will be decreased in the data that more channels are included than the number of independent sources of variety within the information. (Haaland et al.,1988)

3.1.2.3. Genetic Inverse Least Squares (GILS)

Genetic inverse least squares are mostly used application of Genetic Algorithms (GA) which are global optimization and search methods based upon the principles of natural evolution and selection developed by Darwin. It is utilized for selecting wavelengths to construct multivariate calibration models with reduced data set. (Özdemir et al.,2010)

Agreeing to Darwin's Hypothesis; variety could be a include of common population, and each population produces offspring. The results of this overproduction is that those people with the best genetic fitness for the environment. Thus the later generation will have a higher representation of these offspring, and the population will have evolved. (McClean et al.,1997)

Genetic Algorithm has five key steps, as shown in Figure 3.2.

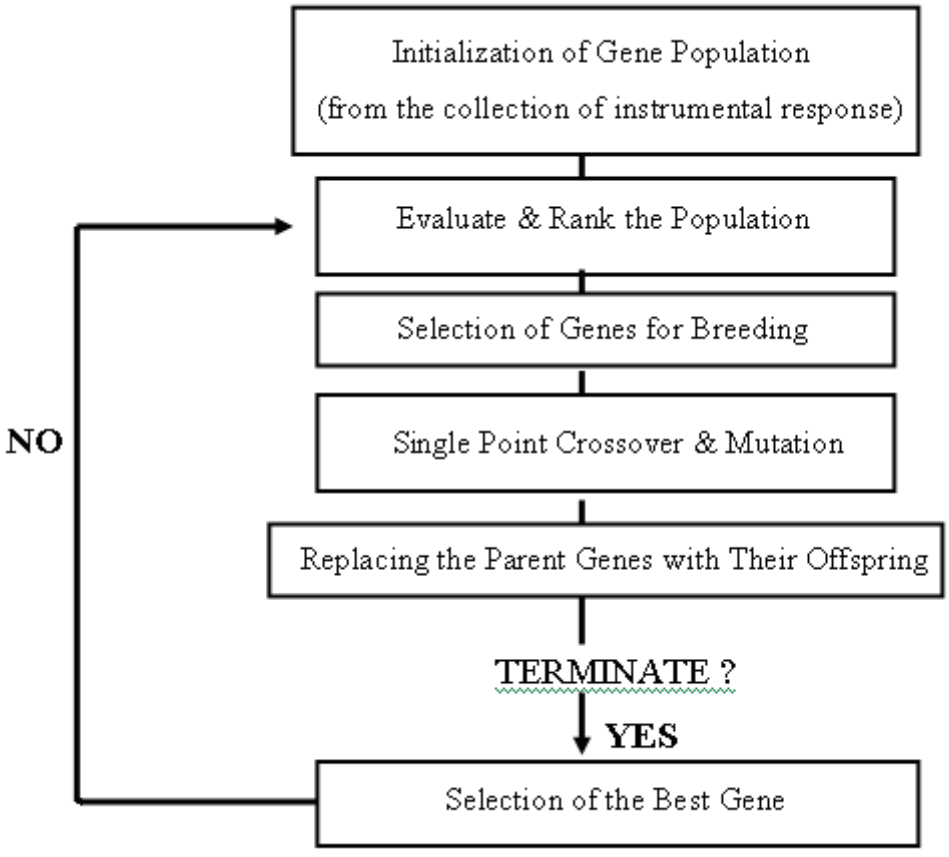

Figure 3.2. Flow chart of genetic algorithm used in GILS

These steps comprise of initialization of a gene population, evaluation of the population, selection of the parent genes for breading, and mating, crossover and replacing parents with their offspring, respectively. The title of these steps starts within the biological feature of the genetic algorithm.

3.1.2.3.1. Initialization

A gene which is shown as the collection of instrumental response at the determined wavelength range according to the data set is a possible solution of given problems. The population is the gathering of individual genes in the current generation.

A gene is shown in the following formula 3.16

$$\mathbf{S = [A_{1754}\ A_{926}\ A_{2268}\ A_{596}\ A_{1255}\ A_{3500}]} \tag{3.16}$$

where **S** is a gene and **A** is the absorbance.

In this step, first generation genes are produced with fixed population size. Genes are selected randomly, then bias is minimized, and the number of possible recombination are maximized. The estimating time depends on the size of the number of genes. If the population size is small, shorted estimating time is required. Therefore, it is important to maintain the number of the gene pool size.

3.1.2.3.2. Evaluate and Rank the Population

After the first step as gene population is produced, the next step would be to evaluate and rank the population the genes. In arrange to assess each each gene's success using a fitness function. The fitness function is formulated as the inverse of the standard error of calibration (SEC):

$$\text{Fitness} = 1/\text{SEC} \tag{3.17}$$

SEC is evaluated from the equation:

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

where c_i is the reference and $\hat{c}_i$ is the predicted values of concentration of i^{th} sample and m is the number of samples. Degree of freedom is m-2 that is the parameters to be derived, where the slope of the actual plot vs. reference concentration plot and the respective intercept.

3.1.2.3.3. Selection of Genes for Breeding

This step based on the fundamental principle of natural evolution. The parent genes are selected from the present population for breeding. The aim is to generate the genes, which are the best performing members of the population will survive in the long run. Their information will be passed to the next generation. Thus, better offspring will be generated with the genes which are better suited for the problems. The genes with high fitness values will be given higher chance to breed, and thus, several will be able to survive.

3.1.2.3.4. Crossover and Mutation

The genes are broken at points randomly. Offspring genes are formed by cross-coupling. S_1 and S_2 are parent genes; S_3 and S_4 are their corresponding off-springs.

Parents:

$S_1 = [A_{452}\ A_{3732}\ \#\ A_{1237}\ A_{2890}]$

$S_2 = [A_{923}\ A_{1457}\ A_{1743}\ \#\ A_{832}\ A_{3022}]$

Offspring:

$S_3 = [A_{452}\ A_{3732}\ A_{832}\ A_{3022}]$

$S_4 = [A_{923}\ A_{1457}\ A_{1743}\ A_{1237}\ A_{2890}]$

As shown above, the points where the genes are cut for mating are indicated by # and the place where crossover takes place. This process is called single point crossover, and it is commonly used in GILS.

3.1.2.3.5. Replacing the Parent Genes with Their Offspring

After cross over and mutation step, the offspring are evaluated by replacing with the parent genes. The modeled concentration of component is predicted in the validation step, and the success of that model is utilized using the standard error of prediction (SEP). standard error of prediction can be calculated with the formula:

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

where m is the number of independent validation samples.

3.1.2.3.6. Termination

The final step is the termination of the algorithm. It is done by setting a predefined iteration number for the number of breeding cycles. If the lowest SEC for the calibration set, means lowest SEC has the highest fitness. Then the lowest SEC is selected for model building; the concentrations of component are estimated by using this model. In the prediction of test sets, the model's success is approximated using the standard error of prediction (SEP).

CHAPTER 4

EXPERIMENTATION

In this study, quantitative composition determination of personal care and cleaning products which are produced by Dalan Kimya Endüstri A.Ş. was carried out by using FTIR Spectroscopy combined with a genetic algorithm based multivariate calibration method called Genetic Inverse Least Squares (GILS). Two different product categories are chosen namely liquid soap and shower gel.

4.1. Sample Preparation

In the first part of the study, quantitative composition determination of liquid soap which is one of the most sold products of Dalan Kimya Endüstri A.Ş, was carried out. Because of the confidentiality requirement, names of the components contained in the particular liquid soap are coded as surfactant 1, surfactant 2 etc. As shown in the Table 4.1 below, the liquid soap composed of 17 different components, including water which contributes more than 78% (w/w) of the overall product. Among the remaining other 16 components, some of them were found around 2% (w/w), some of them around 0.5 % (w/w) but some coloring components like the dyes, are the minor components of the product as their concentrations even lower than 0.0001 % (w/w). Four different surfactants are used where two of them are found to be around 2 % where as one of them 11% (w/w) and the other one around 0.5 % (w/w %). In addition to these components, there is also 2% (w/w) salt which is added in order to modify the rheology of the system. Moreover, there are some other minor ingredients like chelating agent, preservative, perfumes and so on.

Table 4.1. The composition of liquid soap along with coded names of the components and their concentrations (w/w %).

Name of the Component	Code of the Component	(w/w %)
Water	A	78.6891
Surfactant 1	B	11.00
Salt	C	2.10
Surfactant 2	D	2.00
Surfactant 3	E	2.00
Additive 1	F	2.00
Surfactant 4	G	0.50
Additive 2	H	0.50
pH adjuster	I	0.46
Parfum	J	0.25
Preservative	K	0.20
Oil	L	0.15
Chelating agent	M	0.10
UV absorber	N	0.05
Dye 1	O	0.00070
Dye 2	P	0.00017
Dye 3	R	0.00005

As the number of components which forms the product are 17, where some of them are the major and some of them minor components, it is not practical (If not possible) to determine all of these components quantitatively not only by FTIR spectroscopy but also with some other conventional instrumental methods such as chromatography. Therefore, a screening experimental design study was carried out in order to determine which of these factors (components) has the most significant importance in the number of different chemical and physical properties of the product stability such as pH, viscosity, density, color, odor, and texture. The screening design chosen here was 19-factor standard Placket Burman Design in which two dummy factors were not used so that only 17 components has been taken from Table 4.1 in order to construct design shown in Table 4.2.

Table 4.2.Concentration of the samples for the 17 components for Placket Burman design (w/w %)

No	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	R	Total
1	79.73	9.90	1.89	2.20	1.80	2.20	0.45	0.55	0.51	0.28	0.22	0.14	0.09	0.06	0.000775	0.0001541	0.000052	100
2	77.48	12.10	1.89	2.20	1.80	2.20	0.55	0.55	0.51	0.23	0.18	0.17	0.11	0.05	0.000775	0.0001883	0.000043	100
3	77.51	12.10	2.31	2.20	1.80	1.80	0.55	0.55	0.41	0.28	0.22	0.14	0.09	0.05	0.000634	0.0001883	0.000043	100
4	79.32	9.90	2.31	1.80	2.20	2.20	0.55	0.55	0.41	0.23	0.22	0.17	0.09	0.06	0.000775	0.0001541	0.000043	100
5	77.14	12.10	2.31	2.20	2.20	1.80	0.45	0.55	0.51	0.23	0.22	0.17	0.09	0.05	0.000634	0.0001541	0.000052	100
6	80.11	9.90	1.89	1.80	2.20	1.80	0.55	0.45	0.51	0.28	0.22	0.17	0.09	0.05	0.000775	0.0001883	0.000043	100
7	79.46	9.90	2.31	2.20	1.80	2.20	0.55	0.45	0.41	0.23	0.18	0.17	0.09	0.06	0.000634	0.0001883	0.000052	100
8	77.99	12.10	2.31	1.80	1.80	1.80	0.45	0.55	0.41	0.28	0.18	0.17	0.11	0.06	0.000775	0.0001541	0.000043	100
9	80.18	9.90	1.89	1.80	1.80	2.20	0.45	0.55	0.41	0.28	0.22	0.17	0.11	0.05	0.000634	0.0001883	0.000052	100
10	77.56	12.10	2.31	1.80	1.80	2.20	0.55	0.45	0.51	0.28	0.18	0.14	0.09	0.05	0.000775	0.0001541	0.000052	100
11	79.63	9.90	2.31	1.80	2.20	1.80	0.55	0.55	0.51	0.28	0.18	0.14	0.11	0.06	0.000634	0.0001883	0.000052	100
12	78.33	12.10	1.89	1.80	1.80	1.80	0.55	0.45	0.51	0.23	0.22	0.17	0.11	0.06	0.000634	0.0001541	0.000052	100
13	79.84	9.90	2.31	2.20	1.80	1.80	0.45	0.45	0.51	0.23	0.22	0.14	0.11	0.06	0.000775	0.0001883	0.000043	100
14	79.80	9.90	1.89	2.20	2.20	1.80	0.55	0.55	0.41	0.23	0.18	0.14	0.11	0.05	0.000775	0.0001541	0.000052	100
15	77.62	12.10	1.89	1.80	2.20	2.20	0.45	0.55	0.51	0.23	0.18	0.14	0.09	0.06	0.000634	0.0001883	0.000043	100
16	77.34	12.10	2.31	1.80	2.20	2.20	0.45	0.45	0.41	0.23	0.22	0.14	0.11	0.05	0.000775	0.0001883	0.000052	100
17	79.01	9.90	2.31	2.20	2.20	2.20	0.45	0.45	0.51	0.28	0.18	0.17	0.11	0.05	0.000634	0.0001541	0.000043	100
18	77.73	12.10	1.89	2.20	2.20	1.80	0.45	0.45	0.41	0.28	0.18	0.17	0.09	0.06	0.000775	0.0001883	0.000052	100
19	77.20	12.10	1.89	2.20	2.20	2.20	0.55	0.45	0.41	0.28	0.22	0.14	0.11	0.06	0.000634	0.0001541	0.000043	100
20	80.82	9.90	1.89	1.80	1.80	1.80	0.45	0.45	0.41	0.23	0.18	0.14	0.09	0.05	0.000634	0.0001541	0.000043	100

As shown in Table 4.2., concentration of each component has been varied around ±10% of the product formula amount except water which is amount is depend on the other 16 component as the total amount of the each sample must add up to 100%. This may cause slight disturbance to Placket Burman Design where the water composition changes slightly from sample to sample which makes it difficult to code as ±1 in the coded design table. However, considering more than 78% of the total composition is formed by water in all samples, it is assumed that the disturbance should not have a significant effect on the general Placket Burman Design analysis. Table 4.3. shows the coded Placket Burman design table for a linear model including an intercept term.

Table 4.3. Coded Placket Burman design table

No	bo	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	R
1	1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1
2	1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1
3	1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1
4	1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1
5	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1
6	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1
7	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1
8	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1
9	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1
10	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1
11	1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1
12	1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1
13	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1
14	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1
15	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1
16	1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1
17	1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1
18	1	1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1
19	1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	-1
20	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

After evaluating the statistical design of experiment study with 17 factors Placket Burman screening design, only the 7 components (factors) which are A, C, D, E, F, J and L were chosen as to be varied their concentrations in multivariate calibration and G and K components were used as either presence and not available in the samples. The remaining 8 components were kept constant in product composition at the level given in Table 4.1. In order to better understand how the concentrations of the components were changed in the 50 multicomponent samples for multivariate calibration modelling, only the first two samples compositions are given in Table 4.4. with the coded component names. The concentration table of the all of the 50 samples is shown in Table 4.5.

Table 4.4. Components, codes and their concentrations ranges (w/w %) used in Sample 1 and Sample 2 of 'Liquid Soap'.

Code	Sample 1 (w/w %)	Sample 2 (w/w %)	Condition
A*	**77.1837**	**82.0530**	**Change**
B	0.1000	0.1000	Constant
C*	**8.6740**	**6.7048**	**Change**
D*	**4.9772**	**3.3876**	**Change**
E*	**2.9018**	**2.1392**	**Change**
F*	**1.0922**	**0.9471**	**Change**
G	0.0000	0.0500	Presence / NA
H	0.1500	0.1500	Constant
I	2.0000	2.0000	Constant
J*	**0.5261**	**0.2263**	**Change**
K	0.0000	0.2000	Presence / NA
L*	**1.6440**	**1.2910**	**Change**
M	0.2500	0.2500	Constant
N	0.5000	0.5000	Constant
O	0.0007000	0.0007000	Constant
P	0.0000477	0.0000477	Constant
R	0.0001700	0.0001700	Constant

*These are the components which were varied in their concentrations.

Table 4.5. Concentration profile of 50 liquid soap samples as mass percent (w/w %) with coded names of the components that are varied during multivariate calibration.

No	A	C	D	E	F	J	L	No	A	C	D	E	F	J	L
1	77.18	8.67	4.98	2.90	1.09	0.53	1.64	**26**	80.53	8.13	2.23	3.25	0.84	0.34	1.44
2	82.05	6.70	3.39	2.14	0.95	0.23	1.29	**27**	77.17	12.83	3.12	2.13	0.27	0.31	1.11
3	76.99	10.18	2.16	3.70	1.10	0.56	2.07	**28**	75.79	11.65	3.80	3.05	0.49	0.34	1.63
4	74.46	12.29	2.25	4.67	1.39	0.58	1.10	**29**	73.67	12.94	4.16	3.18	1.34	0.24	1.23
5	75.37	12.19	2.23	4.19	0.38	0.54	2.11	**30**	81.03	7.79	2.14	2.07	1.23	0.40	2.13
6	72.82	12.71	3.37	4.95	0.59	0.42	1.88	**31**	73.88	12.66	4.57	2.84	0.34	0.36	2.09
7	79.43	6.10	3.53	4.13	1.32	0.21	2.03	**32**	79.03	6.37	4.87	3.07	1.47	0.53	1.61
8	81.37	8.85	1.66	2.19	0.79	0.50	1.39	**33**	74.45	10.94	3.34	4.96	1.44	0.30	1.32
9	78.10	12.04	2.63	1.83	0.88	0.22	1.04	**34**	72.19	11.72	4.74	4.65	0.99	0.32	2.14
10	79.84	6.97	3.29	4.31	0.36	0.50	1.72	**35**	74.30	12.81	2.43	4.48	0.98	0.49	1.33
11	79.46	7.72	2.37	3.88	1.12	0.21	2.00	**36**	77.63	9.26	3.82	3.72	0.85	0.38	1.09
12	78.37	6.32	3.32	4.84	1.41	0.58	1.91	**37**	76.58	8.10	3.84	4.47	1.29	0.47	2.20
13	82.05	6.23	3.69	2.17	0.59	0.50	1.52	**38**	76.89	11.25	2.42	3.54	0.51	0.28	1.85
14	78.65	7.15	4.96	3.15	0.34	0.41	2.09	**39**	78.88	8.46	3.30	3.28	0.90	0.21	1.72
15	80.98	7.54	1.60	3.59	1.27	0.26	1.76	**40**	77.90	11.43	3.29	1.65	0.91	0.51	1.12
16	82.17	6.12	3.21	1.59	1.05	0.46	2.15	**41**	79.98	6.52	2.21	4.71	0.72	0.38	2.23
17	76.99	8.00	4.90	4.22	0.26	0.48	1.91	**42**	79.32	7.39	4.35	2.02	1.08	0.45	2.34
18	77.81	8.40	3.75	3.46	1.43	0.30	1.59	**43**	81.48	6.45	2.31	3.36	0.90	0.52	1.73
19	78.38	9.88	3.65	2.02	1.38	0.21	1.24	**44**	76.57	8.51	4.25	4.13	1.00	0.43	1.86
20	78.19	8.50	2.13	4.64	0.81	0.40	2.33	**45**	77.45	9.41	3.19	3.94	0.88	0.59	1.35
21	78.34	8.60	2.39	3.67	1.17	0.32	2.27	**46**	78.15	9.58	3.09	2.96	0.85	0.42	1.70
22	78.94	8.49	3.50	3.36	0.41	0.30	1.76	**47**	76.85	8.61	3.17	4.84	1.02	0.22	2.24
23	74.20	12.37	4.72	1.87	1.17	0.54	1.88	**48**	74.03	12.90	4.11	3.68	0.52	0.38	1.13
24	77.67	9.26	4.04	2.58	0.87	0.50	1.82	**49**	81.05	6.28	4.86	2.07	0.68	0.57	1.24
25	76.24	8.98	4.85	4.45	0.49	0.21	1.57	**50**	79.87	7.62	3.76	2.37	1.35	0.50	1.34

In order to develop multivariate calibration models based on FTIR spectra of the samples, the calibration and independent validation sets were perapared from the samples given in Table 4.5. in a ramdom manner where 34 samples were chosen as calibration set and the remaining 16 samples were chosen as the independent validation set.

In the second phase of the multivariate calibration study, a shower gel product is selected in order to analyze composition. The shower gel components, codes, and their percentages are shown in Table 4.6.

Table 4.6. Components, codes and their concentrations (w/w %) used in training set of 'Shower Gel'

Name of the Compound	Code of the Component	(w/w %)
Dye 1	A	0.000233
Dye 2	B	0.000040
Dye 3	C	0.000720
Solubilizer	D	2.0000
Salt*	**E**	**1.2000**
Surfactant 1*	**F**	**16.000**
Oil	G	0.1000
Ph adjuster	H	0.2300
Surfactant 2*	**I**	**1.0000**
Surfactant 3*	**J**	**3.0000**
Preservative	K	0.2000
UV light stabilizer	L	0.0500
Parfum	M	0.7000
Surfactant 4*	**N**	**0.4000**
Surfactant 5*	**O**	**2.0000**
Surfactant 6*	**P**	**7.0000**
EDTA	R	0.1000
Water*	**S**	**66.019**
	Total	100

*These are the components which were varied in their concentrations.

As mentioned above, the particular shower gel composed of 18 different components, including water which forms 66% of the product. Among the remaining other 17 components, one of them is found around 16%, a number of them around 1 to 3% but some coloring components like the dyes, are minor component of the product as their concentrations even lower than 0.0001% by mass. Five different surfactants are used as their concentrations around 16%, 7%, 3%, 2% and 1% by mass. In addition to them, 1.2% salt is added in order to modify the rheology of the system. Moreover, there are some other ingredients like a chelating agent, preservative, perfumes and so on. A total of 30 different synthetic mixtures were prepared shown in Table 4.7, where the concentrations of 8 components were ranged around ±10% of the standard formula of the product.

Table 4.7. Concentration profile of 30 shower gel samples as mass percent (w/w %) with coded names of the components that are varied during multivariate calibration.

No	S	F	P	O	J	N	I	E
1	66.44	14.39	6.71	2.13	3.06	0.39	1.28	1.31
2	65.78	15.92	6.74	1.83	2.81	0.41	0.99	1.14
3	65.43	14.51	7.90	1.91	3.28	0.38	0.92	1.30
4	65.10	15.32	7.19	2.13	3.37	0.37	0.99	1.20
5	63.04	17.50	6.87	2.08	3.12	0.41	1.26	1.26
6	65.70	15.38	6.84	2.26	2.83	0.37	1.09	1.15
7	66.26	15.00	6.73	1.85	3.02	0.43	1.10	1.26
8	65.86	14.56	7.50	1.86	3.21	0.40	1.01	1.26
9	64.87	16.04	6.68	2.22	3.11	0.37	1.05	1.25
10	66.34	14.69	6.36	2.02	2.88	0.45	0.98	1.29
11	65.33	15.99	6.28	2.22	3.15	0.42	1.02	1.24
12	64.96	15.23	7.55	2.12	3.32	0.34	0.96	1.20
13	62.83	18.32	6.82	2.03	2.79	0.41	1.26	1.27
14	64.44	17.02	6.66	2.05	2.76	0.46	1.03	1.17
15	65.55	15.29	7.25	2.13	2.65	0.34	1.28	1.22
16	65.91	14.64	7.41	1.95	3.17	0.37	0.90	1.39
17	64.04	17.09	7.09	2.10	2.80	0.41	1.08	1.26
18	63.23	17.15	7.34	2.16	2.78	0.34	1.25	1.31
19	66.38	15.03	7.44	1.80	2.66	0.36	0.92	1.16
20	63.54	17.38	6.96	2.15	3.06	0.41	1.08	1.11
21	65.61	15.17	7.07	2.03	2.86	0.43	1.17	1.31
22	65.67	14.60	7.10	2.06	3.35	0.43	1.16	1.20
23	65.72	14.65	7.14	1.91	3.44	0.42	1.13	1.22
24	65.39	16.08	6.77	1.84	2.95	0.41	1.11	1.10
25	63.14	16.86	7.74	2.05	3.18	0.40	1.09	1.20
26	65.65	15.10	7.05	2.12	3.16	0.41	0.96	1.21
27	64.44	16.33	7.39	1.94	2.75	0.44	1.16	1.32
28	63.04	17.48	7.41	1.98	3.07	0.43	1.05	1.18
29	64.93	14.39	7.85	2.23	3.24	0.38	1.27	1.19
30	64.26	16.18	7.33	2.20	2.90	0.33	1.10	1.22

The following percent concentration graphs are drawn in order to make the above table more understandable. Figure 4.1 shows the concentration changes for the water component of the second phase of the study.

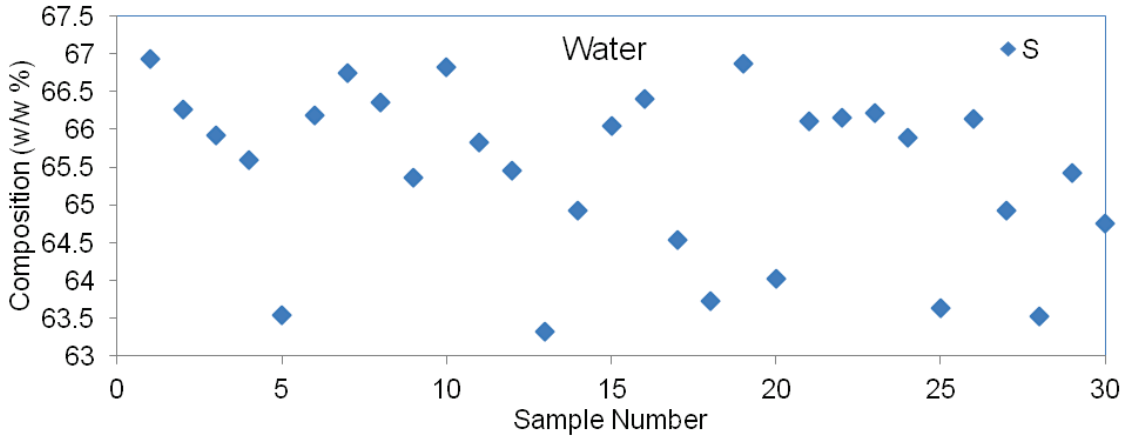

Figure 4.1. Percent concentration of the water contents (w/w %) of the samples in the second phase of the study.

As seen from the Figure 4.1, the concentration of water was changed from 63% to 67% (w/w %). Figure 4.2 shows the concentration changes for the O, N, I and E components of the shower gel in the second phase of the study.

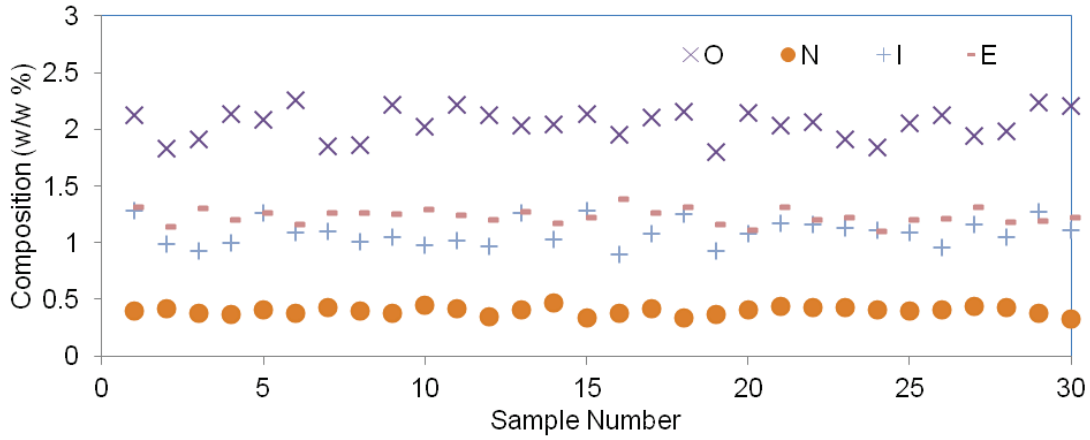

Figure 4.2. Percent concentration of the O, N, I and E contents (w/w %) of the samples in the second phase of the study.

As seen from the Figure 4.2, the concentration of O was changed from 1.8 % to 2.2% (w/w %), N from 0.3 % to 0.5 % (w/w %), I from 0.9 to 1.3 % and E from 1.1 % to 1.4 %. Figure 4.3 shows the concentration changes for the F, P, and stock solution components of the second phase of the study.

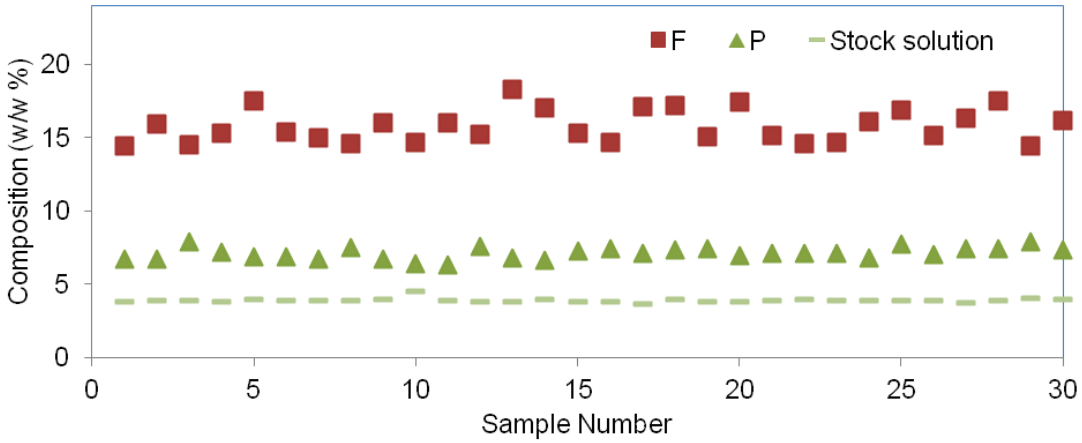

Figure 4.3. Percent concentration of the F, P, and stock solution contents (w/w %) of the samples in the second phase of the study.

As seen from the Figure 4.3, the concentration of F was changed from 14.4 % to 18.3 % (w/w %), P from 6.3 % to 7.9 % (w/w %) and stock solution around 3.8 % (w/w %). As shown in Figure 4.1 above, the amount of water in the samples varies between 63 - 67.5% whereas the F component changes in the range of 13-17% and the P component is changed in the range of approximately 4.5-5.5%. On the other hand, the other components of the mixture O, N, I and E varied in the concentration range ±10% in the product. Finally, the amount of the rest of the components (A + B + C + D + G + H + K + L + M + R) are kept constant at 3.8% by mass.

4.2. Instrumentation and Data Processing

4.2.1. Data Collection

In this study, FTIR spectral data were collected by Perkin Elmer Spectrum 100 spectrometer equipment with universal three reflection ATR accessory between 4000 – 650 cm^{-1}. In order to generate multivariate calibration models for each component, FTIR spectra are recorded as log (1/R) against to not only air but also water background. The spectra are saved as ASCII file format and then transferred to another PC after collected on the FTIR instrument.

4.2.2. Data Processing

Calibration and validation sets are prepared as text files with the aid of Microsoft Excel (MS Office 2010, Microsoft Corporation) program. After the text files are organized by using MS Excel, they are used to build chemometrics multivariate calibration models with GILS method which is implemented by MATLAB R2013a (MathWorks Inc., Natick, MA). These models were used to predict the compositions of liquid soap and shower gel samples whose concentrations were unknown with the models generated in liquid soap study an shower gel study, respectively.

CHAPTER 5

RESULTS AND DISCUSSION

As mentioned in previous chapters, in this study, FTIR spectroscopy coupled with three reflection diamond attenuated total reflectance (ATR) accessory was used to collect spectral data which have been used to develop multivariate calibration models for two personal care and cleaning products namely liquid soap and shower gel. The following section gives the results obtained from these studies starting with the results of liquid soap.

5.1. Results of Liquid Soap

5.1.1. Placket Burman Statistical Experimental Design Results

As mentioned in Chapter 4, Placket Burman design has been carried out in order to determine which of the chemical components (factors) have significant importance in the number of different chemical and physical properties of the product stability such as pH, viscosity, density, color, odor, and texture. For this, a 19 factors standard Placket Burman screening design given in Table 4.2 and 4.3 has been prepared where 2 of the factors was assigned as dummy factor and not included in the tables. Therefore, a total of 17 chemical components which forms the standard formula composition of the liquid soap are included in the design table. After generating the design table composed of 20 independent experiments, 2 kg of liquid soap samples had been produced for the each formulation given in Table 4.2 and viscosity, pH, density and foam tests had been performed on each sample. Additionally, foam tests had been done at five different periods of time. Results of these analyses are shown in Table 5.1. Regression analysis of Placket Burman screening design had been carried out with a linear model equation containing 17 terms given in Table 4.3 along with an intercept term based on coded data for each response variables given in Table 5.1. The regression coefficients and p-values from analysis of variance (ANOVA) and regression analysis are shown in Table 5.2.

Table 5.1. Viscosity, density, pH and foam test results of Liquid Soap samples given in Table 4.2.

				Foam Test (cm)				
No	**Viscosity (cP)**	**pH**	**Density (g/cm3)**	**0 minute**	**5 minute**	**10 minute**	**15 minute**	**20 minute**
1	2660	6.07	1.022	1090	1080	1060	1050	1040
2	4480	5.27	1.037	1180	1160	1140	1130	1130
3	9750	6.11	1.039	1160	1140	1130	1120	1100
4	5580	5.07	1.040	1180	1170	1150	1130	1120
5	9850	6.20	1.038	1180	1160	1140	1130	1120
6	2000	6.54	1.020	1170	1160	1140	1120	1100
7	9870	5.32	1.039	1170	1160	1140	1130	1100
8	4860	5.63	1.038	1080	1060	1060	1040	1020
9	1100	5.57	1.020	1100	1040	1040	1040	1020
10	8720	5.37	1.036	1120	1120	1100	1080	1080
11	8460	6.95	1.039	1120	1100	1100	1080	1080
12	3490	6.16	1.031	1080	1080	1070	1060	1050
13	3610	6.77	1.032	1150	1140	1120	1100	1080
14	5190	6.05	1.035	1150	1120	1100	1100	1080
15	5090	6.07	1.034	1160	1150	1140	1120	1100
16	8680	5.90	1.037	1090	1080	1070	1060	1060
17	6490	5.62	1.035	1170	1160	1150	1120	1100
18	6900	6.53	1.035	1150	1140	1120	1100	1100
19	10350	5.81	1.039	1100	1080	1070	1060	1040
20	4500	5.89	1.032	1130	1120	1100	1100	1080

Table 5.2. Regression coefficients of Placket Burman design for the linear model along with p-values.

	Viscosity (cP)		pH		Density (g/cm^3)		Foam test (0 min)		Foam test (5 min)		Foam test (10 min)		Foam test (15 min)		Foam test (20 min)	
Terms	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
b0	6081.5	**0.001**	5.945	**0.000**	1.0339	**0.000**	1136.500	**0.000**	1121.000	**0.000**	1107.0	**0.000**	1093.5	**0.000**	1080.0	**0.000**
A	-489.5	0.127	-0.013	0.270	-0.0002	0.836	5.500	0.243	5.000	0.130	5.0	0.087	3.5	0.145	3.0	0.641
B	1135.5	**0.028**	-0.040	**0.043**	0.0025	0.099	-6.500	0.192	-4.000	0.184	-3.0	0.198	-3.5	0.145	0.0	1.000
C	1505.5	**0.016**	-0.051	**0.027**	0.0034	**0.057**	5.500	0.243	8.000	**0.057**	9.0	**0.030**	5.5	0.067	6.0	0.391
D	833.5	**0.050**	0.030	0.073	0.0012	0.294	13.500	**0.057**	13.000	**0.023**	10.0	**0.024**	10.5	**0.020**	9.0	0.245
E	777.5	**0.057**	0.129	**0.004**	0.0013	0.266	10.500	0.089	11.000	**0.032**	11.0	**0.020**	8.5	**0.030**	10.0	0.212
F	220.5	0.373	-0.338	**0.001**	0.0000	1.000	-0.500	0.895	-1.000	0.667	-1.0	0.592	-1.5	0.423	-1.0	0.873
G	707.5	0.068	-0.080	**0.011**	0.0016	0.201	6.500	0.192	8.000	**0.057**	7.0	**0.047**	7.5	**0.038**	8.0	0.284
H	-379.5	0.189	-0.046	**0.033**	0.0003	0.758	3.500	0.406	-3.000	0.272	-1.0	0.592	0.5	0.771	1.0	0.873
I	-596.5	0.091	0.157	**0.003**	-0.0015	0.220	5.500	0.243	10.000	**0.038**	9.0	**0.030**	5.5	0.067	8.0	0.284
J	47.5	0.829	0.075	**0.013**	-0.0016	0.201	-10.500	0.089	-13.000	**0.023**	-10.0	**0.024**	-12.5	**0.014**	-12.0	0.162
K	-374.5	0.193	0.075	**0.013**	-0.0021	0.132	-6.500	0.192	-8.000	**0.057**	-8.0	**0.037**	-6.5	**0.049**	-7.0	0.333
L	-619.5	0.085	-0.154	**0.003**	-0.0006	0.554	9.500	0.105	8.000	**0.057**	8.0	**0.037**	6.5	**0.049**	6.0	0.391
M	-410.5	0.168	0.028	0.083	0.0004	0.685	-14.500	**0.050**	-19.000	**0.011**	-15.0	**0.011**	-14.5	**0.011**	-14.0	0.127
N	5.5	0.980	0.093	**0.008**	0.0010	0.361	-8.500	0.127	-5.000	0.130	-4.0	0.127	-6.5	**0.049**	-7.0	0.333
O	-813.5	**0.052**	-0.025	0.101	-0.0007	0.497	-0.500	0.895	2.000	0.423	-1.0	0.592	-2.5	0.238	1.0	0.873
P	-87.5	0.696	0.158	**0.003**	-0.0007	0.497	8.500	0.127	6.000	0.095	7.0	**0.047**	6.5	**0.049**	7.0	0.333
R	410.5	0.168	0.067	**0.016**	-0.0007	0.497	-11.500	0.076	-13.000	**0.023**	-13.0	**0.014**	-10.5	**0.020**	-7.0	0.333

As seen from Table 5.2., many terms in the linear model equation are not significant at 95% confidence level for most of the responses except for pH and foam test at 5, 10 and 15 minutes. Those terms, which are significant at given confidence level, are highlighted with bold red font where 14 terms including intercept were significant for pH. For the foam test, there were 13 terms that are significant for 10 minute whereas for 5 and 15 minutes the number of significant terms were around 12. Actual versus predicted plots of Placket Burman design analysis are given in Figure 5.1 for the response variables given in Table 5.1.

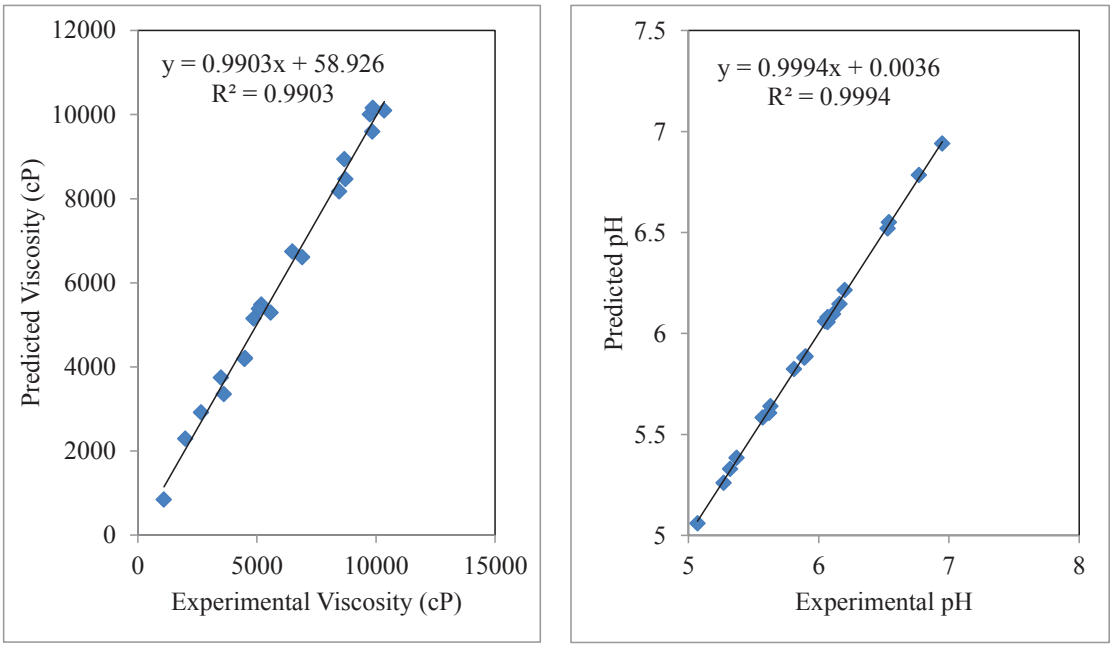

Figure 5.1. Actual vs. predicted plots of viscosity, pH, density and foam test (0, 5, 10, 15 and 20 min.) results from Placket Burman design. (cont. on next page)

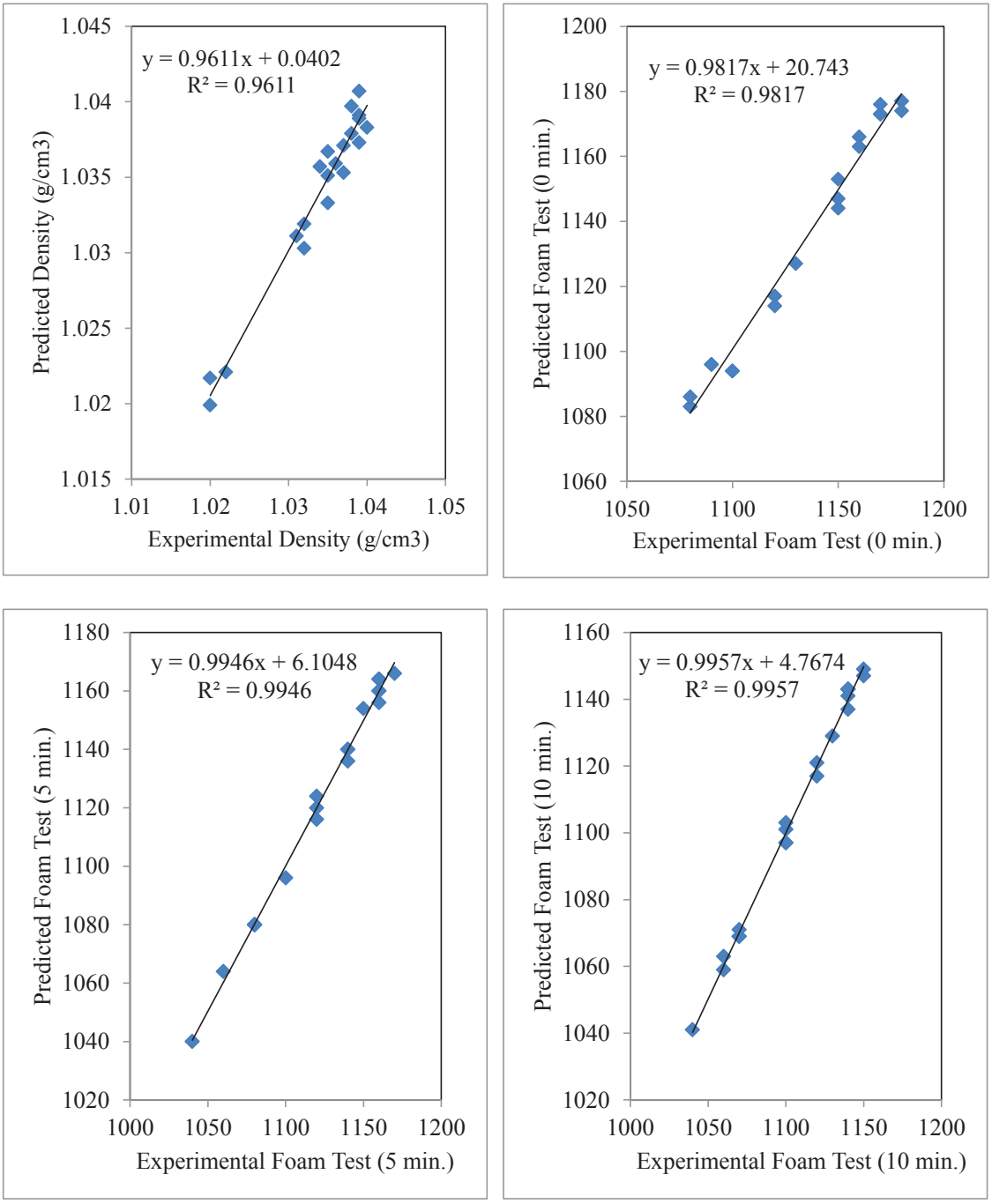

Figure 5.1. Actual vs. predicted plots of viscosity, pH, density and foam test (0, 5, 10, 15 and 20 min.) results from Placket Burman design. (cont. on next page)

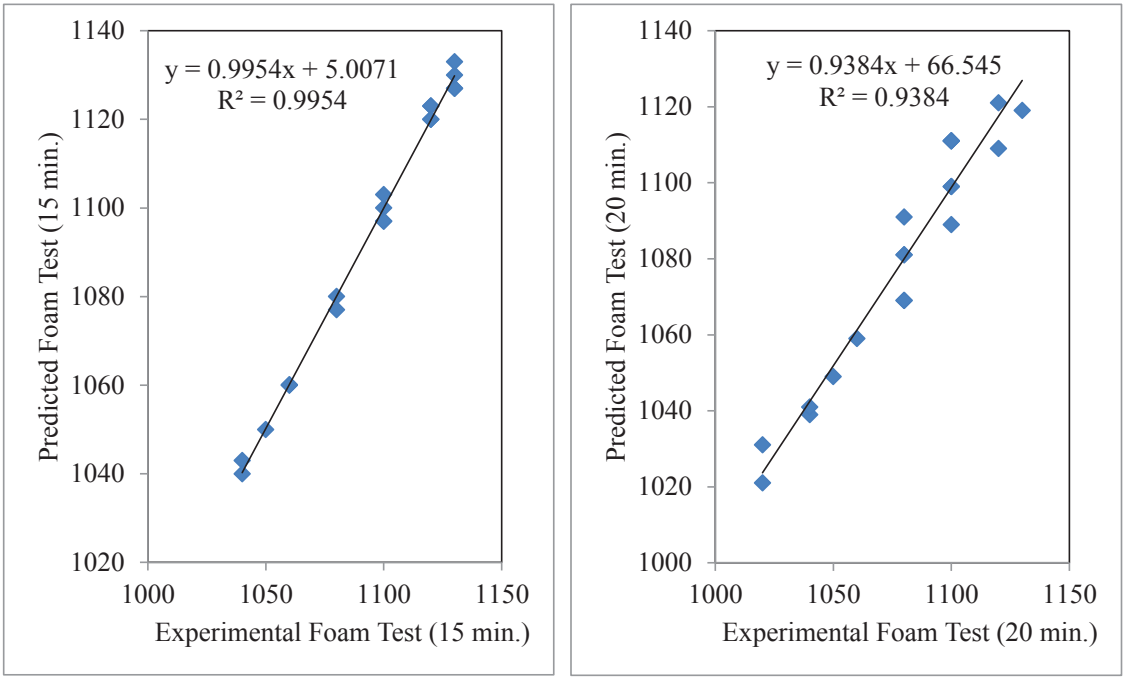

Figure 5.1. Actual vs. predicted plots of viscosity, pH, density and foam test (0, 5, 10, 10, 15 and 20 min.) results from Placket Burman design (cont.).

As can be seen from Figure 5.1, the highest regression coefficients are found in viscosity, pH and foam test parameters at 5, 10 and 15 minutes whereas somewhat lower R^2 values are observed for viscosity and foam test at 20 minutes. Figure 5.2 shows the residual plots of these response variables.

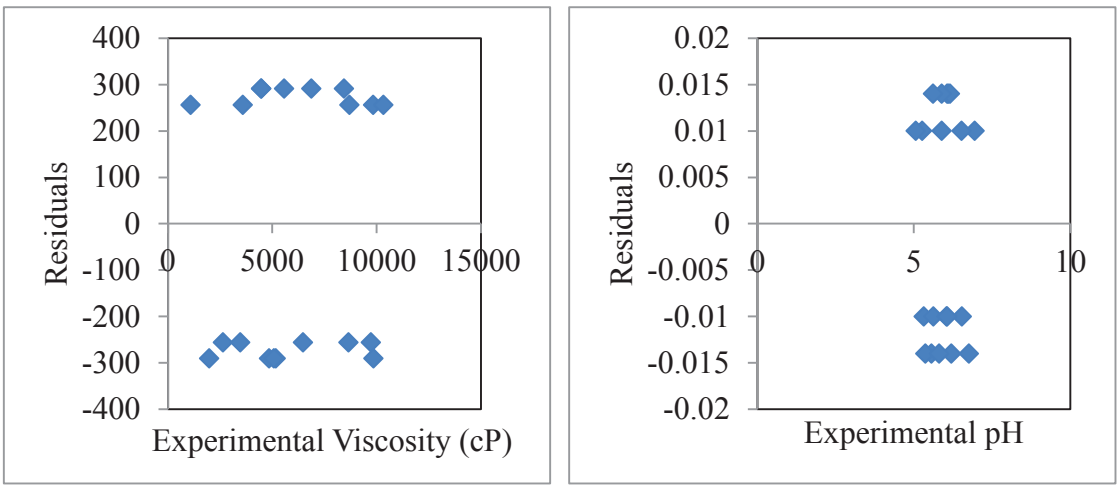

Figure 5.2. Residual plots of the response variables (cont. on next page).

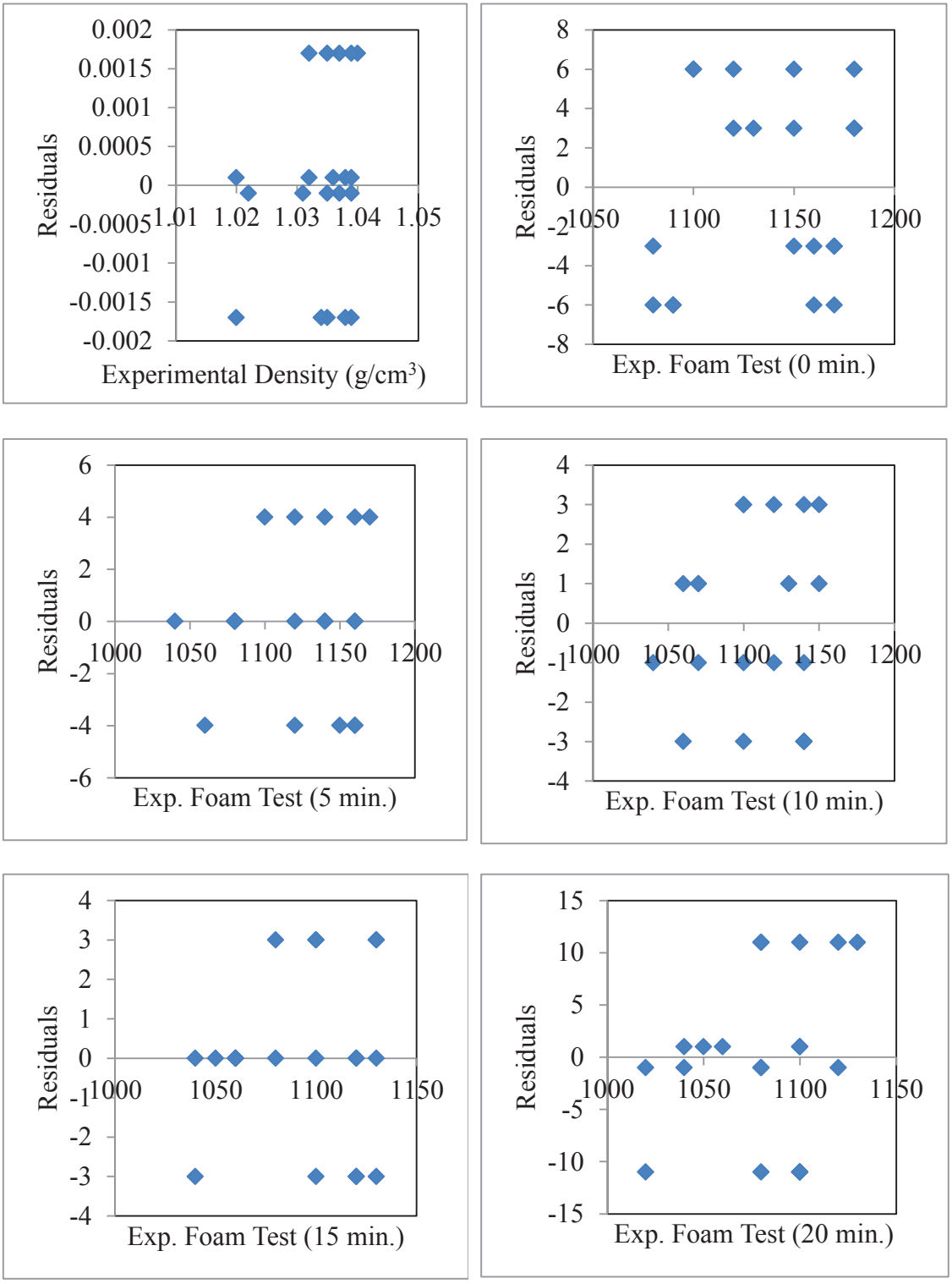

Figure 5.2. Residual plots of the response variables (cont.)

A close examination of the residual plots given in Figure 5.2. indicates that the residuals resulting from Placket Burman design analysis for all of the response variables are demonstrating normal distribution and no significant indication of any trend in the residual plots.

As a result of this Placket Burman screening experimental design study, it was decided to investigate the possibilities of measuring the component composition of the liquid soap product by using Fourier transform infrared spectroscopy (FTIR) combined with multivariate calibration. Among the 17 components studied, 7 components (factors) which are A, C, D, E, F, J and L were chosen as important components that are aimed to be modelled with multivariate calibration based on FTIR spectral data. In addition, the components G and K were used as either included as the standard formula amount or not presence in the samples. The remaining 8 components were kept constant in product composition at the level given in Table 4.1.

5.1.2. FTIR-ATR Results

The ATR – FTIR spectra of 11 components of the Liquid Soap studied here are shown in Figure 5.3. against water background in order to demonstrate the spectral differences and/or similarities among these major components. As can be seen in the collected FTIR spectra of the liquid soap samples, certain similarities are possible and clearly seen in these spectra of eleven components of 'liquid soap'. However, remarkable differences are also observed in the wavelength range of 1800-600 cm^{-1} which is called fingerprint region. The names of the components shown in Figure 5.3. are surfactant 1, surfactant 2, surfactant 3, additive 1, surfactant 4, additive 2, pH adjuster, perfume, preservative, oil and UV absorber.

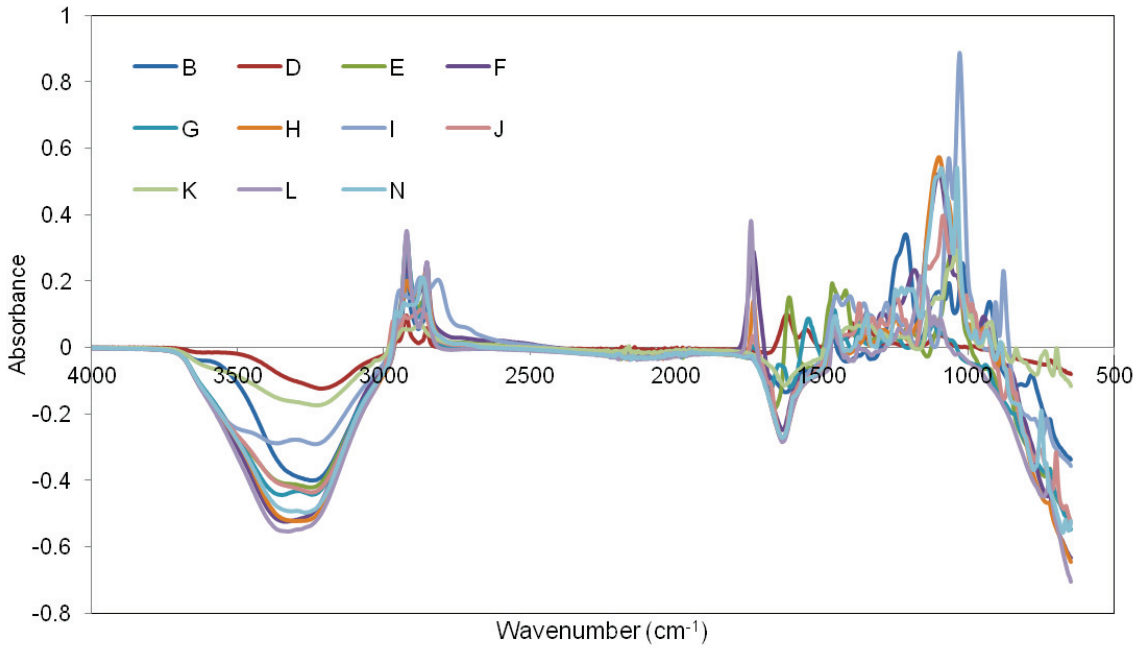

Figure 5.3. The spectra of 11 different compounds that are B, D, E, F, G, H, I, J, K, L and N collected by using an FTIR spectrometer coupled with ATR accessory against water background.

In these spectra, significant spectral differences are apparent among the 11 components of the liquid soap selected for this study. As can be seen in the collected spectra, there are negative peaks for almost all components from 3600 to 2800 cm^{-1} region since these components are already contains some amount of water in their stock solutions. Through the multivariate calibration, it is expected that these differences could provide sufficient differentiating power for the successful models at least some of the major components if not all of them selected for quantitative determination. In Figure 5.4, a total of 50 FTIR spectra collected against pure water background are shown for the samples given in Table 4.5. The reason for water back ground is that water accounts approximately 80% of liquid soap used in this study. Also, FTIR spectra of the same samples collected against air back ground are illustrated in Figure 5.4. As can be seen here, the spectra have very broad water peaks leading into other components' peaks making them almost invisible. Therefore, multivariate calibration models are generated with the spectra against water back ground.

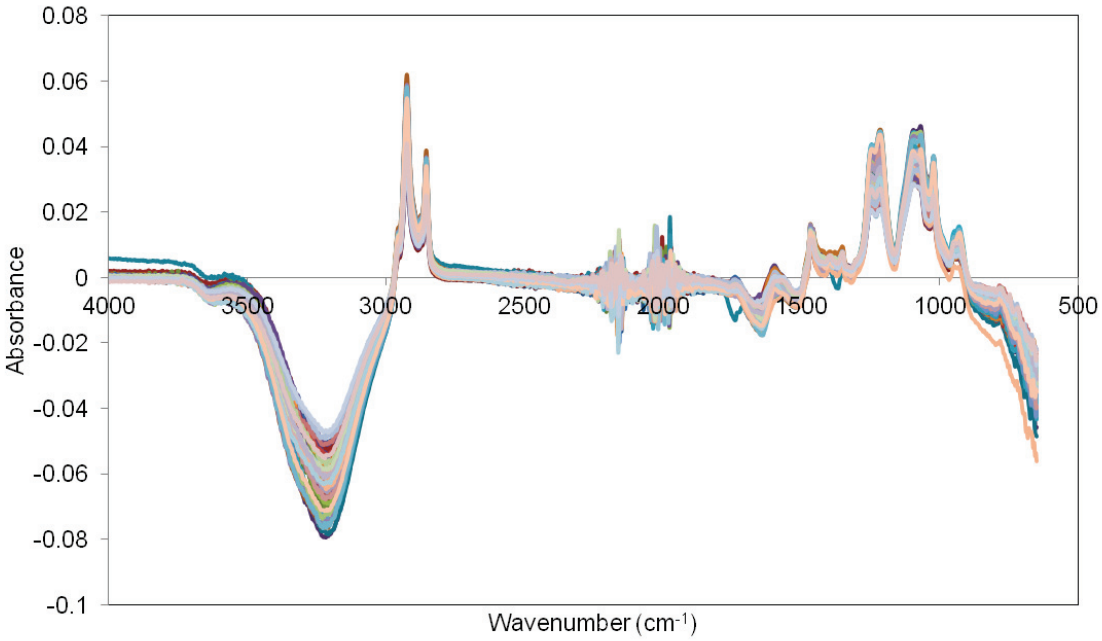

Figure 5.4. The spectra of 50 different liquid soap sample collected by using FTIR spectrometer coupled with ATR accessory by taking water as the background.

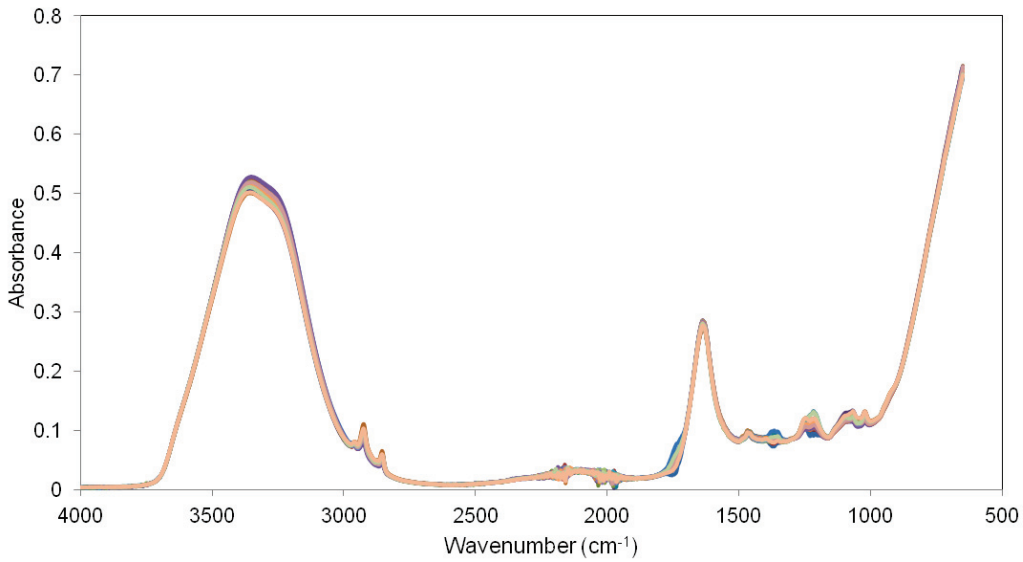

Figure 5.5. The spectra of 50 different liquid soap sample collected by using FTIR spectrometer coupled with ATR accessory by taking air as the background.

5.1.3. Multivariate Calibration Results of Liquid Soap with Genetic Inverse Least Squares (GILS)

GILS method is based on an iterative algorithm which allows to select the best combination of the wavelengths or wavenumbers which correlates most for the selected component in model building step of multivariate calibration using ILS approach. Due to the iterative nature of the algorithm, it is possible that the model can easily be over-fitted for the samples in the calibration set but fails for the predictions of the samples in the independent validation set. In order to avoid such over fitting problems, the algorithm is set to run with a leave one out cross validation method for calibration set. The fact that the variable selection is done in a complete random manner in GILS, the algorithm produces a different solution in each run, so the algorithm is set to run with a predefined number of gene which is 30 in this study and with 100 iteration number in each run. In order to enhance the averaging effect of the algorithm the program was also set to run 250 times with above settings and the average of the best solutions are used to evaluate the success of the model for calibration set and independent validation set. These settings for GILS were kept same for all the models generated in this thesis study not only for the liquid soap but also for the shower gel study.

The performance of the models was determined based on the standard error of cross validation (SECV) and standard error of prediction (SEP) for the calibration and independent validation sets, respectively. However, it is also good practice to examine the regression coefficients for models obtained from actual vs predicted plots of the selected mixture components. As mentioned in experimental section, among the 50 synthetic liquid soap samples, 34 randomly selected samples were used as calibration set and the remaining 16 samples were reserved for independent validation set. Figure 5.6. shows the actual vs predicted plots of the liquid soap components as A, C, D, E, F, J and L obtained from averaged GILS models.

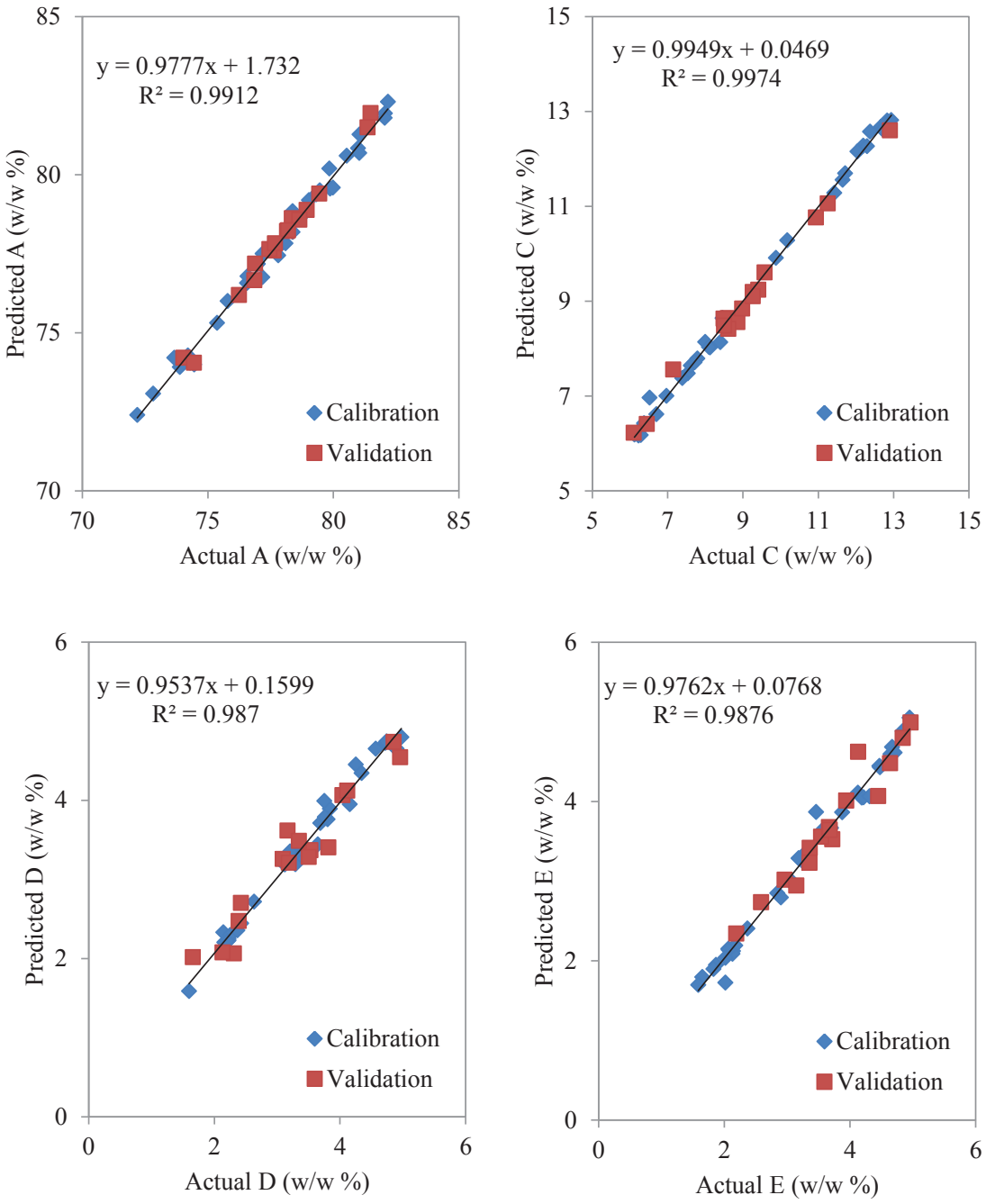

Figure 5.6. Actual versus predicted plots of A, C, D, E, F, J and L contents of liquid soap resulting from GILS models in the first phase of the study.

(cont. on next page)

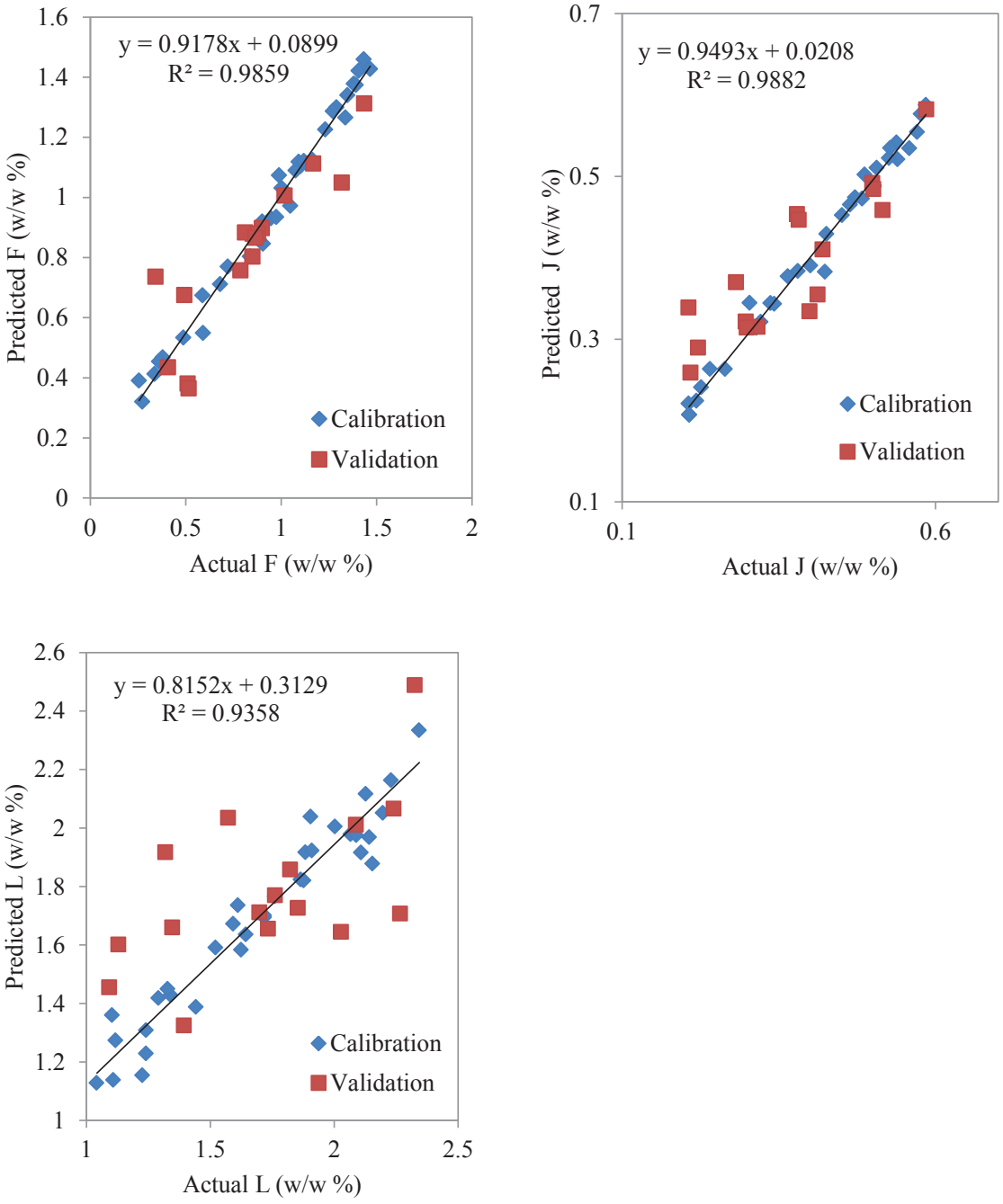

Figure 5.6. Actual versus predicted plots of A, C, D, E, F, J and L contents of liquid soap resulting from GILS models in the first phase of the study (cont.).

As can be seen in Figure 5.6., the highest regression coefficients (R^2) are obtained for the components A and C for both calibration sets and independent validation sets implying that the predictive performance for these components are more successful than the others. Moreover, relatively good model performances were also obtained for

components D and E in which their concentrations ranged approximately from 1.5% to 5.0% (w/w %) resulting in 3.5% dynamic range. On the other hand, the models for the components F and G look like somewhat over-fitted as the prediction results of the independent validation set are much more scattered when compared to the calibration set even though cross validation is used in modelling step. Finally, the model for the component L was the weakest one not only for calibration set but also for the independent validation set. However, it must be noted that the concentrations of these components were quite low compared to other four components. Due to these low concentrations of the components F, J and L modelling performances are relatively worse compare to other components.

Among the calibration plots that are shown in Figure 5.6., the regression coefficients in descending order are: C, A, J, E, D, F and L. On Table 5.3, standard errors of cross validation (SECV) and standard error of prediction (SEP) values along with minimum and maximum ranges of components are shown. While evaluating the magnitude of the SECV and SEP values of the models for each component, the width of the dynamic range of the calibration set and the magnitude of the concentration values must be compared.

Table 5.3. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum values of the components (Max and Min) and regression coefficient (R^2) of GILS models for liquid soap study.

	SECV (w/w %)	**SEP (w/w %)**	**Min (w/w %)**	**Max (w/w %)**	**R^2**
A	0.278	0.211	72.187	82.171	0.991
C	0.128	0.188	6.102	12.939	0.997
D	0.118	0.246	1.597	4.977	0.987
E	0.123	0.189	1.585	4.965	0.988
F	0.084	0.143	0.256	1.467	0.956
J	0.005	0.058	0.206	0.586	0.988
L	0.111	0.314	1.041	2.343	0.936

As can be seen from Table 5.3., agreement between SECV and SEP values of the models A, C, D and E were quite good but for the component J, the SEP value is almost 4 times larger than SECV. On the other hand, the R^2 values for the calibration sets were all around 0.99 except the component F and L where R^2 for F was 0.96 and it was 0.94 for L. As mentioned in experimental section, there were seven components whose concentration values were changed and the two of them (G and K) were either present or absent in the 50 multicomponent liquid soap samples. Therefore no model were built for these two components as it is impossible to generate a model from just two concentration values in which one of them already zero (absent).

The major success of those variations basically come from the ingredient of the product components. In this context, FTIR spectral data can be easily integrated into product components. On the other hand, the components that have a less weight of average on product cannot succeed enough on validation prediction even they have a limited calibration achievement. These results indicate that water, salt, surfactant 2 and surfactant 3 models are slightly better than additive 1 and parfum models.

5.1.4. Multivariate Calibration Results of Commercial Liquid Soap with Genetic Inverse Least Squares (GILS)

The models generated for water (A), salt (C), surfactant 2 (D), surfactant 3 (E), additive 2 (E) and parfum (J) were also tested which consist of 15 commercial liquid soap samples. The predicted concentrations for 15 different brands of liquid soap are given in Table 5.4. The fact that the model performance was the worst for component L, no results for this component were given here for the commercial liquid soap samples.

Table 5.4. Code of brands and their predicted concentrations (w/w %) according to the liquid soap model

Code of Brands	A	C	D	E	J	Total
Brand 1	78.69	11.00	2.00	2.00	0.46	94.15
Brand 2	85.11	8.58	1.18	1.40	0.35	96.62
Brand 3	83.44	9.37	2.31	1.10	0.34	96.56
Brand 4	84.68	9.47	0.78	1.19	0.28	96.41
Brand 5	86.26	7.38	1.29	1.18	0.28	96.40
Brand 6	92.20	3.88	-0.08	-0.22	0.27	96.04
Brand 7	91.56	3.96	0.01	-0.32	0.26	95.47
Brand 8	90.94	4.77	-0.09	0.35	0.21	96.18
Brand 9	91.88	4.85	-0.45	0.31	0.23	96.81
Brand 10	89.67	7.10	-0.28	0.08	0.25	96.82
Brand 11	81.30	10.27	4.49	0.26	0.61	96.92
Brand 12	82.46	9.35	5.83	-0.90	0.71	97.45
Brand 13	87.65	7.09	1.54	0.21	0.36	96.84
Brand 14	86.13	7.66	1.73	1.19	0.33	97.04
Brand 15	86.07	7.74	1.76	1.12	0.29	96.99

As it can be seen from the Table 5.4., A total of 15 different commercial liquid soap samples sol on the market have been tested with the models generated in this study. Predicted concantrations of water (A), salt (C) and parfum (J) showed that all the commercial liquid soap samples had these components with different amounts but the model generated with GILS was unable to detect surfactant 2 (D) in brands 6, 8, 9 and 10. In addition, the surfactant 3 (E) is not observed in brands 6, 7 brand 12.

5.2. Shower Gel Results

5.2.1. FTIR - ATR Results

The FTIR spectra recorded by 3 reflection diamond ATR are given in Figure 5.7. The spectra are recorder with air background for the 17 components except the water which is the last component of the 18 components that form the formulation of the shower gel product in Table 4.7.

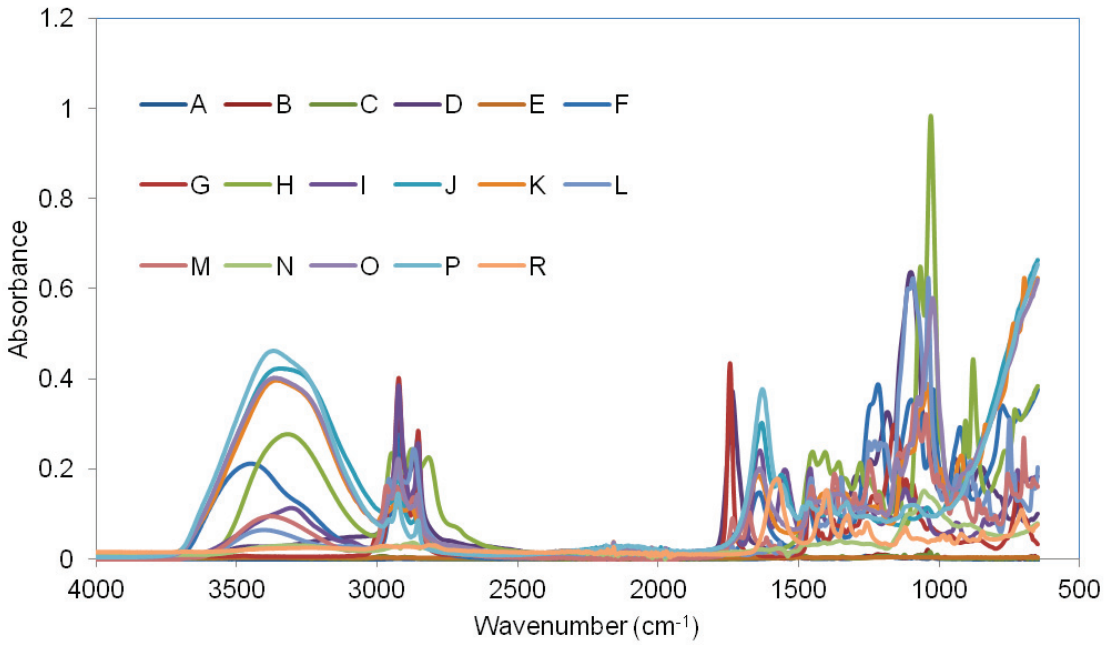

Figure 5.7. The FTIR spectra of 17 different compounds that are A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P and R collected by using three reflection diamond ATR accessory against air background.

As can be seen in the collected FTIR spectra of the shower gel samples, certain similarities are possible and clearly seen in these spectra of seventeen components of shower gel. As it is observed for liquid soap, remarkable differences are seen in the wavelength range of 1800-600 cm^{-1} which is called fingerprint region. The names of the components shown in Figure 5.7. are surfactant 1, surfactant 2, surfactant 3, surfactant 4, surfactant 5, surfactant 6, pH adjuster, perfume, preservative, oil, EDTA, dye 1, dye 2,

dye 3, solubilizer, salt and UV absorber. In order to give a better view of the spectral features, Figure 5.8. shows the FTIR spectra of the components E, F, I, J, N, O and P which are used in multivariate calibration modelling.

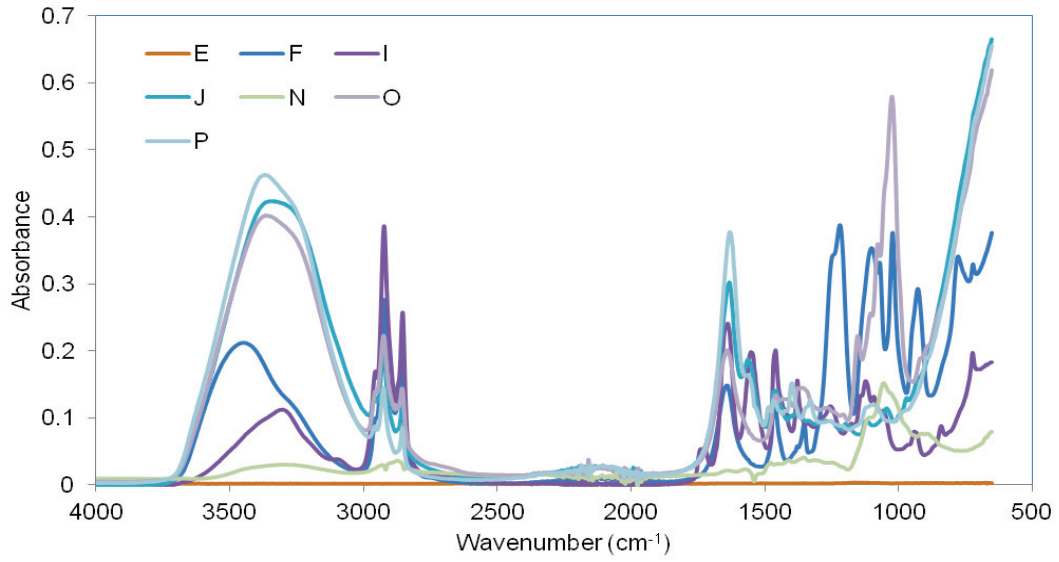

Figure 5.8. The spectra of 7 different compounds that are E, F, I, J, N, O and P collected by using three reflection diamond ATR accessory against air background.

The component labelled as E is the salt therefore the orange color line has almost no spectral feature as from Figure 5.8. However, the other components are all have strong peaks both in fingerprint region and 3500 – 2800 cm^{-1} region as most of the other components are used as their water solutions. To be able to better interpret fingerprint region of the FTIR spectra, Figure 5.9. shows an enlarged view of the spectra from 1800 – 600 cm^{-1} region. On the other hand, Figure 5.10. shows the FTIR spectrum of component E as in its pure form against air background.

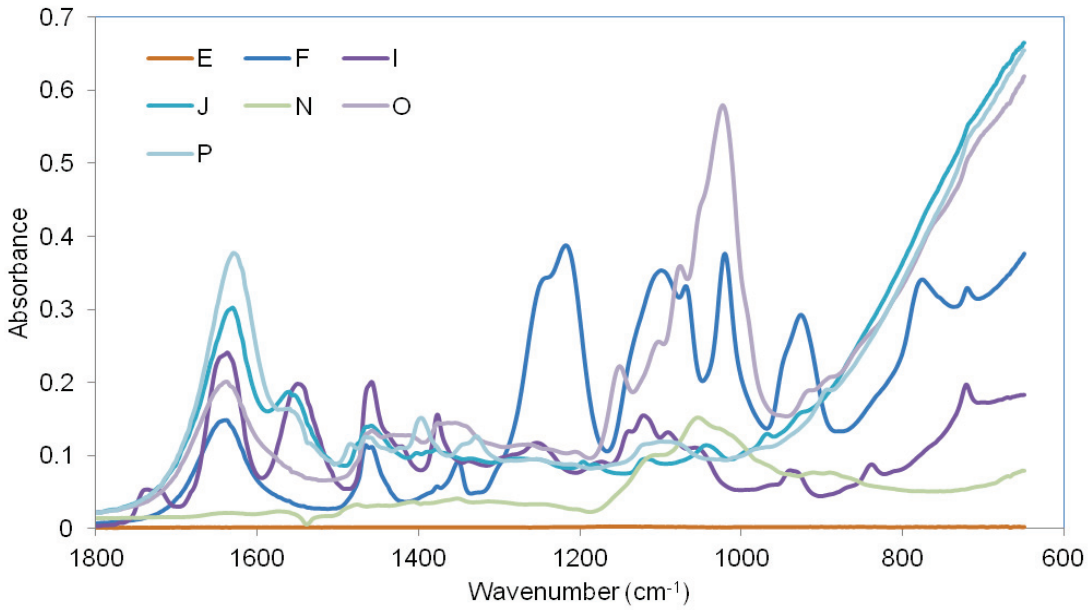

Figure 5.9. Fingerprint region of FTIR spectra between 1800 and 600 cm^{-1} wavenumber range of 7 different compounds that are E, F, I, J, N, O and P collected by using three reflection diamond ATR accessory against air background.

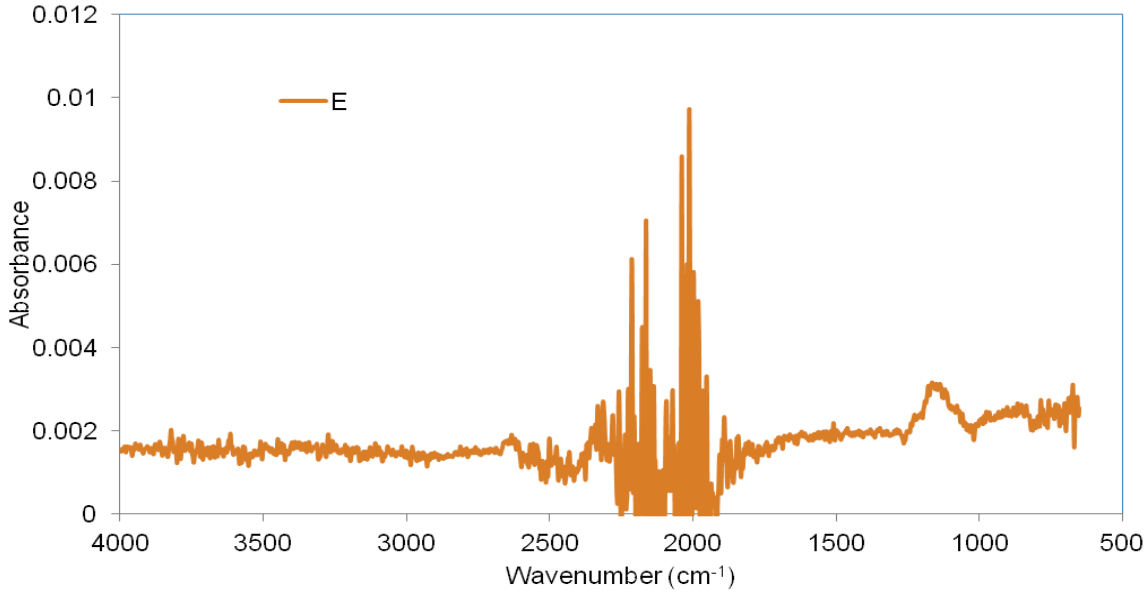

Figure 5.10. The FTIR spectra of component E (salt) collected by using three reflection diamond ATR accessory against air background.

As it is explained in detail in Chapter 2, FTIR spectroscopy has been used to characterize functional groups, bonding types and nature of compounds. Despite having a complex chemical structure for shower gel, it is obvious that all components have somewhat similar peaks around 3600-2800 cm^{-1} region except the component N. The reason behind this is the water contents of the ingredients which have various amount of water whereas the component N has the least amount of water. When the fingerprint area is considered, it is clear that there are some distinguishing features of the components.

Shower gel includes approximately 65 % water, therefore in Figure 5.11 shows the FTIR spectra collected against pure water background for the 30 samples given in Table 4.7. In addition, the same 30 samples FTIR spectra collected against air background are given in Figure 5.12. As it is known from experience in liquid soap analysis, the spectra have very broad water peaks and other components are nearly invisible. Because of this, multivariate calibration models are generated with the spectra that were collected against water background.

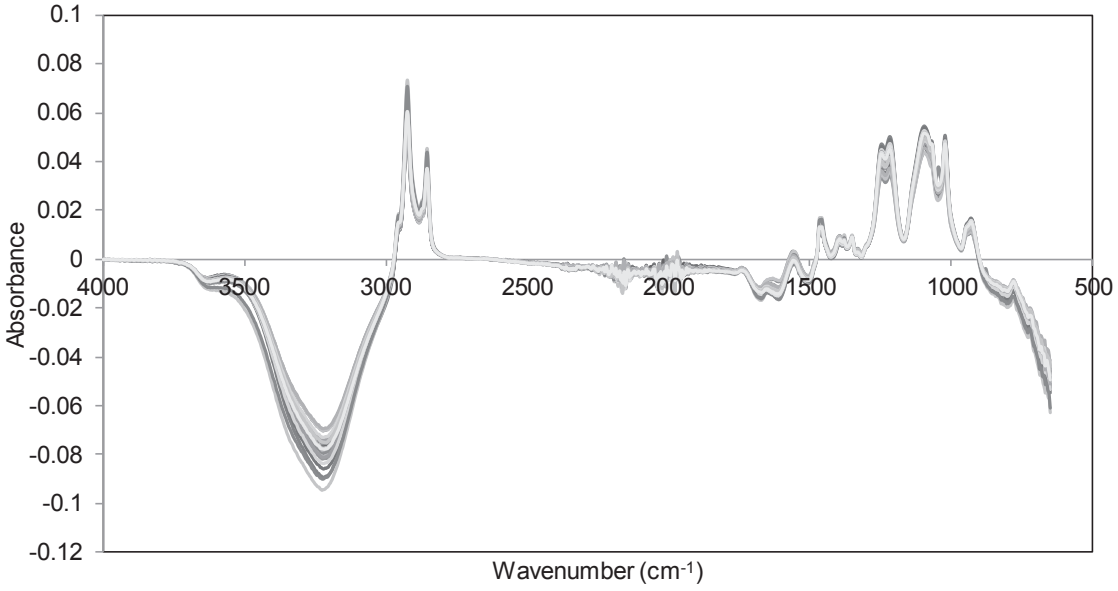

Figure 5.11. The FTIR spectra of 30 samples collected by using three reflection diamond ATR accessory against water background.

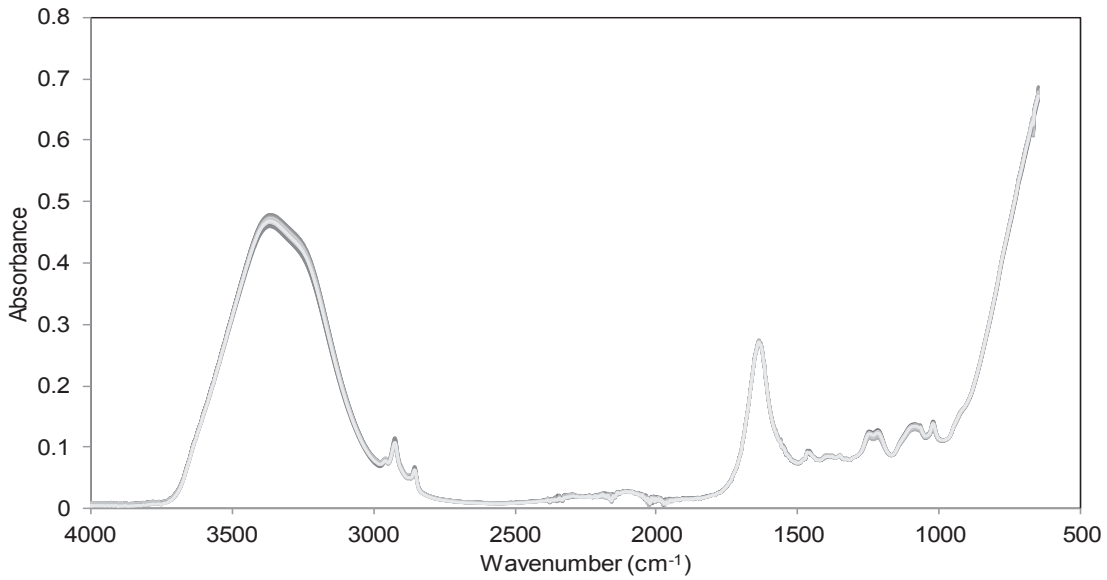

Figure 5.12. The FTIR spectra of 30 samples collected by using three reflection diamond ATR accessory against air background

.

5.2.2. Multivariate Calibration Results of Shower Gel with Genetic Inverse Least Squares (GILS)

As mentioned in Multivariate Calibration Results of Liquid Soap with Genetic Inverse Least Squares section, GILS method is used in order to generate models for the selected shower gel compounds. The regression coefficients for models which are obtained from actual vs predicted plots of the shower gel components is considered in order to determine the performance of the models. Moreover, based on the standard error of cross validation (SECV) and standard error of prediction (SEP) for the calibration and validation sets, the performance of the models was concluded. As given in experimental section, among the 30 synthetic shower gel samples, 23 randomly selected samples were used as calibration set and the remaining 7 samples were reserved for independent validation set. Figure 5.13. shows the actual vs predicted plots of the shower gel components as E, F, I, J, N, O, P and S obtained from averaged GILS models.

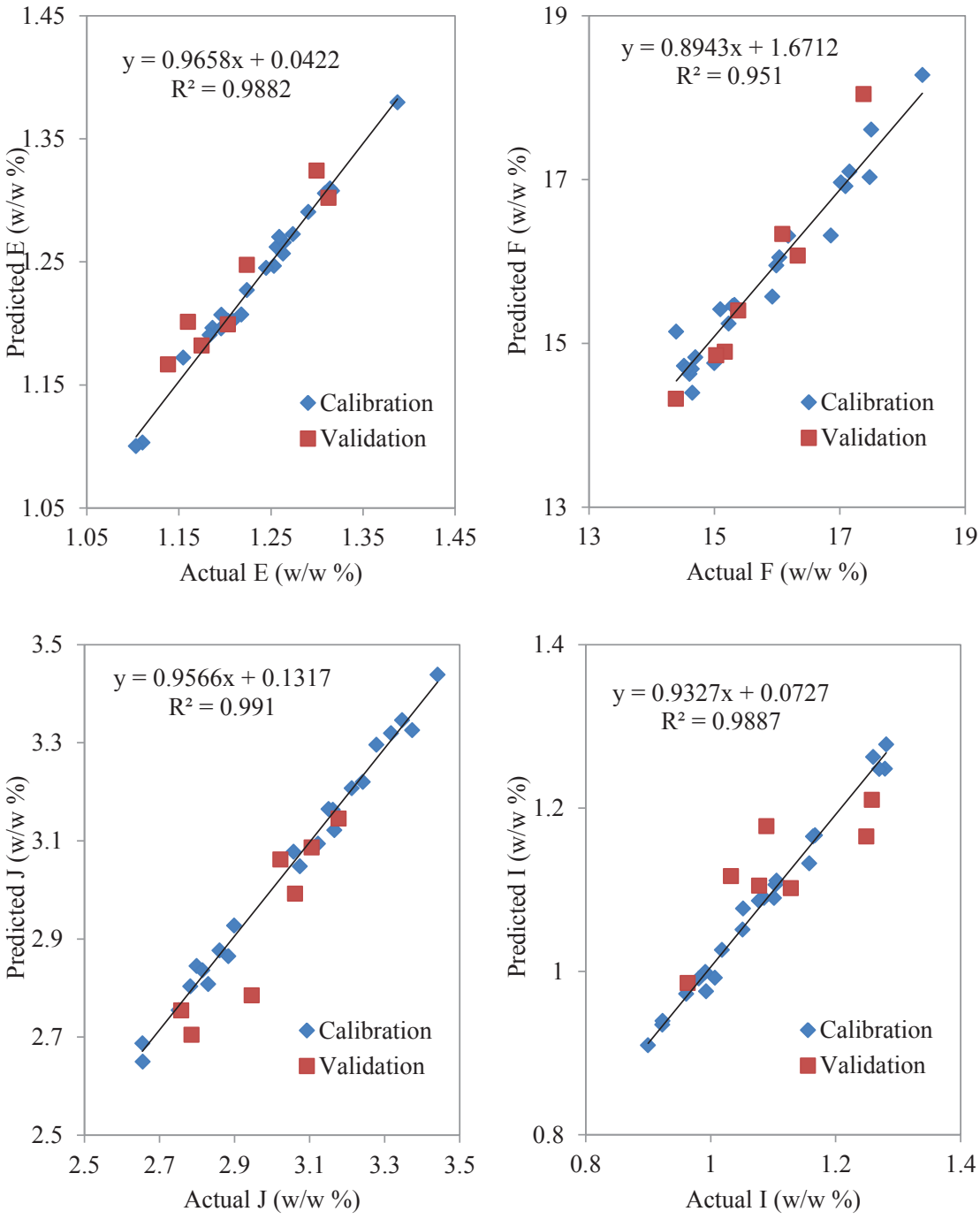

Figure 5.13. Actual versus predicted plot of E, F ,I, J, N, O, P and S contents resulting from GILS in the second phase of the study.

(cont. on next page)

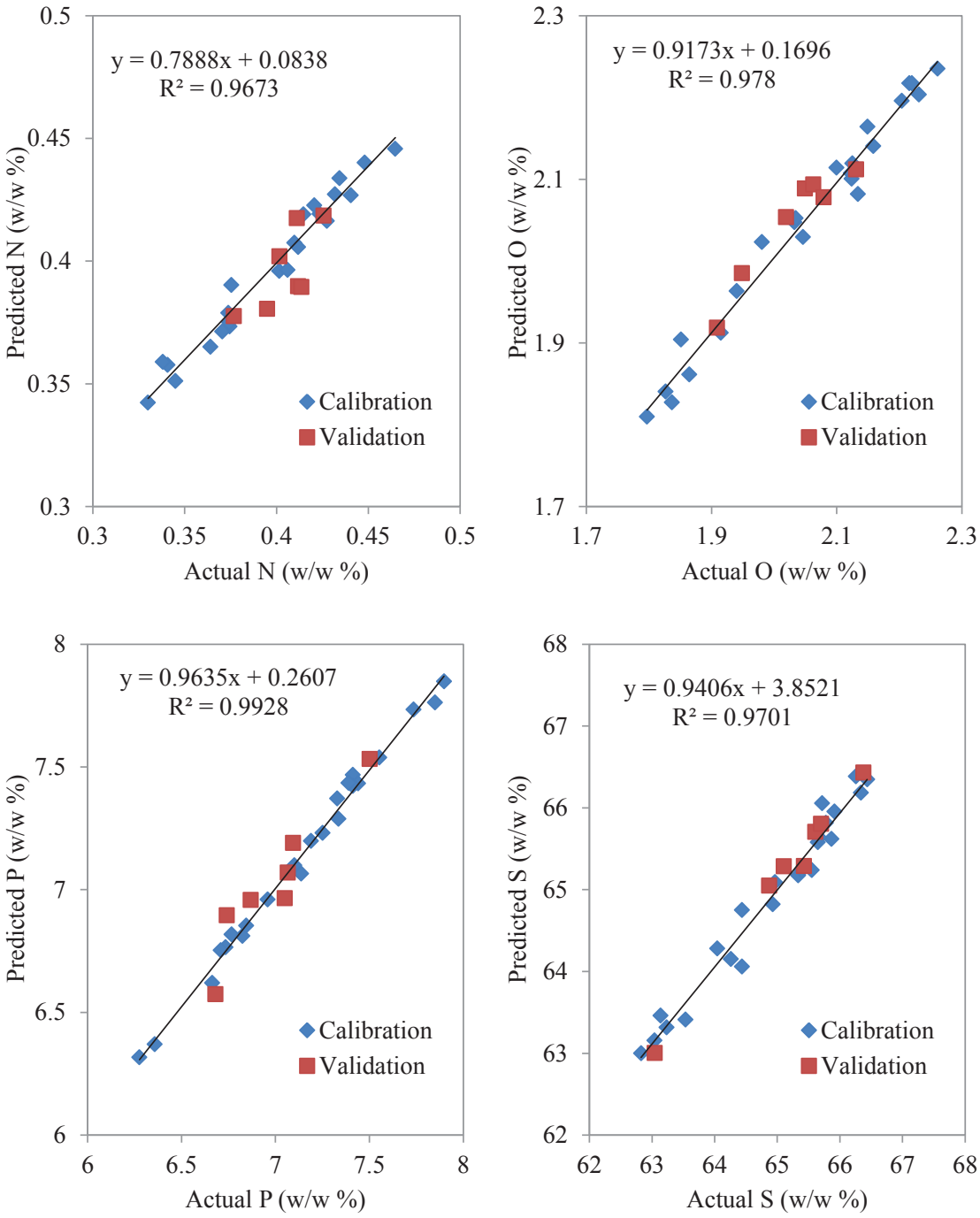

Figure 5.13. Actual versus predicted plot of E, F ,I, J, N, O, P and S contents resulting from GILS in the second phase of the study (cont.)

Among the calibration plots that are shown in Figure 5.13., the correlation coefficients in descending order are: E, F, I, J, N, O, P and S. In order to interpret calibration models, the SECV, SEP and R^2 values along with their operating ranges of each model are shown in Table 5.5.

Table 5.5. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum values of the components (Min and Max) and correlation coefficient (R^2) of GILS models belong to the second phase of the study.

	SECV (w/w %)	**SEP (w/w %)**	**Min (w/w %)**	**Max (w/w %)**	**R^2**
E	0.009	0.010	1.100	1.340	0.985
F	0.147	0.213	14.200	18.500	0.987
I	0.015	0.062	0.900	1.300	0.989
J	0.019	0.097	2.600	3.500	0.995
N	0.004	0.011	0.330	0.450	0.995
O	0.016	0.025	1.800	2.300	0.989
P	0.031	0.150	6.300	8.000	0.996
S	0.154	0.112	63.000	66.500	0.984

As can be seen from Table 5.5., all the models for the 8 components were considered to be quite good regardless of the dynamic ranges of each ingredient. The R^2 values were all around 0.99 for calibration sets. In addition, the agreement between SECV and SEP values were also good for the components E, F and O. In fact for the water content (S), the ratio of SEP over SECV was even lower than 1.0 indicating no overfitting problem despite the relatively low number of calibration samples. On the other hand, models for the components I, J, N and P were resulted in slightly higher SEP values when compared to SECV of the models.

CHAPTER 6

CONCLUSION

In this thesis, calibration models were developed for 2 main products these are liquid soap and shower gel by the combination of FTIR-ATR spectroscopy and multivariate calibration methods. Samples were gathered and analyzed in mid- infrared region by the method of FTIR-ATR spectroscopy. A genetic algorithm based calibration method (GILS) was applied to each component of both liquid soap and shower gel data sets.

Quite good results were determined for both liquid soap study and shower gel study by using GILS. Reliability of the calibration models were calculated by SECV and SEP values as well as with the R^2 values from the reference vs. predicted content plots. For liquid soap study, A (water) and C (surfactant 1) is the best modelled compounds where coefficients are higher. Moreover, a successful modelling study was also carried out in the analysis of liquid soap benchmark products. On the other hand, the R^2 values for the calibration sets were around 0.99 for the components as J (surfactant 3), N (surfactant 4) and P (surfactant 6) which are quite good modelled in shower gel formulation.

In conclusion, within the scope of this thesis, a rapid and simple molecular spectroscopy based analytical method was developed for the determination of the compounds in the liquid soap and shower gel formulations by the use of chemometrics multivariate calibration methods.

REFERENCES

Afkhami, A. *et al.* (2009) 'Simultaneous spectrophotometric determination of binary mixtures of surfactants using continuous wavelet transformation', *Journal of Hazardous Materials*, 166(2–3), pp. 770–775. doi: 10.1016/j.jhazmat.2008.11.118.

Bazel, Y. R. *et al.* (2014) 'Methods for the determination of anionic surfactants', *Journal of Analytical Chemistry*, 69(3), pp. 211–236. doi: 10.1134/s1061934814010043.

Brereton, R. G. (2000) 'Introduction to multivariate calibration in analytical chemistry', *Analyst*, 125(11), pp. 2125–2154. doi: 10.1039/b003805i.

Carolei, L. and Gutz, I. G. R. (2005) 'Simultaneous determination of three surfactants and water in shampoo and liquid soap by ATR-FTIR', *Talanta*, 66(1), pp. 118–124. doi: 10.1016/j.talanta.2004.10.005.

Derrick, M.R; Stulik, D; Landry, J.M; *Infrared Spectroscopy in Conservation Science;* CIP; 99-37860; 1999; 5-18.

Du, X., Lu, Y. and Liang, Y. (1998) 'FTIR and UV-vis spectroscopic studies of black soap film', *Journal of Colloid and Interface Science*, 207(1), pp. 106–112. doi: 10.1006/jcis.1998.5767.

Hoyt, L. F. and Verwiebe, A. (1926) 'Determination of the Concentration of Liquid Soaps by the Immersion Refractometer', *Industrial and Engineering Chemistry*, 18(6), pp. 581–582. doi: 10.1021/ie50198a011.

Im, S. H., Jeong, Y. H. and Ryoo, J. J. (2008) 'Simultaneous analysis of anionic, amphoteric, nonionic and cationic surfactant mixtures in shampoo and hair conditioner by RP-HPLC/ELSD and LC/MS', *Analytica Chimica Acta*, 619(1), pp. 129–136. doi: 10.1016/j.aca.2008.03.058.

Kargosha, K. et al. (2008) 'Simultaneous determination of one nonionic and two anionic surfactants using Fourier transform infrared spectrometry and multivariate

analysis', *Talanta,* 75(2), pp. 589–593. doi: 10.1016/j.talanta.2007.11.065.

Kazemipour, M., Noroozian, E. and Saber, M. (2002) 'Short communication A new second-deri v ati v e spectrophotometric method for the determination of permethrin in shampoo', 30, pp. 1379–1384.Islam, M. T. *et al.* (2004) 'Fourier transform infrared spectroscopy for the analysis of neutralizer-carbomer and surfactant-carbomer interactions in aqueous, hydroalcoholic, and anhydrous gel formulations', *The AAPS Journal*, 6(4), pp. 61–67. doi: 10.1208/aapsj060435.

Kenkel, J.; *Analytical Chemistry for Technicians*; CRC Press; 3rd Edition; 2003; 161.

López-Sánchez, M. *et al.* (2008) 'Assessment of dentifrice adulteration with diethylene glycol by means of ATR-FTIR spectroscopy and chemometrics', *Analytica Chimica Acta*, 620(1–2), pp. 113–119. doi: 10.1016/j.aca.2008.05.032.

Mirghani, M. E. S. *et al.* (2002) 'FTIR spectroscopic determination of soap in refined vegetable oils', *JAOCS, Journal of the American Oil Chemists' Society*, 79(2), pp. 111–116. doi: 10.1007/s11746-002-0443-4.

Samardžić, M., Sak-Bosnar, M. and Madunić-Čačić, D. (2011) 'Simultaneous potentiometric determination of cationic and ethoxylated nonionic surfactants in liquid cleaners and disinfectants', *Talanta*, 83(3), pp. 789–794. doi: 10.1016/j.talanta.2010.10.046.

Setle, F.A. 1997. *Handbook of Instrumental Techniques for Analytical Chemistry*. 1st Edition ed: Prentice Hall PTR.

Settle, F. (1998) 'Handbook Of Instrumental Techniques For Analytical Chemistry - Fran A.Settle.pdf', *Journal of Liquid Chromatography Related Technologies*, pp. 3072–3076. doi: 10.1080/10826079808006889.Conservation, I. (no date) *Scientific_Tools_for_Conservation1996*.Lynch, M. L. *et al.* (2002) 'Intermolecular Interactions and the Structure of Fatty Acid−Soap Crystals', *The Journal of Physical Chemistry B*, 105(2), pp. 552–561. doi: 10.1021/jp002602a.

Settle, F. A. "Handbook Of Instrumental Techniques For Analytical Chemistry - Fran A.Settle." Prentice Hall PTR: 241.

Skoog, D. A., F. J. Holler, and S. R. Crouch. 1998. *Principles of Instrumental Analysis*: Brooks/Cole, Cengage Learning.

Smith, B.C. 1996. Fundamentals of Fourier transform infrared spectroscopy. New York: CRC Press.

Stuart, B; *Infrared Spectroscopy: Fundamentals and Applications,* Wiley Publishing, 1st Edition; 2004; 46-49.

Wolf, R., Wolf, D. and Tu, B. (no date) 'Soaps, Shampoos, and Detergents', (01).Leblanc, A., Dumas, P. and Lefebvre, L. (1999) 'Trace element content of commercial shampoos: Impact on trace element levels in hair', *Science of the Total Environment*, 229(1–2), pp. 121–124. doi: 10.1016/S0048-9697(99)00059-5.

Worden, M. (2019) 'Molecular and Atomic Spectroscopy', pp. 1–97.Khanmohammadi, M. *et al.* (2009) 'Quantitative monitoring of the amidation reaction between coconut oil and diethanolamine by attenuated total reflectance fourier transform infrared spectrometry', *Journal of Surfactants and Detergents*, 12(1), pp. 37–41. doi: 10.1007/s11743-008-1101-7.