

**DEVELOPMENT OF GENETIC ALGORITHMS
BASED MULTIVARIATE CALIBRATION MODELS
FOR THE DETERMINATION OF LUBRICATING
OIL COMPOSITIONS USING FOURIER
TRANSFORM INFRARED SPECTROSCOPY**

**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
İrem ANIL**

**September 2013
İZMİR**

We approve the thesis of **İrem ANIL**

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. Figen TOKATLI

Department of Food Engineering, İzmir Institute of Technology

26 September 2013

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. R. Tuğrul SENGER

Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGMENTS

It is a pleasure to thank the many people who made this thesis possible.

I would like to express my sincere gratitude to my advisor Prof. Dr. Durmuş Özdemir. With his enthusiasm, his inspiration, and his great efforts to explain things clearly and simply, he helped to make chemometrics fun for me. 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 Yalçın and Assoc. Prof. Dr. Figen Tokatlı for their encouragement, insightful comments, and hard questions.

I wish to thank my best friends in all my IZTECH life (Ezgi Tulukçuoğlu, Sinem Dümen, Duygu Turşucu, Cüneyt Çalışkan, Merter Örken, Mithat Boz, İbrahim Çelebi, Emir Güneri, Nihan Toygar and Selen Ünalın) for helping me get through the difficult times, and for all the emotional support, friendship, entertainment, and caring they provided.

Lastly, I would like to thank my family for their motivation, patience and supporting me spiritually throughout my life.

ABSTRACT

DEVELOPMENT OF GENETIC ALGORITHMS BASED MULTIVARIATE CALIBRATION MODELS FOR THE DETERMINATION OF LUBRICATING OIL COMPOSITIONS USING FOURIER TRANSFORM INFRARED SPECTROSCOPY

Engine oils consist of base oils and additives. There have been a number of changes in the composition of engine oils depending on the time and using conditions like the decrease in the additive amounts and increase in the amount of oxidation products. Although there are many physical and chemical standard test methods for the determination of oil quality, none of these methods alone can be used to determine the whole composition of engine oils. The main objective of this work, is to develop a single analytical method that is simple, rapid and accurate for the quantitative determination of lubricating oil compositions using Fourier transform infrared (FTIR) spectroscopy combined with chemometric multivariate data analysis. For this study, a number of most intensively produced engine oils are chosen in an industrial lubrication oil plant and then synthetic mixtures of oil components were prepared in order to develop multivariate calibration models. The FTIR spectra of these samples were recorded using a three reflection attenuated total reflectance (ATR) accessory. The collected spectral data and the reference concentration values of the samples are then used in multivariate calibration modelling step using a genetic algorithm based inverse least squares (GILS) calibration method. It was observed that the correlation coefficients between the reference concentration values and the GILS predicted concentrations were around 0.99. As a result, FTIR spectroscopy combined with multivariate calibration can be a rapid method for the quantitative determination of engine oil compositions.

ÖZET

FOURIER DÖNÜŞÜMLÜ KIZILÖTESİ SPEKTROSKOPİSİ KULLANILARAK MADENİ YAĞ İÇERİKLERİNİN BELİRLENMESİ İÇİN GENETİK ALGORİTMALARA DAYALI ÇOK DEĞİŞKENLİ KALİBRASYON MODELLERİNİN GELİŞTİRİLMESİ

Motor yağları, baz yağı ve katkılardan oluşur. Yağların kullanımı esnasında kullanım koşulları ve süresine bağlı olarak oksidasyon ürünlerinin artması ve katkıların miktarlarının azalması gibi yağ kompozisyonlarında bazı değişimler meydana gelmektedir. Yağ kalitesinin belirlenmesine yönelik çeşitli fiziksel ve kimyasal testlerden oluşan bir çok standart metod olmakla beraber, bu yöntemlerin hiçbiri tek başına yağların kompozisyonunun tamamını tespit etme imkanı sunamamaktadır. Bu çalışmada, madeni yağların kantitatif olarak içeriklerinin belirlenmesi amacıyla Fourier dönüşümlü kızılötesi (FTIR) spektroskopisine ve kemometrik çok değişkenli veri analizine dayalı hızlı ve uygulama kolaylığı olan bir yöntem geliştirilmesi hedeflenmiştir. Bu amaçla, işletmede yoğun olarak üretimi gerçekleştirilen bazı ürün gamları belirlenerek bu ürünlere yönelik çoklu karışımlar hazırlanmış ve bu örneklerin FTIR spektrumları kaydedilmiştir. Elde edilen spektral veriler ile referans konsantrasyon değerleri bir genetik algoritma temelli ters en küçük kareler (GILS) metodu yardımı ile analizlenmiş ve kemometrik çok değişkenli kalibrasyon modeli oluşturulmuştur. Geliştirilen modeller ile hesaplanan bileşen konsantrasyon değerleri ile referans konsantrasyon değerleri arasında 0.99 oranında bir uyum olduğu görülmüştür. Sonuç olarak, FTIR spektroskopisi ve çok değişkenli kalibrasyon ile madeni yağların kompozisyonlarının kantitatif olarak tayin edilebileceği tespit edilmiştir.

TABLES OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES	x
CHAPTER 1. INTRODUCTION	1
CHAPTER 2. INFRARED SPECTROSCOPY	4
2.1. Infrared Region	4
2.2. Infrared Instruments	5
2.2.1. Dispersive Infrared Instruments	6
2.2.2. FTIR Instruments	6
2.3. Attenuated Total Reflectance Fourier Transform Infrared (ATR - FTIR) Spectroscopy	7
2.4. Quantitative Analysis	10
CHAPTER 3. MULTIVARIATE CALIBRATION METHODS	11
3.1. Overview	11
3.2. Classical Least Squares (CLS)	12
3.3. Inverse Least Squares (ILS)	13
3.4. Genetic Inverse Least Squares (GILS)	14
CHAPTER 4. EXPERIMENTATION & INSTRUMENTATION	19
4.1. Experimentation	19
4.2. Instrumentation	22
4.3. Data Analysis	22
CHAPTER 5. RESULTS AND DISCUSSION	24
5.1. 'Engine Oil 1' Results	24
5.1.1. GILS Results	27
5.1.2. PLS Results	44
5.2. 'Engine Oil 2' Results	52

5.2.1. GILS Results	53
5.2.2. PLS Results	57
CHAPTER 6. CONCLUSIONS	63
REFERENCES	65

LIST OF FIGURES

Figure	Page
Figure 2.1. The Peak ID and Fingerprint Regions of the IR spectrum	5
Figure 2.2. A schematic representation for an Dispersive Infrared instrument.....	6
Figure 2.3. A schematic representation for an FT instrument	7
Figure 2.4. A schematic diagram of an ATR accessory	9
Figure 3.1. Flow chart of genetic algorithm used in GILS	16
Figure 5.1. FTIR-ATR spectra of six components of ‘Engine Oil 1’ samples (Res=4)	26
Figure 5.2. FTIR-ATR spectra of six components of ‘Engine Oil 1’ samples (Res=16)	26
Figure 5.3. FTIR-ATR spectra of six components of ‘Engine Oil 1’ samples (Res=32)	27
Figure 5.4. Reference vs. FTIR-ATR predicted ‘Base Oil 1’ contents for the data set of ‘Engine Oil 1’, when resolution was (a) 4 cm ⁻¹ , (b) 16 cm ⁻¹ , (c) 32 cm ⁻¹	28
Figure 5.5. Reference vs. FTIR-ATR predicted ‘Base Oil 2’ contents for the data set of ‘Engine Oil 1’, when resolution was (a) 4 cm ⁻¹ , (b) 16 cm ⁻¹ , (c) 32 cm ⁻¹	29
Figure 5.6. Reference vs. FTIR-ATR predicted Additive 1 contents for the data set of ‘Engine Oil 1’, when resolution was (a) 4 cm ⁻¹ , (b) 16 cm ⁻¹ , (c) 32 cm ⁻¹	31
Figure 5.7. Reference vs. FTIR-ATR predicted B1B2A1 contents for the data set of ‘Engine Oil 1’, when resolution was (a) 4 cm ⁻¹ , (b) 16 cm ⁻¹ , (c) 32 cm ⁻¹	32
Figure 5.8. Reference vs. FTIR-ATR predicted Additive 2 contents for the data set of ‘Engine Oil 1’, when resolution was (a) 4 cm ⁻¹ , (b) 16 cm ⁻¹ , (c) 32 cm ⁻¹	34
Figure 5.9. Reference vs. FTIR-ATR predicted Additive 3 contents for the data set of ‘Engine Oil 1’, when resolution was (a) 4 cm ⁻¹ , (b) 16 cm ⁻¹ , (c) 32 cm ⁻¹	35

Figure 5.10. Reference vs. FTIR-ATR predicted Additive 4 contents for the data set of ‘Engine Oil 1’, when resolution was (a) 4 cm ⁻¹ (b) 16 cm ⁻¹ (c) 32 cm ⁻¹	37
Figure 5.11. Frequency distribution of GILS selected FTIR-ATR wavelengths for (a) Base Oil 1, (b) Base Oil 2, (c) Additive 1, (d) B1B2A1, (e) Additive 2, (f) Additive 3, (g) Additive 4 contents of ‘Engine Oil 1’ samples when the resolution was 4 cm ⁻¹	38
Figure 5.12. Frequency distribution of GILS selected FTIR-ATR wavelengths for (a) Base Oil 1, (b) Base Oil 2, (c) Additive 1, (d) B1B2A1, (e) Additive 2, (f) Additive 3, (g) Additive 4 contents of ‘Engine Oil 1’ samples when the resolution was 16 cm ⁻¹	40
Figure 5.13. Frequency distribution of GILS selected FTIR-ATR wavelengths for (a) Base Oil 1, (b) Base Oil 2, (c) Additive 1, (d) B1B2A1, (e) Additive 2, (f) Additive 3, (g) Additive 4 contents of ‘Engine Oil 1’ samples when the resolution was 32 cm ⁻¹	42
Figure 5.14. PLS Results for the set of ‘Engine Oil 1’ in the range of 4000-600 cm ⁻¹ (Res=4).....	45
Figure 5.15. PLS Results for the set of ‘Engine Oil 1’ in the range of 1500-600 cm ⁻¹ (Res=4).....	47
Figure 5.16. PLS Results for the set of ‘Engine Oil 1’ in the range of 1300-800 cm ⁻¹ (Res=4).....	50
Figure 5.17. Reference vs. FTIR-ATR predicted (a) Base Oil 1, (b) Base Oil 2, (c) Additive 1, (d) Additive 2, (e) Additive 3 contents for the data set of ‘Engine Oil 1’, when resolution was 4 cm ⁻¹	54
Figure 5.18. Frequency distribution of GILS selected FTIR-ATR wavelengths for (a) Base Oil 1,(b) Base Oil 2, (c) Additive 1, (d) Additive 2, (e) Additive 3, contents of ‘Engine Oil 2’ samples, when the resolution was 4 cm ⁻¹	56
Figure 5.19. PLS Results for the set of ‘Engine Oil 2’ in the range of 4000-600 cm ⁻¹ (Res=4).....	57
Figure 5.20. PLS Results for the set of ‘Engine Oil 2’ in the range of 1500-600 cm ⁻¹ (Res=4).....	59
Figure 5.21. PLS Results for the set of ‘Engine Oil 2’ in the range of 1300-800 cm ⁻¹ (Res=4).....	61

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 1.1. Several ASTM Methods Used for Quality Control of Lubricants.....	2
Table 4.1. Components and their concentration ranges (w/w %) used in training set of 'Engine Oil 1'	19
Table 4.2. Concentration profile of the samples in the set of 'Engine Oil 1'	20
Table 4.3. Components and their concentration ranges (w/w %) used in training set of 'Engine Oil 2'	22
Table 5.1. ANOVA Results for the data set of 'Engine Oil 1' ($\alpha = 0.05$).....	44
Table 5.2. Calibration summary for the data set of 'Engine Oil 1' (Res=4)	52
Table 5.3. Calibration summary for the data set of 'Engine Oil 2' (Res=4)	62

CHAPTER 1

INTRODUCTION

Engine oil is a kind of lubricants and it is mostly used for reducing the friction between the moving parts in engines. In these powerful machines (e.g. internal combustion engines, compressors, turbines, aircrafts, cars, boats and electrical generators), there are lots of moving components and when they are in contact relative to each other, the force of friction tends to prevent this relative motion. Then friction converts the kinetic energy to heat. This property can cause striking effects such as decline in the performance of engine and enlargement of fuel consumption. At this point, engine oil creates a thin, separating film between moving surfaces to minimize direct contact between them, transfers the heat caused by friction and protects the engine. There are also a vast number of other applications of lubricants so between 5000 and 10000 different formulations are necessary to satisfy more than 90% of all requirement.

Engine oils are made from heavier, thicker petroleum hydrocarbon bases derived from crude oil and many additives to improve certain properties. The bulk of a typical engine oil has hydrocarbons with between 18 and 34 carbon atoms per molecule. On average, engine oils consist of about 90% base oils (hydrogenated polyolefins, esters, silicones, fluorocarbons and many others) and 10% chemical additives (mostly salts of organic acids and such metallic ions as zinc, barium, magnesium or calcium) [1].

During lubricant use, contamination, loss of additive performance and an increase in oxidation products can occur. The lubricants could have an unlimited lifetime if they are not contaminated by any kind of agents. The quality control of lubricating oil is essential for preservation of the longevity and the performance of industrial machines, automotives, and equipment which depends on hydraulic fluids. Based on this, changes in oil quality need to be detected and potential problems fixed before they become serious. Analysis of oils during use can also help to prevent unnecessary replacement of oil and premature engine overhauls [2].

In lubricants industry, many standard primary chemical methods that American Society for Testing and Materials (ASTM) approved are employed for the determination

of lubricants quality. The analyses that generally employed in engine oil quality control are viscosity, total acid number (TAN) , total base number (TBN), water and fuel contamination and oxidation. The viscosity is the most important parameter to be monitored in quality control, because an increase in viscosity can indicate the presence of insolubles, oxidation products, replacement by degraded oil or water. On the other hand, a decrease can indicate the presence of fuel, replacement by a different oil and additive breakdown. Table 1.1 shows some of these ASTM methods :

Table 1.1. Several ASTM Methods Used for Quality Control of Lubricants

Method	ASTM Description
ASTM D 97	Pour Point of Petroleum Products
ASTM D 445	Kinematic Viscosity of Transparent and Opaque Liquids
ASTM D 664	Acid Number of Petroleum Products by Potentiometric Titration
ASTM D 4739	Base Number of Petroleum Products by Potentiometric Perchloric Acid Titration
ASTM D 129	Sulfur in Petroleum Product (General Bomb Method)
ASTM D 240	Heat of Combustion of Liquid Hydrocarbon Fuels by Bomb Calorimeter
ASTM D 972	Evaporation Loss of Lubricating Oils
ASTM D 5185	Additive Elements ,Wear Metals and Contaminants in Used Oils by ICP-AES
ASTM D 4294	Phosphorus, Sulfur, Calcium, and Zinc in Lubrication Oils by ED-XRF

The ASTM standard methods mentioned above are generally based on the titration of all the basic constituents in an engine oil with standardized acid to a fixed endpoint, the titer, however, being expressed in terms of base number (mg KOH/g oil). As is often the case with solvent-based titrimetric methods, such methods are fine in theory, but problematic in practice. Aside from the time-consuming and labor-intensive nature of the analysis itself, a major drawback is the need for substantial volumes of organic solvents and corrosive reagents that are hazardous and difficult to dispose of. Even with automated potentiometric titration versions of the ASTM methods, the analysis is still time-consuming. Generally, these tests mentioned above are in general slow, time-consuming and they give information only about the partial properties of the lubricants so they are not capable of determining the composition of an oil [3].

The aim of this study, is to develop a fast, simple and accurate procedure for the determination of lubricating oil composition based on the use of the FTIR spectroscopy along with multivariate calibration. In industry, particular lubricating oil mixtures are prepared by technicians and after that they are produced in larger scales, by tones everyday. However, there is no analytical technique to confirm the content of the oils correctly for quality control of lubricants. If personal errors result from the carelessness or inattention occur during preparation of oil mixtures, improper oil produced will reach the customer. In some cases, this may result in loss of the company's profits. To overcome this problems, in this study, the determination of base oils and additive components in engine oils is presented with the aim of developing a fast, simple and accurate procedure based on the diffuse-reflectance mid infrared spectroscopy along with multivariate calibration methods for the routine analysis in engine oil industry. All the oils produced in engine oil industry can be checked using this new method before they reach the customer. For a chemometric point of view, several chemometrical methods with IR spectroscopy were proposed in the literature for analysis of complex engine oil blends. Felkel et al. [4] developed partially least squares (PLS) model for the determination of total acid number (TAN) using IR data measured from monograde mineral gas engine oils. In another study, Al-Ghouti et al. [5] performed PLS, classical least squares (CLS) and principal component regression (PCR) as an alternative way of ASTM D 2896 and ASTM D 445 for the determination of total base number (TBN) and kinematic viscosity. In a comparative study, Borin et al used [6] principal component analysis (PCA) with FTIR data to identify the condition of engine oils based on viscosity. Also, Blanco et al. [7] used partial least squares regression (PLSR) and multiple linear regression (MLR) with mid-infrared (MIR) and near infrared (NIR) data to determine water in lubricating oils which consist of highly additive that introduce large errors by the Karl-Fischer and hydride methods. However, it should be emphasized again that there is not any alternative methods for quantitative determination of lubricants in the literature so this new method will be a significant improvement for the industry.

CHAPTER 2

INFRARED SPECTROSCOPY

2.1. Infrared Region

Infrared (IR) Spectroscopy is a popular technique used for compound identification. Simply, it is the absorption measurement of different IR frequencies by a sample positioned in the path of an IR beam. Different functional groups in sample absorb different frequencies of infrared radiation and these characteristic frequencies are used to determine chemical structures. Generally, almost any solid, liquid or gas sample can be analyzed and it is an important method which has a wide variety of applications such as scientific, industrial, medical, military and forensic analysis [8].

The infrared region of the electromagnetic spectrum is classified into three groups; the near-, mid- and far- infrared, named for their relation to the visible spectrum. The higher-energy near-IR, approximately $12800\text{--}4000\text{ cm}^{-1}$ ($0.78\text{--}2.5\text{ }\mu\text{m}$ wavelength) can excite overtone or harmonic vibrations. The mid-infrared, approximately $4000\text{--}200\text{ cm}^{-1}$ ($2.5\text{--}50\text{ }\mu\text{m}$) may be used to study the fundamental vibrations and associated rotational-vibrational structure. The far-infrared, approximately $200\text{--}10\text{ cm}^{-1}$ ($50\text{--}1000\text{ }\mu\text{m}$), lying adjacent to the microwave region, has low energy and may be used for rotational spectroscopy [9].

According to the literature, chemists divide the IR region into two parts : The region from 4000 cm^{-1} to approximately 1500 cm^{-1} is named as ‘Peak ID Region’ because it is especially useful for correlating peak location with bonds. Secondly, the region from 1500 to 600 cm^{-1} is typically very busy and is not as useful for such correlation but it remains very useful as the molecular fingerprint. For this reason, this region is called as ‘Fingerprint Region’. This means that this region can still be used for peak-for-peak matching with a known spectrum from a library of known spectra. Figure 2.1 shows the Peak ID and the Fingerprint Regions of the IR spectrum [10].

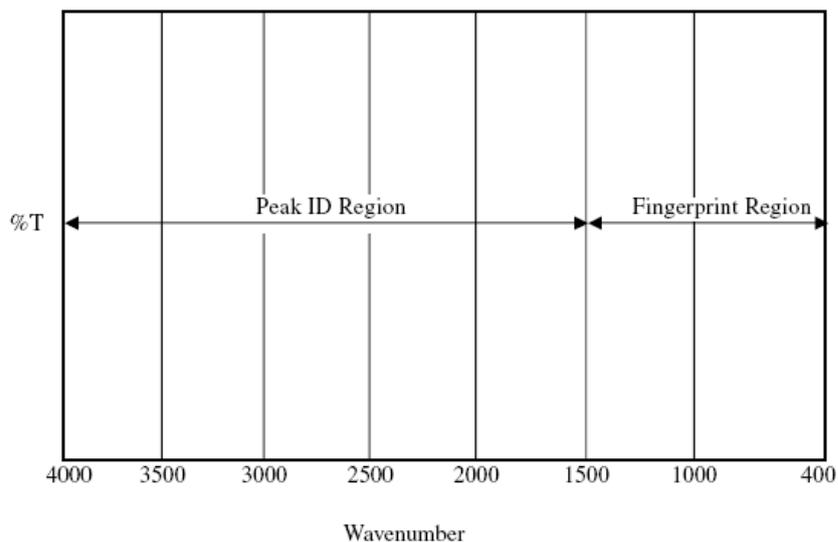


Figure 2.1. The Peak ID and Fingerprint Regions of the IR spectrum [10]

2.2. Infrared Instruments

Generally, an IR instrument contains a source of infrared radiation, a sample container which should be infrared transparent, a wavelength selecting device, a detector and a signal processor, consecutively. The common radiation source for the IR spectrometer is an inert solid heated electrically to 1000 to 1800 °C. Three mostly used sources are Nernst glower (constructed of rare-earth oxides), Globar (constructed of silicon carbide), and Nichrome coil. As a sample holder, quartz cells are used in the NIR region and in the MIR and FIR region potassium bromide (KBr) is used. Wavelength selecting devices are used to disperse a broad spectrum of radiation. Other component, detector used in IR spectrometers can be categorized into two classes: thermal detectors and photon detectors. Thermal detectors include thermocouples, thermistors, and pneumatic devices (Golay detectors). Photon detectors rely on the interaction of IR radiation and a semiconductor material. Lastly, a signal processor amplifies the signal from the detector [11].

Measurement of IR absorption can be made using three types of instruments such as dispersive spectrometers with a grating monochromator, fourier transform spectrometers with an interferometer and nondispersive photometers using a filter or an absorbing gas that are used for analysis of atmospheric gases at specific wavelengths [10].

2.2.1. Dispersive Infrared Instruments

A typical dispersive IR instrument consists of a monochromator which disperse the radiation into component frequencies after it comes from the source and passes through the sample and reference path. Then the wavelengths of light are separated in the spectral range and each wavelength is individually transferred to the detector, which generates an electrical signal and results in a recorder response. Most of the dispersive spectrometers have a double-beam design. A representative figure for a dispersive instrument is shown in Figure 2.2 [12].

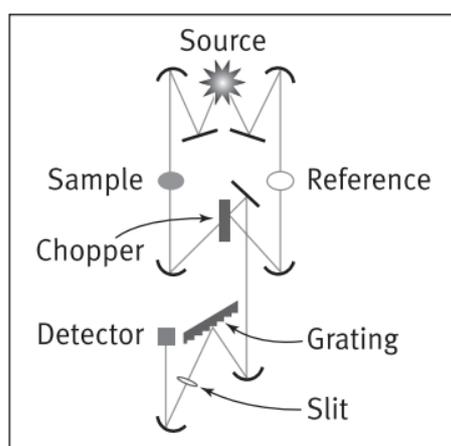


Figure 2.2. A schematic representation for an Dispersive Infrared instrument [12]

2.2.2. FTIR Instruments

A FT system consists of three main spectrometer components: source, interferometer and detector. The same types of sources are used for both dispersive and fourier transform spectrometers. On the contrary of dispersive instruments, the monochromators are replaced by interferometers in FT spectrometers. The most commonly used interferometer is a Michelson interferometer. It consists of three active components: a moving mirror, a fixed mirror, and a beamsplitter. The two mirrors are perpendicular to each other. Figure 2.3 shows parts of a typical FT instruments [8].

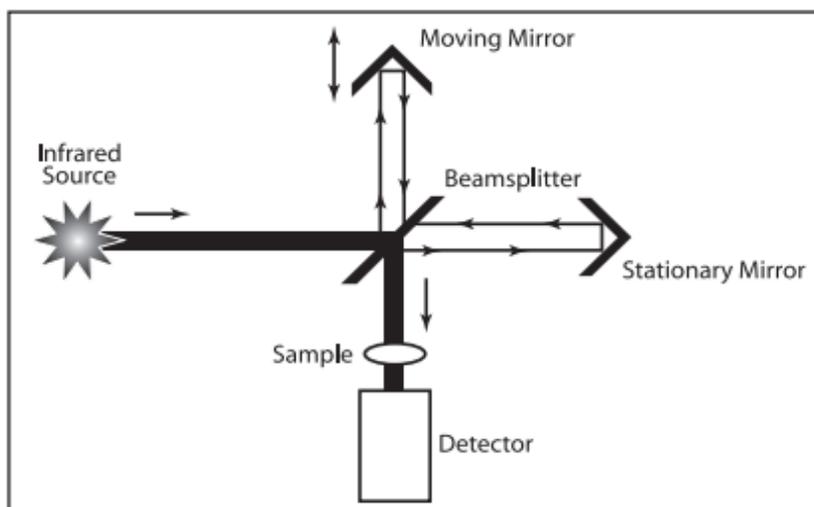


Figure 2.3. A schematic representation for an FT instrument [12]

Collimated light from a broad-band infrared source comes to the beamsplitter which is made from a thin film of germanium. Nearly 50% of the light passes through the film and is reflected back by a fixed mirror, where half of the light intensity (25% of the original light intensity) is reflected by the beamsplitter through the sample cell to the detector. The other 50% of the original light intensity is reflected to the moving mirror. Then half of this light is transferred to the detector after passes through the sample. Finally, 50% of the original light reaches to the detector [10].

FTIR spectrometers are more preferable than dispersive spectrometers because FTIR spectrometers provide the data collection faster and also , in a FTIR spectrometer there are fewer mirror surfaces so there will be less reflection losses and signal-to-noise ratio will be higher in the spectrum [10].

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

Mid-Infrared spectroscopy is a popular and trustable fingerprinting method by scientists for the characterization and also quantification of the substances. Mostly preferable way of obtaining mid-infrared spectra is the transmission technique of sampling, although it has many disadvantages. One of the disadvantages is the “thickness problem”. Samples thicker than 20 microns absorb too much infrared radiation so it will be not possible to get a spectrum. Also, if the sample is thinner than

1 micron, its absorption will be so weak that can not be detected. Another disadvantage is that the sample preparation is time consuming because of the melting, squishing or diluting requirements so that it transmits the appropriate amount of light. For these reasons, reflectance techniques can be more preferable than transmission ones to obtain mid-infrared spectra. Reflectance techniques has no thickness problem so there is no need to worry about the thickness or the concentration of the sample and it is not necessary to spend much time during the sample preparation. This means the sample preparation for reflectance samples is faster and easier than for transmittance samples. A final advantage of some reflectance techniques is that they are nondestructive. The sample is left intact after its spectrum is obtained , which means the sample can be used for other analyses [13].

The attenuated total reflectance (ATR) technique is a kind of reflectance technique used with mid-infrared spectroscopy and it is used to obtain the spectra of solids, liquids, semi solids, and thin films. It is a fast method and also it gives strong signals although the samples are in nanogram quantities [12,14].

The ATR-FTIR spectroscopy is performed using an accessory which is placed onto the sample compartment of an FTIR instrument. Accessory consists of a crystal which is made of infrared transparent material with high refractive index and also mirrors which bring the IR radiation to a focus on the face of the crystal. Infrared radiation passes through the crystal and reaches to its top surface. At this point, 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. That is called total internal reflection. Infrared beam reflects off the crystal surface three times before leaving the crystal. This procedure is shown in the schematic diagram of an ATR accessory in Figure 2.4.

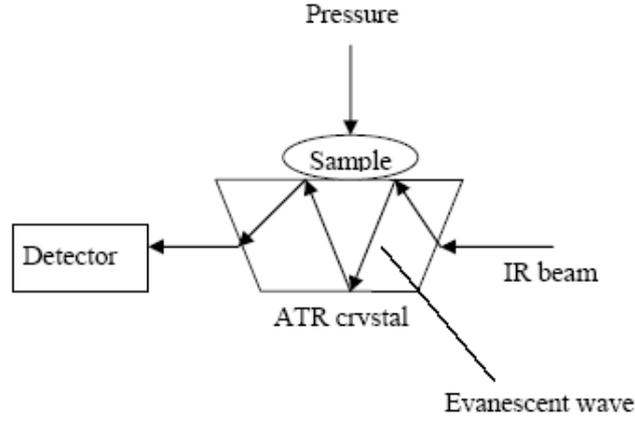


Figure 2.4. A schematic diagram of an ATR accessory [13]

Total refractive index only occurs where the wave travels from a medium with a higher refractive index to one with a lower refractive index. There are several internal total reflections occur within the crystal until the beam reaches the end. The IR radiation reflects with a critical angle, θ_c , depends on the refractive index of the crystal n_1 , and of external medium, n_2 , can be presented as:

$$\theta_c = \sin^{-1} n_{21} \quad (2.1)$$

In addition, when the radiation is inside the crystal, a transmitted wave of radiation is formed which is called evanescent wave. The evanescent wave is made weaker by the sample's absorbance. Evanescent wave is characterized by its amplitude which is rapidly growing with the distance from the interface represented as:

$$E = E_0 e^{-\frac{z}{d_p}} \quad (2.2)$$

where E_0 represents the time averaged electric field intensity at the interface, E is the time averaged field intensity at a distance z from the interface in the rarer medium and d_p is the penetration depth of the evanescent field which is given by:

$$d_p = \frac{\lambda_1}{2\pi (\sin^2 \theta - n_{21}^2)^{1/2}} \quad (2.3)$$

where $\lambda_1 = \lambda/n_1$ and $n_{21} = n_2/n_1$. The larger λ or the smaller θ , the larger the penetration depth [12, 15].

2.4. Quantitative Analysis

Infrared spectroscopy is a reliable and mostly preferred method not only for qualitative analyses but also quantitative analyses without consumption or destruction of the sample. The development of FTIR instruments and strong chemometrical data processing methods improve the performance of quantitative IR work. The main point of the quantitative analysis of absorption spectrometry is Beer's law. For a single compound in a homogeneous medium, the absorbance at any frequency is expressed as

$$A = abc \quad (2.4)$$

where A is the measured absorbance of the sample at the given frequency, a is the molecular absorptivity at the frequency, b is the path length of source beam in the sample, and c is the concentration of the sample. This law generally expresses that the intensities of absorption bands are linearly proportional to the concentration of each component in a homogeneous mixture or solution. For a simple two-component mixture, the total absorbance, A_T , of the mixture at a given frequency is the sum of the absorbance of two component compounds, x and y , at the specified frequency:

$$A_T = A_x + A_y = a_x b c_x + a_y b c_y \quad (2.5)$$

Using matrix algebra it is possible to extend this technique to mixtures containing more than two components. The absorbance of a mixture of n independently absorbing components at a particular frequency n may be expressed in the following equation:

$$A_n = a_1 b c_1 + a_2 b c_2 + \dots + a_n b c_n \quad (2.6)$$

where A_n is the total absorbance of the sample at the frequency n , a_j is the absorptivity of component j at the frequency n ($j = 1, 2, \dots, n$), c_j is the concentration of component j , and b is the sample path length [10].

CHAPTER 3

MULTIVARIATE CALIBRATION METHODS

3.1. Overview

Calibration is a very important part of all analytical procedures and it determines the relationship between the analytical response and the analyte concentration. The simplest form of a linear calibration model is univariate calibration and it is shown in Equation 3.1.

$$y_i = x_i b_1 + e_i \quad (3.1)$$

where y_i represents the concentration of the i^{th} calibration sample, x_i denotes the corresponding instrument reading, b_1 symbolizes the calibration coefficient (slope of the fitted line), and e_i signifies the error associated with the i^{th} calibration sample. A single instrument response, e.g., absorbance at a single wavelength, is measured for each calibration sample.

It should be noted that while other constituents can be present in the calibration samples, the selected wavelength must be spectrally pure for the analyte, i.e., other constituents do not respond at the wavelength. Additionally, matrix effects must be absent at the selected wavelength, i.e., inter- and intramolecular interactions are not present. Values in y and x are used to estimate the model parameter b_1 by the least squares procedure. This least-squares estimate, \hat{b}_1 , is computed by;

$$\hat{b}_1 = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y} \quad (3.2)$$

In Equation 3.2, the symbol \hat{b}_1 is used to emphasize its role as an estimate of b_1 . The resulting calibration model is used to predict the analyte concentration for an unknown sample, is expressed by;

$$\hat{y}_{unk} = x_{unk} \hat{b}_1 \quad (3.3)$$

where x_{unk} represents the response for the unknown sample measured at the calibrated wavelength. This kind of calibration is called univariate calibration because only one response variable is used [16].

However, it is not possible to analyze multiple components in a sample simultaneously. When univariate method is used, there has to be one measurement for each component and more time will be spent. Also, univariate calibration causes wrong prediction of concentration of analyte because of the presence of interferences. To prevent this problem, physical separation of analyte from interfering material or using selective measurements is needed and this means necessity of a great effort. At this point, multivariate calibration will be better alternative because it has fault-detection capabilities and unknown interferences in the sample can be overcome by multivariate calibration.

Multivariate calibration develops the equations in two ways as classical calibration case and inverse calibration case. According to first one, absorbance is a function of concentration and second one is that concentration is a function of absorbance. Differently from univariate calibration, absorbance values in the full spectrum of one sample are used. For this reason, the absorbance vector in univariate calibration becomes a matrix. Also, concentration matrix is a matrix because more than one component can be used [17].

In this study, a hybrid method of genetic algorithm and inverse least squares, genetic inverse least squares method is used. Before discussing this method, it is necessary to explain classical least squares and inverse least squares methods as an introduction to the multivariate calibration methods.

3.2. Classical Least Squares (CLS)

Classical Least Squares is also known as K-matrix method or Beer's Law method and is modeled by the following equation:

$$\mathbf{A} = \mathbf{C} \times \mathbf{K} + \mathbf{E}_A \quad (3.4)$$

where \mathbf{A} is the matrix which consists of absorbance values of the samples at different wavelengths, \mathbf{C} is the matrix which consists of concentrations of multi-component samples \mathbf{K} is the matrix of absorptivity coefficients multiplied by path

length and E_A is the matrix of spectral errors or residuals not fit by the model. Here K matrix represents the first order estimates of the pure component spectra at unit concentration and unit path length. K matrix can be determined by the following formula:

$$\hat{K} = (C^t C)^{-1} C^t A \quad (3.5)$$

Then the concentrations of unknown sample can be predicted in the following equation;

$$\hat{c} = (\hat{K} \hat{K}^t)^{-1} \hat{K} a \quad (3.6)$$

where a is the spectrum of the unknown sample and c is the vector of the predicted component concentrations. Lastly, the residual is the difference between the reference and predicted concentration values;

$$e = c - \hat{c} \quad (3.7)$$

As a result, CLS is a very popular method because it can be applied to simple systems where all of the pure-component spectra can be measured. However, it has the disadvantage of requirement that all interfering chemical components must be known and if there is another components which we ignore in the sample, this method will fail [18].

3.3. Inverse Least Squares (ILS)

Inverse Least Square or P Matrix method applies the inverse of Beer's Law and assumes that component concentration is a function of absorbance as shown in the following equation:

$$C = A \times P + E_C \quad (3.8)$$

where C is the concentration matrix, A is the absorbance matrix and E_C is the matrix of errors in the concentrations not fit by the model as in CLS. Model error is assumed to derive from error in the measurement of component concentration, whereas no error is assumed to be inherent in the absorbance values. The matrix P involves the

unknown calibration coefficients relating component concentrations to the spectral intensities and can be determined by:

$$\hat{p} = (A^t A)^{-1} A^t \times c \quad (3.9)$$

The greatest advantage of ILS is that the Equation 3.8 can be reduced for the analysis of a single component at a time and this reduced model is shown in the following equation;

$$c = Ap + e_c \quad (3.10)$$

Lastly, A predicted concentration of a multi-component sample can be obtained by;

$$\hat{c} = a^t \times \hat{p} \quad (3.11)$$

where \hat{c} is the scalar estimated concentration of component that is being analyzed, \hat{p} is the vector of calibration coefficients and a is the spectrum of unknown sample.

An ILS model has a significant advantage in that it does not need to know and include all components in the calibration set. This means that ILS assumes that the intensities for each measured variable in the analysis all behave perfectly independently. Additionally, you are restricted from using all of the spectral channels in making the model. The number of channels of spectral information used cannot exceed the number of calibration standards. Precision will be reduced if more channels are included than the number of independent sources of variation in the data. [19].

3.4. Genetic Inverse Least Squares (GILS)

Genetic inverse least squares (GILS) is an important form of a genetic algorithms (GA). It is used for selecting wavelengths to build multivariate calibration models with reduced data set. GA are global search and optimization methods based on the principles of natural evolution and selection as developed by Darwin [20]. According to Darwin's theory; variation is a feature of natural populations and every population produces offspring. The consequences of this overproduction is that those

individuals with the best genetic fitness for the environment will produce offspring that can more successfully compete in that environment. Thus the later generation will have a higher representation of these offspring and the population will have evolved [21].

In the last several years there has been widespread interaction among researchers studying various evolutionary computation methods. In the 1950s and the 1960s, several computer scientists independently started to study evolutionary systems with the idea that evolution could be used as an optimization tool for engineering problems. The idea in all these systems was to evolve a population of solutions to a given problem, using operators inspired by natural genetic variation and natural selection. Box 1957, Friedman 1959, Bledsoe 1961, Bremermann 1962, and Reed, Toombs, and Baricelli 1967 all worked in the 1950s and the 1960s to develop evolution-inspired algorithms for optimization and machine learning. In addition, a number of evolutionary biologists, Baricelli 1957, 1962; Fraser 1957, Martin and Cockerham 1960 used computers to simulate evolution for the purpose of controlled experiments. Genetic Algorithms (GAs) were invented by John Holland in the 1960s and were developed by Holland and his students and colleagues at the University of Michigan in the 1960s and the 1970s. In contrast with evolution strategies and evolutionary programming, Holland's original goal was not to design algorithms to solve specific problems, but rather to formally study the phenomenon of adaptation as it occurs in nature and to develop ways in which the mechanisms of natural adaptation might be imported into computer systems. Over the years, GA were used for wavelength selection in many applications especially in chemometrics [22].

Genetic algorithms (GA) have five basic steps including initialization of a gene population, evaluation of the population, selection of the parent genes for breeding and mating, crossover and mutation and replacing parents with their offspring. These steps have taken their names from the biological foundation of the algorithm. GILS has the same steps namely initialization, breeding, mutation and evaluation as other GAs to select a subset of wavelengths. In contrast to other methods, only GILS encodes genes. Figure 3.1 shows the steps of a typical GA.

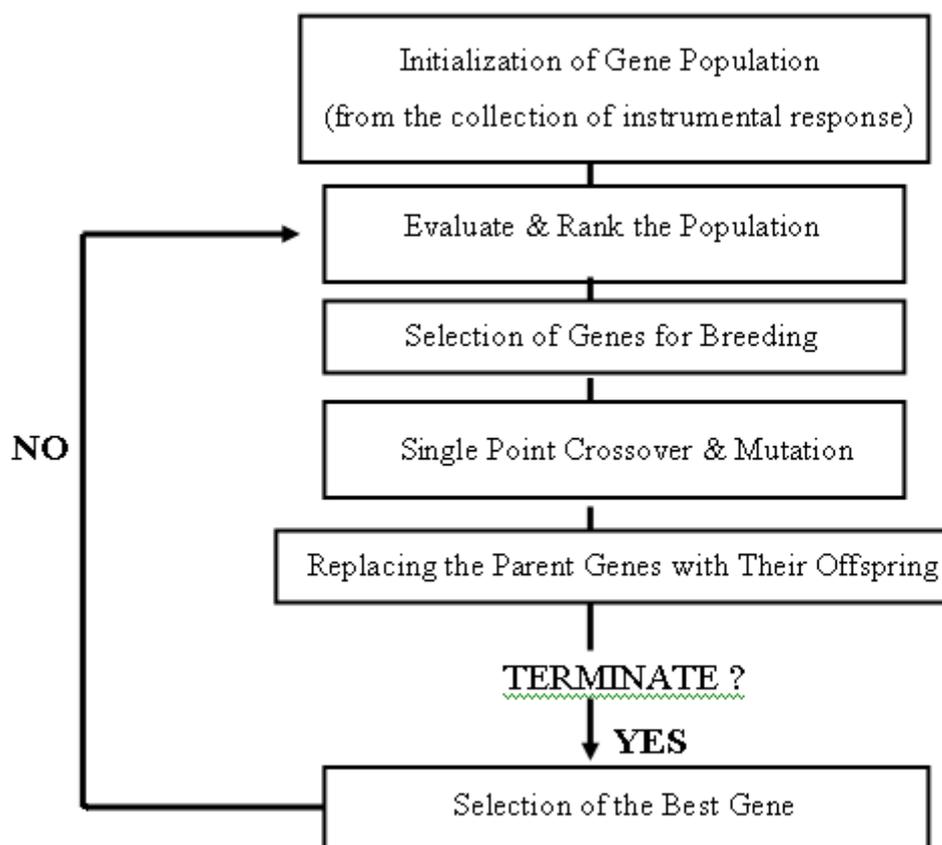


Figure 3.1. Flow chart of genetic algorithm used in GILS

A gene represents a potential solution to a given problem and here, this term is used to describe the randomly selected collection of instrumental response at the wavelength or wavenumber range given in the data set. The collection of individual genes in the current generation is called as ‘population’. In the population, a gene consists of absorbance values at randomly chosen wavelengths or wavenumber between a predefined lower and upper limit. A gene is shown as the following :

$$S = [A_{2134}A_{435}A_{785}] \quad (3.12)$$

where S is so-called a gene, A is the absorbance measured at the indicated wavelength. In the ‘initialization of gene population’ step, many solutions are randomly generated to form an initial population. The population size depends on the nature of the problem, but typically contains several hundreds or thousands of possible solutions. These solutions are the genes of the first generation. The total number of genes in the population should be even in order to allow breeding of each gene in the population. Once the initial gene population is created, the next step is to evaluate and rank the

genes in order to evaluate each gene's success using a fitness function such as the reciprocal of *Standard Error of Calibration with cross Validation (SECV)*. *SECV* is calculated from the following equation;

$$SECV = \sqrt{\frac{\sum_{i=1}^m (c_i - \hat{c}_i)^2}{m - 2}} \quad (3.13)$$

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. Degrees of freedom is $m - 2$ because when a linear model is assumed, there are only two parameters to be extracted which are the slope of the actual vs. reference concentration plot and the intercept. In each step, increase in the fitness value is targeted. In the next step, for each successive generation, a proportion of the parents with high fitness function are selected to breed a new generation. The goal is that only the best performing members of the population will survive in the long run. Here, it is expected that the genes better suited for the problem will generate better off-springs. The genes with low fitness values will be given lower chance to breed and hence most of them will be unable to survive [20]. There are lots of selection methods such as top down method, roulette wheel method and tournament selections. The one which is used in GILS is the roulette wheel selection method. In this method, each gene is represented as different parts on the wheel. The gene with the highest fitness has the slot that has the largest area and the gene with the lowest fitness has the slot that has the smallest area. Therefore, when the wheel is rotated, there is a higher chance of selection for a gene with high fitness than for a gene with a low fitness. There will also be the genes which are selected multiple times and some of the genes will not be selected at all and will be thrown out from the gene pool. After all the parent genes are selected, they are allowed to mate top-down, whereby the first gene S_1 mates with the second gene S_2 ; S_3 with S_4 and so on until all the genes mate. In the 'crossover and mutation step', the genes are broken at random points and the offspring genes are formed by cross-coupling.

Consider S_1 and S_2 are parent genes which are to breed; S_3 and S_4 are their corresponding off-springs shown in the following example:

Parents

$$S_1 = (A_{1245}, A_{872}, \#A_{2170}A_{2118})$$

$$S_2 = (A_{1725}, A_{1578}, \#A_{1333}, A_{940}, A_{1461}, A_{2368}, A_{941}, A_{1258})$$

The points where the genes are cut for mating are indicated by #.

Offspring

$$S_3 = (A_{1245}, A_{872}, A_{1333}, A_{940}, A_{1461}, A_{2368}, A_{941}, A_{1258})$$

$$S_4 = (A_{1725}, A_{1578}, A_{2170}, A_{2118})$$

where A_{1245} represents the instrument response at the wavelength given in subscript. Here S_1 and S_2 are the first and second parent genes. Also, S_3 and S_4 are the offspring. Parent genes are broken at the points indicated by #. The first part of S_1 combines with the second part of the S_2 and create the first offspring (S_3). Then the rest of the parent genes gather and other offspring (S_4) is formed. This process is called the single point crossover and is common in GILS.

After crossover, the parent genes are replaced by their off-springs. The ranking process based on their fitness values follows the evolution step. Then the selection for breeding and mating starts again. This is repeated until a predefined number of iterations are reached. At the end, the gene with the lowest *SECV* (highest fitness) is selected for model building. This model is used to predict the concentrations of component being analyzed in the independent validation set. The success of the model in the prediction of the independent validation set is evaluated using *Standard Error of Prediction (SEP)* which is calculated as:

$$SEP = \sqrt{\frac{\sum_{i=1}^m (c_i - \hat{c}_i)^2}{m}} \quad (3.14)$$

where m is now, the number of independent validation samples. At the end of the process, the termination of the algorithm is done by setting predefined iteration number for the number of breeding/mating cycles. Then, the gene with the lowest *SECV* (this means it has highest fitness) is selected for the model building and this model is used to predict the concentrations of component being analyzed in the prediction (test) sets. The success of the model in the prediction of the test sets is evaluated using standard error of prediction (*SEP*) [23].

CHAPTER 4

EXPERIMENTATION & INSTRUMENTATION

4.1. Experimentation

Firstly, the engine oil which is mostly sold by Opet Fuchs Madeni Yağ Sanayi ve Ticaret A.Ş was determined. It was coded as ‘Engine Oil 1’ not to disclose its real name because of the confidentiality requirement. Then in our laboratory, according to the original engine oil (Engine Oil 1) formulation, a training set of 60 samples were prepared mixing the components which were supplied by Opet Fuchs Madeni Yağ Sanayi ve Ticaret A.Ş. In the training samples, concentrations of the components are not the same as in the original formulation but they are very close with real ones. Components and their concentration ranges used in training samples are shown in Table 4.1. All concentrations are given in grams.

Table 4.1. Components and their concentration ranges (w/w %) used in training set of ‘Engine Oil 1’

Name of the Component	(w/w %) (g/g)
Base Oil 1	34.62 – 47.26
Base Oil 2	23.60 – 31.87
Additive 1	10.60 – 25.80
Additive 2	0.34 – 1.34
Additive 3	0.11 – 0.58
Additive 4	6.89 – 18.56

According to the original formulation of ‘Engine Oil 1’ , major components Base Oil 1 and Base Oil 2 are base oils and minor components Additive 1, Additive 2, Additive 3 and Additive 4 are the additives. Although base oils are used in larger amounts than additives , the most important characteristics of engine oils are provided by additives. Table 4.2. illustrates the concentrations (w/w%) of the components in the set of ‘Engine Oil 1’.

Table 4.2. Concentration profile of the samples in the set of 'Engine Oil 1'

Sample #	Base Oil 1	Base Oil 2	Additive 1	Additive 2	Additive 3	Additive 4
1	9.2254	7.4621	4.7447	0.1649	0.048	3.4909
2	10.0102	7.1574	4.7247	0.3377	0.033	2.8824
3	10.1811	7.5935	4.0156	0.2685	0.107	2.9318
4	9.6063	7.2268	4.6145	0.2495	0.049	3.2981
5	9.9166	6.7889	5.3365	0.1761	0.034	2.7623
6	10.4132	6.4767	5.78	0.2037	0.076	2.0868
7	10.304	6.34	4.7064	0.2297	0.099	3.471
8	9.6606	7.4597	4.8417	0.2658	0.102	2.6552
9	9.9391	7.1432	5.9693	0.1339	0.073	1.7222
10	9.3627	7.1332	6.2235	0.1563	0.048	2.1222
11	10.3296	7.0069	4.1349	0.1894	0.077	3.4109
12	10.4108	6.7548	4.8433	0.1736	0.032	2.8761
13	10.4455	7.3655	4.1986	0.2232	0.053	2.7341
14	9.4052	6.5819	6.3364	0.1553	0.041	2.8681
15	10.4456	7.1959	3.3084	0.2364	0.066	4.0505
16	9.8548	6.3782	5.1557	0.161	0.085	3.3799
17	9.2093	7.2616	6.256	0.1762	0.057	2.0546
18	10.4055	6.6892	4.2762	0.182	0.095	3.4734
19	9.1211	6.3085	4.6848	0.1702	0.1	4.6468
20	10.5853	6.6879	3.4046	0.1973	0.05	4.0031
21	9.0215	6.9363	4.7956	0.2689	0.103	3.9901
22	9.5328	7.203	5.1415	0.2169	0.036	2.8845
23	11.7258	6.3906	2.648	0.1561	0.109	3.9473
24	11.8441	6.1288	3.7853	0.2281	0.064	3.0121
25	9.8891	7.2243	4.7307	0.1233	0.089	2.9681
26	9.5332	6.8864	5.6667	0.1376	0.087	2.7139
27	9.1791	7.1447	5.8633	0.2544	0.072	2.7952
28	9.5928	6.7011	5.5615	0.16	0.046	2.8463
29	10.1335	6.9841	4.0035	0.2393	0.035	3.5575
30	10.8089	7.3128	2.7146	0.1977	0.048	3.8637
31	10.2625	6.2877	5.5932	0.1666	0.088	2.7624
32	10.3626	7.9904	4.5608	0.1691	0.052	1.935
33	10.1865	6.7952	5.8529	0.218	0.052	1.9442
34	9.5	7.4416	5.379	0.2245	0.036	2.6404
35	9.4448	7.3483	3.9993	0.2304	0.102	3.9887
36	9.971	6.7067	4.6857	0.2175	0.081	3.6352
37	10.7543	7.4237	2.8475	0.2574	0.073	3.7071

(cont. on next page)

Table 4.2 (cont.)

38	9.5284	7.1994	5.2233	0.212	0.043	2.9615
39	9.3813	7.1669	5.649	0.0859	0.045	2.8079
40	10.4396	5.9521	6.5059	0.1868	0.082	2.0515
41	10.8096	6.716	4.2062	0.2015	0.046	2.9243
42	10.2879	6.8987	3.4693	0.2466	0.066	4.127
43	10.3423	7.208	4.4162	0.2417	0.053	2.9217
44	11.8253	6.7656	4.1847	0.1827	0.05	2.4859
45	10.4835	6.3613	5.6059	0.2661	0.059	2.2247
46	9.6576	7.6295	3.0499	0.2006	0.038	4.6325
47	10.624	6.5913	3.7298	0.2604	0.072	4.2318
48	8.7276	6.6579	5.8894	0.1138	0.051	3.7662
49	10.1093	7.5459	4.7382	0.2361	0.139	2.6324
50	11.0442	6.5534	4.0698	0.2551	0.116	3.366
51	9.7463	6.0609	5.8073	0.2186	0.115	3.3888
52	9.5434	7.2465	5.5722	0.1718	0.146	2.6833
53	9.812	6.0661	5.488	0.2622	0.078	3.4827
54	9.625	6.8756	5.8961	0.1894	0.113	2.3744
55	9.9635	7.2987	3.788	0.1846	0.042	3.7417
56	10.007	7.1177	3.9044	0.2634	0.145	3.695
57	9.8796	7.0887	4.0905	0.2617	0.09	3.886
58	9.3069	7.4482	5.8001	0.2204	0.116	2.4468
59	9.8284	6.5569	5.3451	0.2595	0.027	2.9701
60	10.3265	6.9158	4.7018	0.2759	0.062	3.0492

During the preparation period, components were weighed according to the engine oil formulation and put into a glass beaker . Base oils were flowing liquid with very low viscosity but dditives were very thick liquidss with high viscosity so after the components were added into the glass beakers , each beaker was heated , sample was mixed using a magnetic stirring bar and homogeneous mixtures were had .Then samples were poured into the plastic falcon tubes. They were stored in a dark and dry cupboard.

In the second part of this project, a new training set of another engine oil which contains 50 new training samples were prepared in the laboratory. This new engine oil was coded as ‘Engine Oil 2’. In contrast, this new original oil has 5 components and formulation is distinctly different from ‘Engine Oil 1’. Table 4.3 shows components and their concentration ranges used in 50 training sample. (All concentrations are given in grams.)

Table 4.3. Components and their concentration ranges (w/w %) used in training set of 'Engine Oil 2'

Name of the Component	(w/w%) (g/g)
Base Oil 1	3.37 - 18.90
Base Oil 2	58.07 - 81.27
Additive 1	0.32 - 11.84
Additive 2	8.98 - 22.46
Additive 2	0.03 - 0.40

In the Table 4.3. shown above, Base Oil 1 and Base Oil 2 are base oils and Additive 1, Additive 2 and Additive 3 are the additives. In both training sets, samples were categorized into two classes, calibration and independent validation sets. The first set of 60 samples has 45 calibration, 15 independent validation and the second set of 50 samples has 38 calibration, 12 independent validation samples.

4.2. Instrumentation

Mid-Infrared spectra were recorded on a Perkin Elmer Spectrum 100 FT-IR Spectrometer (PerkinElmer Inc., MA, USA) equipped with a three reflection diamond ZnSe ATR accessories, tungsten lamp as the source and FR-DTS (fast recovery deuterated triglycine sulfate detector) as the detector. Data was collected in the absorbance mode in the range of $4000\text{-}600\text{ cm}^{-1}$ by taking air as the background. Samples were delivered onto the ZnSe crystal using Pasteur pipettes individually. Resolution was fixed to 4, 16 and 32 cm^{-1} for the first training set and for the second training set it was optimized to only 4 cm^{-1} . Also, during the measurements scan number was 4. At the end, totally 3401 data points gathered for each samples.

4.3. Data Analysis

The spectra were saved as ASCII file format and then transferred to another PC after collection on the FTIR instrument . Then the text files were arranged using MS Excel (Microsoft Office Excel 2003) .After that, these text files were used to construct models with GILS by the Matlab 7.0 (MathWorks Inc, Natick, MA) to predict the

unknown concentrations of the components in the training sets. Also, for making a statistical comparison, another chemometric method, Partial Least Square (PLS) was performed by the Minitab 15 (Minitab Statistical Software, Release 15 for Windows, State College, Pennsylvania).

PLS has the approach to factor construction that provides the description of available data using minimum number of adjustable parameters and, consequently, maximum precision and stability of regression model. However, inclusion of excessive factors in the model increases the accuracy of description but may decrease the predictivity as model starts to represent not just the true pattern of relation between descriptors and activity but also random noise and individual features of the training set. Because of this, during construction of the model its predictivity is monitored after including each successive factor by means of cross-validation procedure. Due to the need of using cross validation, in this study, this procedure was used for PLS and the optimal number of PC's was fixed to 15 for each model.

CHAPTER 5

RESULTS AND DISCUSSION

All samples were analyzed by FTIR-ATR and the data were collected to predict the concentrations of the components in sets of engine oils. Each set was divided into two groups as calibration set and independent validation (prediction) set. The set of 'Engine Oil 1' has 45 calibration samples and 15 independent validation samples. Also, the set of 'Engine Oil 2' has 38 calibration and 12 independent validation samples. The samples in calibration or independent validation set were chosen randomly. Then the concentrations of all constituents were predicted using GILS. Also, PLS was used to make a statistical comparison between the results at the end of the study.

5.1. 'Engine Oil 1' Results

In FTIR spectral region, it is possible to verify that similarities are obviously seen in the spectra of six components of 'Engine Oil 1' but they have noticeable differences in the wavelength range of $1800\text{-}600\text{ cm}^{-1}$, mainly due to the absorption bands related to the additives and to hydrocarbons. The Figures 5.1 to 5.3 show the spectra of six components, Base Oil 1, Base Oil 2, Additive 1, Additive 2, Additive 3 and Additive 4 at resolutions 4, 16 and 32 cm^{-1} .

The reason for analysing all the samples in the set of 'Engine Oil 1' at different resolutions is able to determine if there is any difference in success of the models at different resolutions. Because better resolution means richer the spectral information but longer analysis time to collect the spectra. At lower resolution, spectra collection is made in a quicker way but narrow peaks on the spectrum may be lost. This prevents to get information about the small peaks on the spectrum of the components. Because of this, after the models were developed by GILS, using a statistical test method of ANOVA (Analysis of Variance) with single factor, a decision was made whether the results are different or not at different resolutions for each component.

In Figures 5.1 to 5.3, it is clearly observable that the components of 'Engine Oil 1' have high absorbance values in the range of $3100\text{-}2800\text{ cm}^{-1}$ and $1500\text{-}600\text{ cm}^{-1}$. In

the range of 1500-600 cm^{-1} , we can observe absorptions due to hydrocarbons and from Zn, Ca and Mg salts of organic acids, such as alkylthiophosphate, sulphonate, phenolate. In the region of 2800 to 3000 cm^{-1} , the spectra of all the components show consistency with each other because they consist of strong C-H and C-C vibrations. However, the additives most typically used are calcium or magnesium sulphonates, phenates and salicylates have differentiated peaks in the 1500 and 600 cm^{-1} region of the infrared spectrum based on their different chemical structures. There is a peak around 3300 cm^{-1} which belongs to Additive 2. It is an antioxidant additive and contains amine so a peak of N-H vibrations is seen at this wavelength. Also, water, ethylene glycol and butylated hydroxy toluene (BHT), commonly used antioxidant additive components, all absorb light in the 3600 to 3400 cm^{-1} region due to the O-H functional group present in each molecule. On the spectra, it is not possible to get information about components in the range of 1800-2000 cm^{-1} because ZnSe masks this small part of mid-infrared spectra. This is only disadvantage of ZnSe crystal. Also, there are baseline shifts in the absorbance scale. This type of baseline shifts is common in reflectance spectroscopy and it occurs because of the thickness differences of the components. Lastly, it is evident that the absorption values get lower when the resolution gets higher from Figures 5.1 to 5.3. Spectral resolution is a measure of how well a spectrometer can distinguish closely spaced spectral features. In a 4 cm^{-1} resolution spectrum, spectral features only 4 cm^{-1} apart and when the resolution is increased, spectral features will be more apart can be distinguished and data matrix can be thought as smaller. When resolution goes from 4 cm^{-1} to 32 cm^{-1} , resolution decreases and some information will be lost on spectrum. Figures 5.1 to 5.3 are shown below illustrate the spectra of six components at resolutions 4, 16 and 32 cm^{-1} .

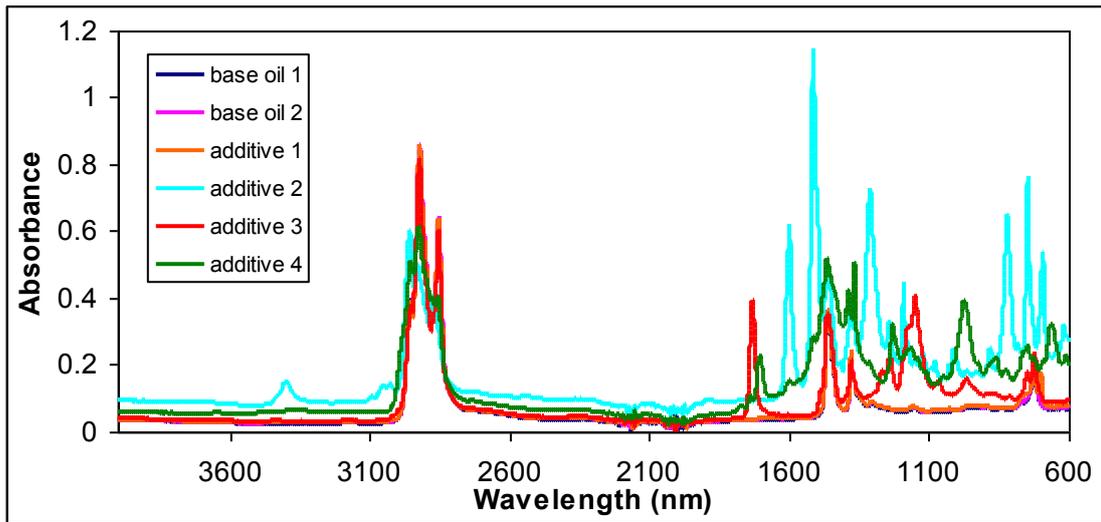


Figure 5.1. FTIR-ATR spectra of six components of 'Engine Oil 1' samples (Res=4)

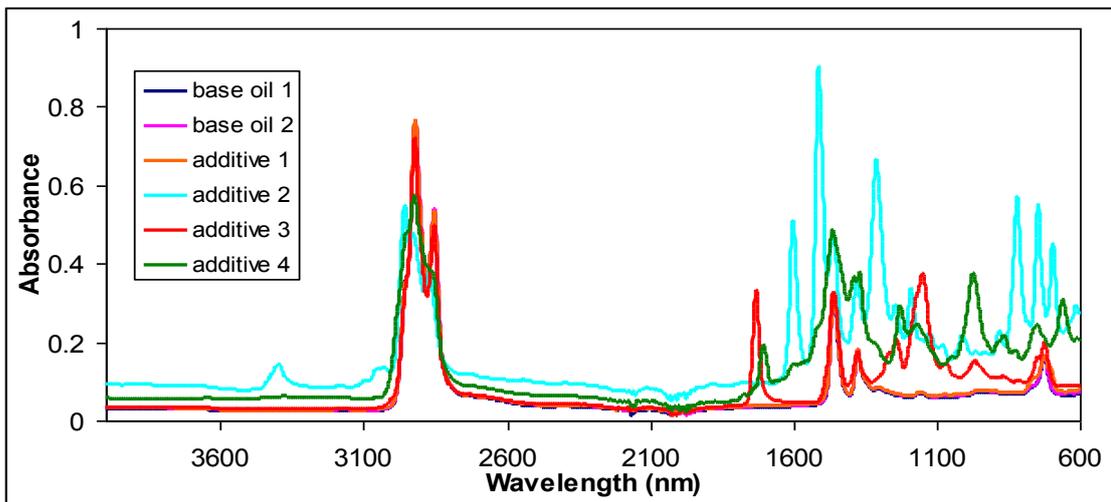


Figure 5.2. FTIR-ATR spectra of six components of 'Engine Oil 1' samples (Res=16)

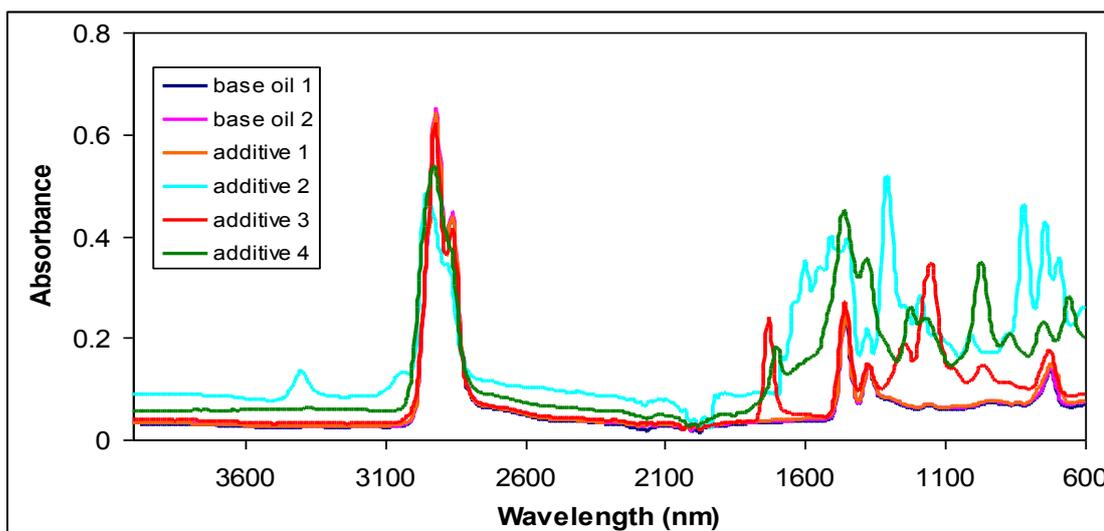


Figure 5.3. FTIR-ATR spectra of six components of 'Engine Oil 1' samples (Res=32)

In this study, a genetic algorithm based calibration method GILS and another multivariate calibration method of PLS were tested with the aim of establishing calibration models that have a high predictive capacity for the simultaneous determination of base oils and additives in two different engine oil mixtures using the FTIR-ATR spectroscopic technique.

5.1.1. GILS Results

Reference Base Oil 1 contents versus predicted values based on FTIR-ATR spectra using GILS method are shown in Figure 5.4 for the first data set. Calibration models for Base Oil 1 gave standard error of calibration with cross validation (SECV) and standard error of prediction (SEP) values as 0.616 g/g (w/w%) and 1.038 g/g (w/w%) for calibration and independent test sets, respectively when resolution was 4 cm^{-1} . When the resolution was fixed to 16 cm^{-1} , the SECV and SEP values were 0.795 g/g (w/w%) and 0.918 g/g (w/w%) and when the resolution was 32 cm^{-1} the SECV and SEP values reached 1.360 g/g (w/w%) and 1.292 g/g (w/w%). When resolution decreases, the prediction ability of the model decreases. This is because some of the information in the finger print region got lost. The R^2 values of regression lines for Base Oil 1 were 0.943, 0.906, 0.724 shown in Figure 5.4 (a), (b), (c) respectively. When compared the R^2 value for Base Oil 1, it decreased with the increasing resolution from 4

to 32 cm^{-1} . This is an expected outcome because there were more data points for resolution 4 cm^{-1} than others and this provided a more successful model for the prediction.

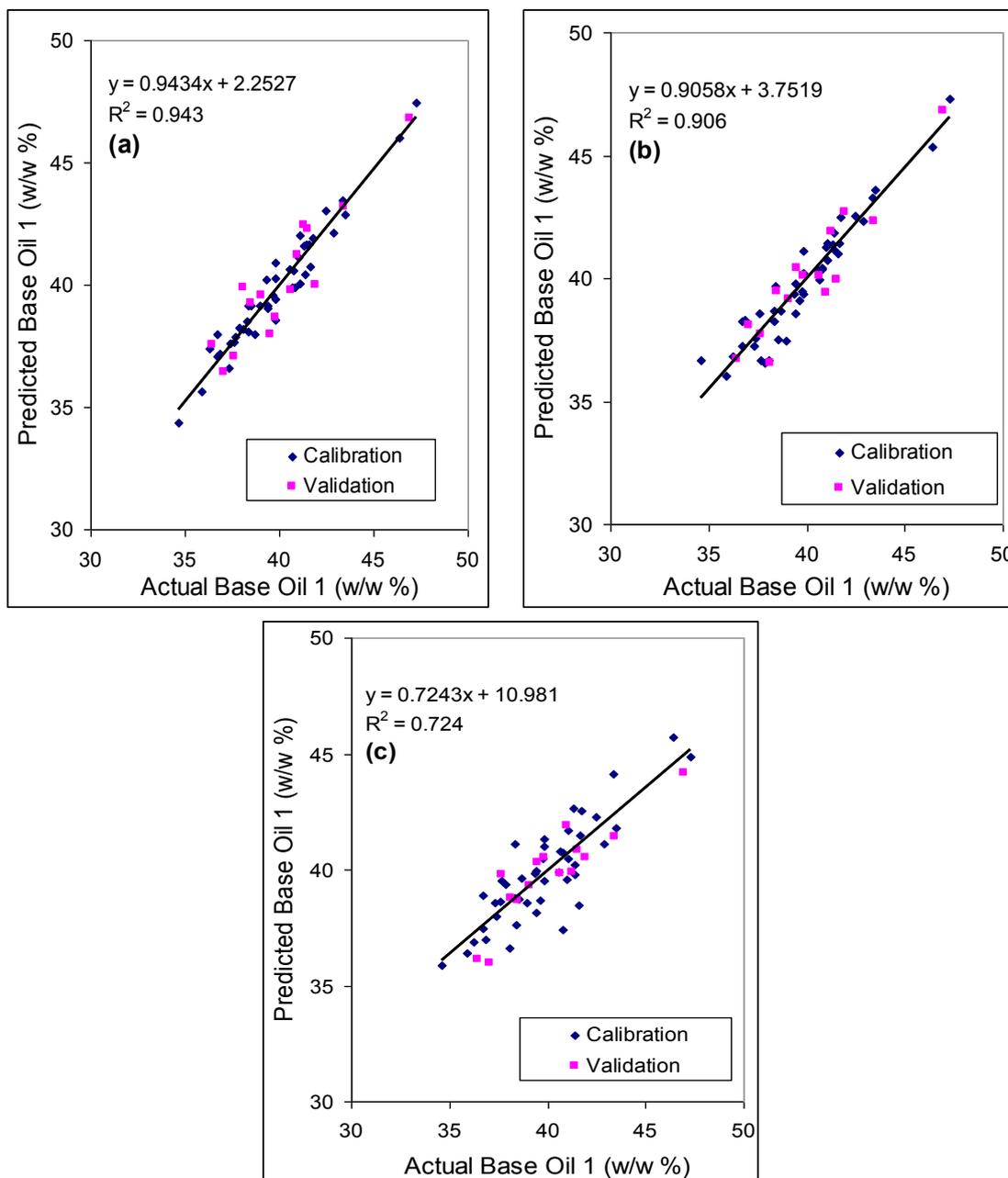


Figure 5.4. Reference vs. FTIR-ATR predicted 'Base Oil 1' contents for the data set of 'Engine Oil 1', when resolution was (a) 4 cm^{-1} , (b) 16 cm^{-1} , (c) 32 cm^{-1}

Figure 5.5 shows Base Oil 2 contents versus predicted values by GILS for the data set of 'Engine Oil 1'. The SECV values for Base Oil 2 were 0.479 g/g (w/w\%) , 0.693 g/g (w/w\%) and 0.894 g/g (w/w\%) , respectively. Also, the SEP values were 1.315

g/g (w/w%), 0.779 g/g (w/w%) and 1.468 g/g (w/w%). As it was stated in previous paragraph, for Base Oil 1, when resolution decreases, the prediction ability of the model decreases because of the loss of some of the information in the finger print region. The R^2 values of regression lines for Base Oil 2 were 0.933, 0.860 and 0.767 for the resolutions 4 cm^{-1} , 16 cm^{-1} and 32 cm^{-1} . SECV and SEP values became higher normally when the resolution got higher and thus regression became smaller. For this component, SECV and SEP values are still comparable but they are not so good. The main reason is that the FTIR spectra of base oils generally exhibits numerous bands, many of which overlap extensively. Thus, few of acidic or basic constituents that may be present in Base Oil 2 will give rise to readily identifiable bands in its spectrum so directly developing of models for the prediction is difficult.

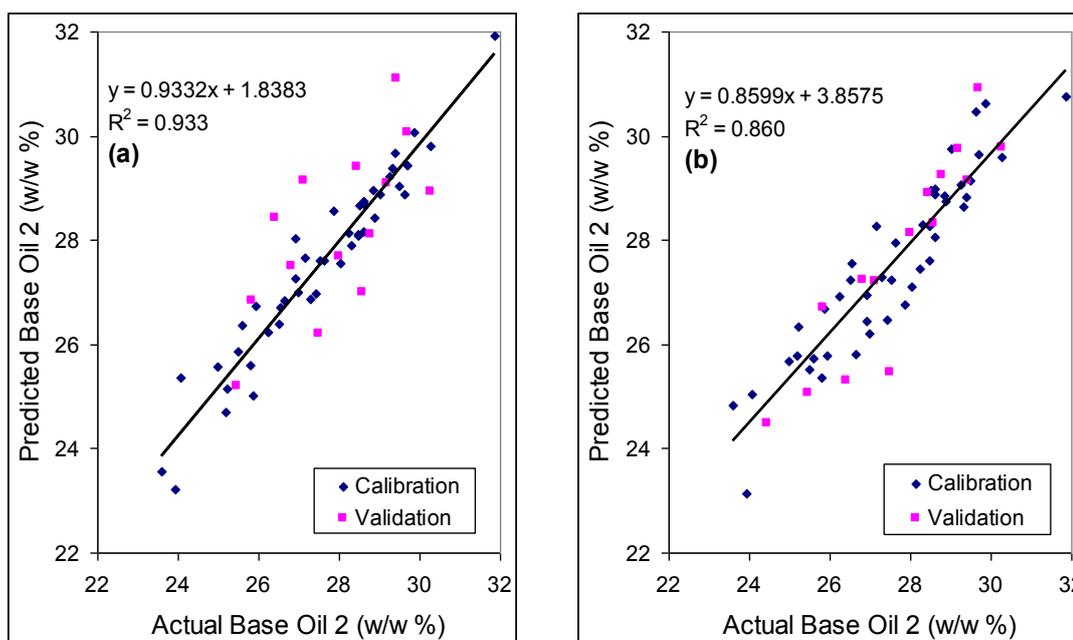


Figure 5.5. Reference vs. FTIR-ATR predicted 'Base Oil 2' contents for the data set of 'Engine Oil 1', when resolution was (a) 4 cm^{-1} , (b) 16 cm^{-1} , (c) 32 cm^{-1}

(cont. on next page)

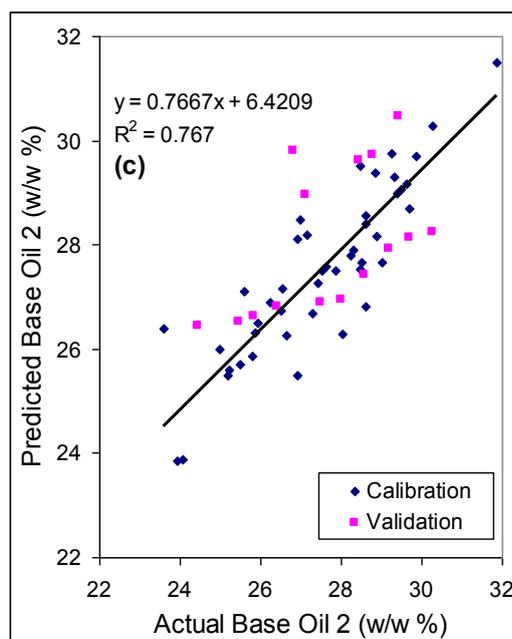


Figure 5.5. (cont.)

Figure 5.6 shows the reference versus predicted concentrations of Additive 1 content in ‘Engine Oil 1’ data set. The SECV values that are calculated for Additive 1 were 0.613 g/g (w/w%), 0.680 g/g (w/w%) and 0.477 g/g (w/w%). The SEP values for the same component were 0.697 g/g (w/w%), 0.635 g/g (w/w%) and 0.375 g/g (w/w%). R^2 values were 0.992, 0.970 and 0.985. Although Additive 1 has a very similar spectrum with the base oils spectra, its model are better than models of the base oils and generally accuracy for the Additive 1 independent validation samples are high enough at all resolutions. In fact, Additive 1 shows many different characteristic bands than base oils on FTIR-ATR spectra and this differentiation provides a fit model with high R^2 for Additive 1 content.

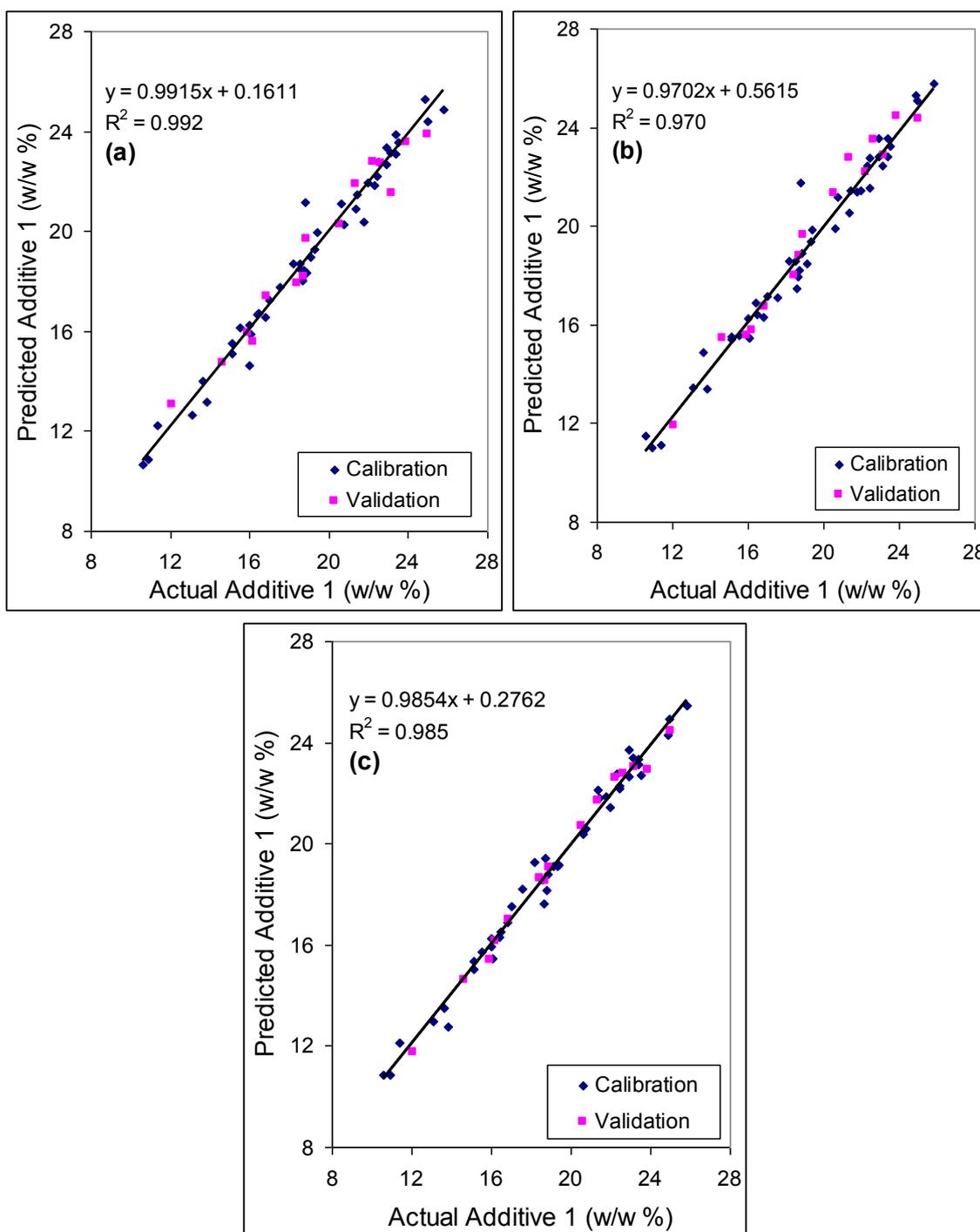


Figure 5.6. Reference vs. FTIR-ATR predicted Additive 1 contents for the data set of 'Engine Oil 1', when resolution was (a) 4 cm^{-1} , (b) 16 cm^{-1} , (c) 32 cm^{-1}

Until now, although the individual results for Base Oil 1, Base Oil 2 and Additive 1 were seen adequate, in fact they were not good enough especially for the base oils, Base Oil 1 and Base Oil 2. The main reason of this situation is that chemical structures of base oils and Additive 1 are very similar and this makes harder to build for

each components individually. For developing better models, these three components were combined to a single one and a new model was developed. Figure 5.7 below shows the new model developed at different resolutions. The SECV values were found as 0.303 g/g (w/w%), 0.321 g/g (w/w%), 0.231 g/g (w/w%) and the SEP values were calculated as 0.424 g/g (w/w%), 0.443 g/g (w/w%) and 0.286 g/g (w/w%), respectively. Also R² regression values were found as 0.989, 0.987 and 0.994. It is clearly seen that this new model (combination of Base Oil 1, Base Oil 2 and Additive 1 contents) made better predictions for the concentration than individual ones. In Figure 5.7, B1 represents Base Oil 1, B2 represents Base Oil 2 and A1 represents Additive 1.

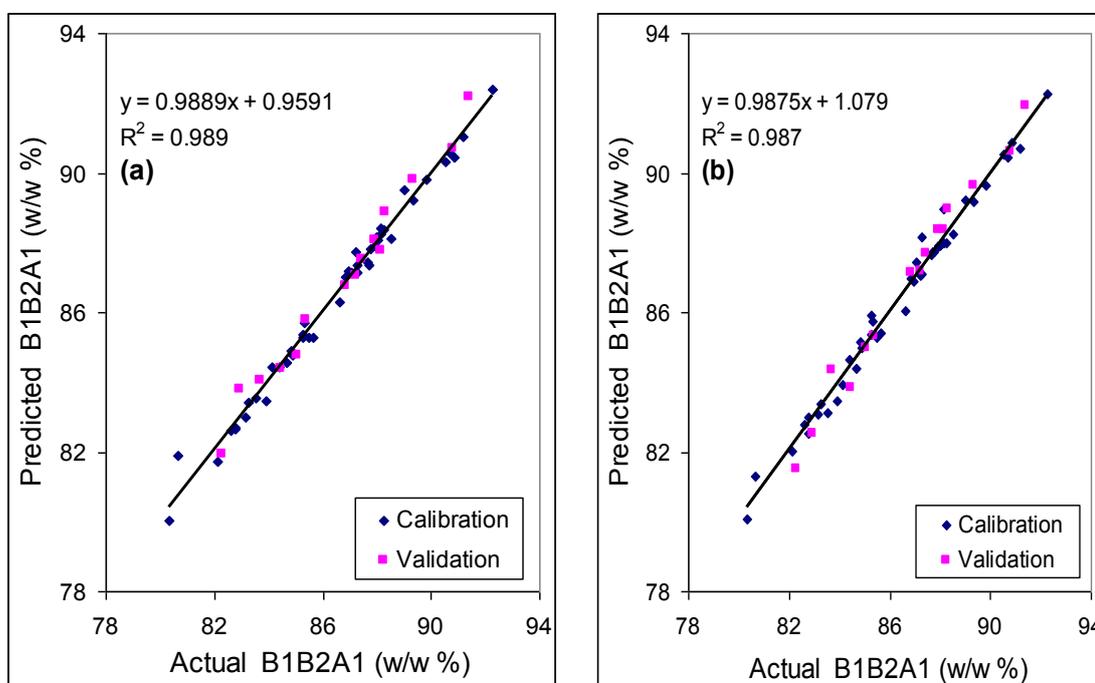


Figure 5.7. Reference vs. FTIR-ATR predicted B1B2A1 contents for the data set of 'Engine Oil 1', when resolution was (a) 4 cm⁻¹, (b) 16 cm⁻¹, (c) 32 cm⁻¹

(cont. on next page)

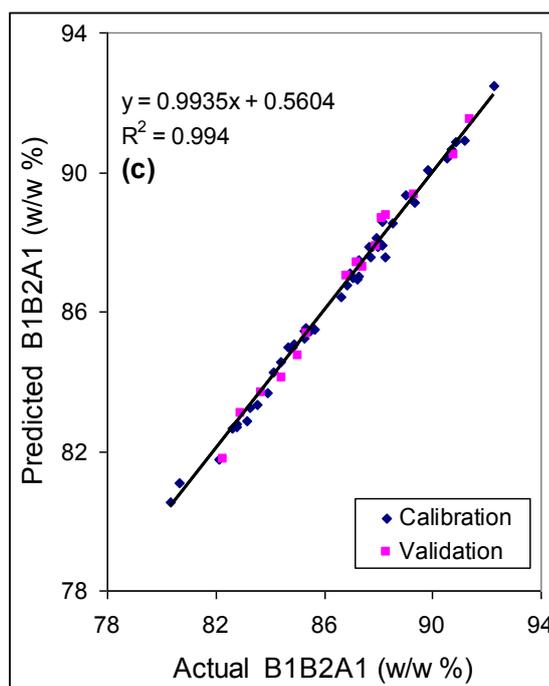


Figure 5.7. (cont.)

Figure 5.8 shows the actual versus predicted concentrations belongs to the fourth component, Additive 2. The SECV values were 0.013 g/g (w/w%), 0.010 g/g (w/w%), 0.009 g/g (w/w%) and SEP values were 0.015 g/g (w/w%), 0.015 g/g (w/w%), 0.024 g/g (w/w%). R^2 values were found as 0.996, 0.998 and 0.998. When these results were compared to other components, the prediction ability of the models for this component are better than others with lowest SECV-SEP and highest regression coefficients. This component, Additive 2 is a an antioxidant additive used in engine oils to prevent the oil from oxidizing. This kind of aminic antioxidants have noticeable absorbance bands in select regions (especially between $1600-1200\text{ cm}^{-1}$ and around 3300 cm^{-1}) of the infrared spectrum, thus enabling FTIR spectroscopy to be an ASTM method preferred means of measurement. These bands are so characteristic, they are often called fingerprint bands of antioxidant additives and its functional groups are automatically detected by the FTIR-ATR spectrometer and modeled successfully by GILS. Also, other reason of this best- fit model is that accurate results are achievable if a carefully preparation of training samples has been used to develop the model.

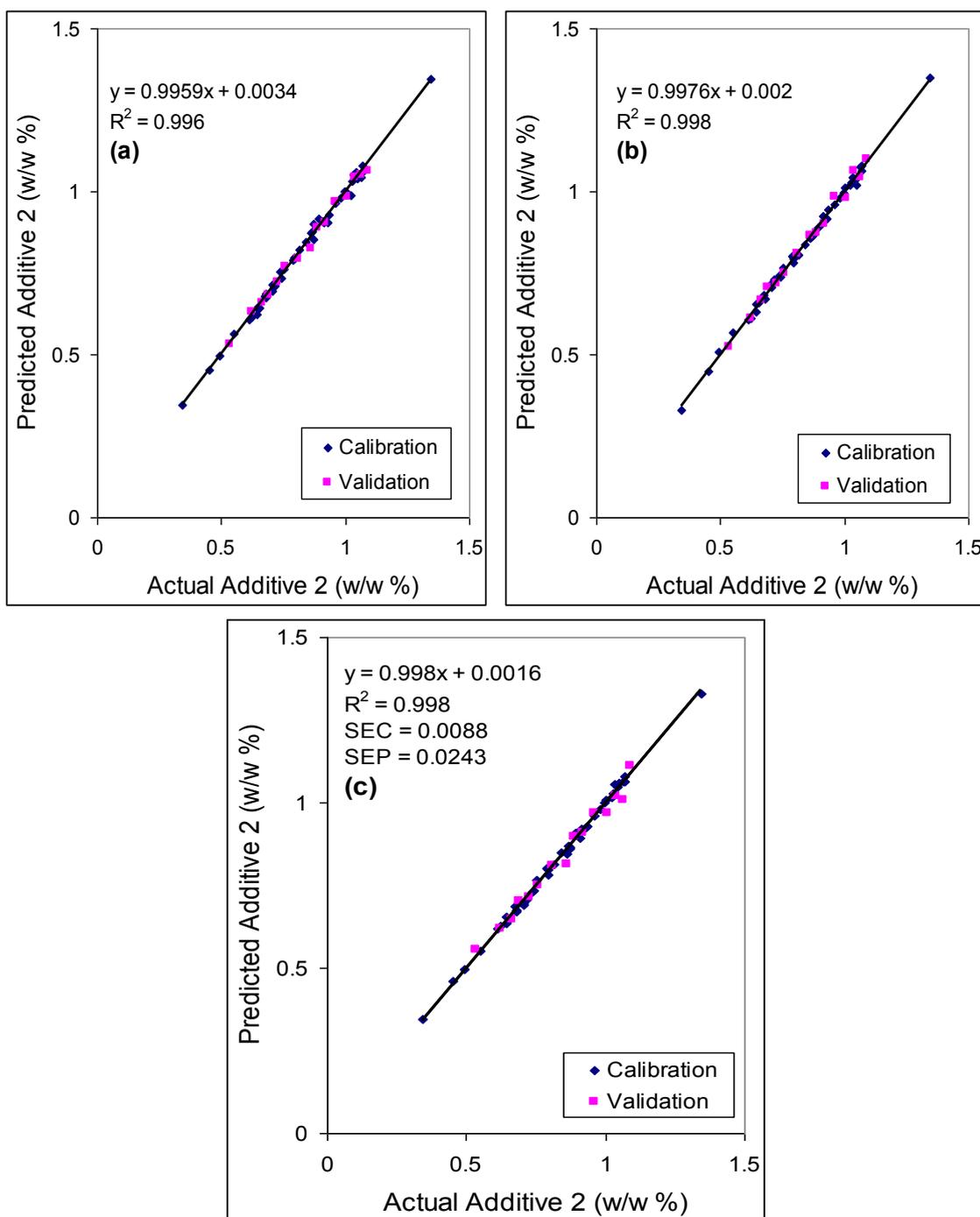


Figure 5.8. Reference vs. FTIR-ATR predicted Additive 2 contents for the data set of 'Engine Oil 1', when resolution was (a) 4 cm⁻¹, (b) 16 cm⁻¹, (c) 32 cm⁻¹

Reference Additive 3 content versus predicted values based on mid infrared spectra using GILS method are shown in Figure 5.9 for the 'Engine Oil 1' data set. Calibration models for Additive 3 content determination gave standard error of calibration with cross validation (SECV) values as 0.022 g/g (w/w%), 0.016 g/g (w/w%) and 0.021 g/g (w/w%) for calibration set. Also, it gave standard error of

prediction (SEP) 0.035 g/g (w/w%), 0.017 g/g (w/w%) and 0.017 g/g (w/w%) for independent test set, respectively. The R^2 values of regression lines for Additive 3 were 0.967, 0.982 and 0.970.

This component was coded as Additive 3 and it is a pour point depressant. It is a critical components that prevent wax fractions in the base oil from forming large crystal networks which inhibit lubricant flow at cold temperatures. Basically, it is a kind of additive which modifies viscosity of engine oil . It is very affective additive although it was used in very small quantities differently from other components. For this reason, the range of concentration (w/w %) is narrower for Additive 3 than others. This makes prediction harder. However, a successful model was obtained with very low SEP values.

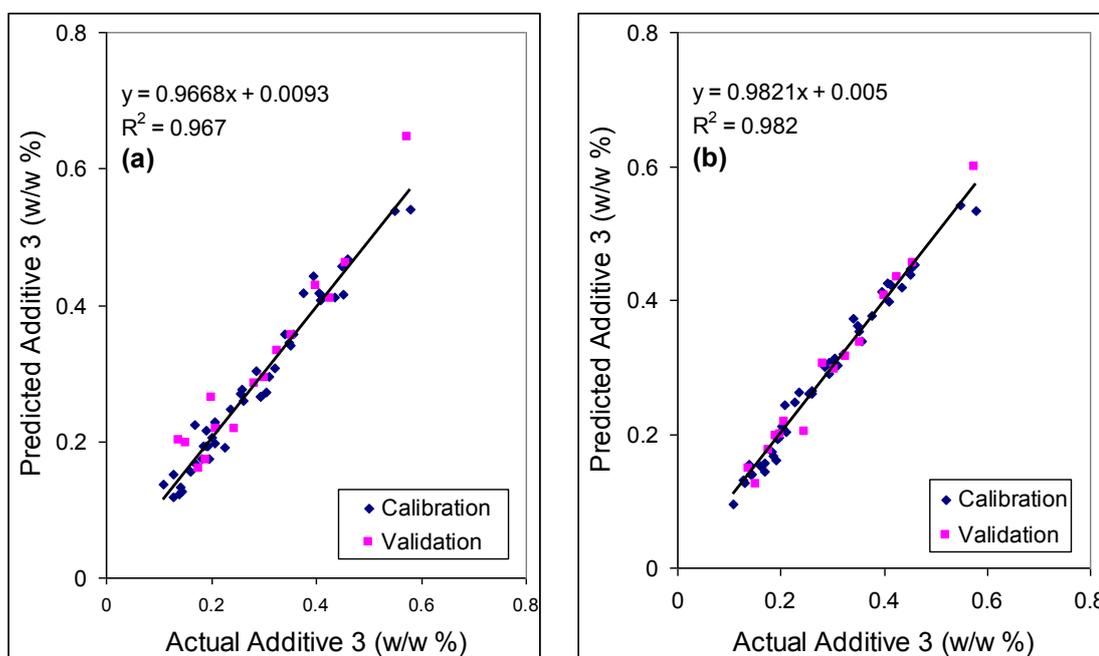


Figure 5.9. Reference vs. FTIR-ATR predicted Additive 3 contents for the data set of 'Engine Oil 1', when resolution was (a) 4 cm⁻¹, (b) 16 cm⁻¹, (c) 32 cm⁻¹

(cont. on next page)

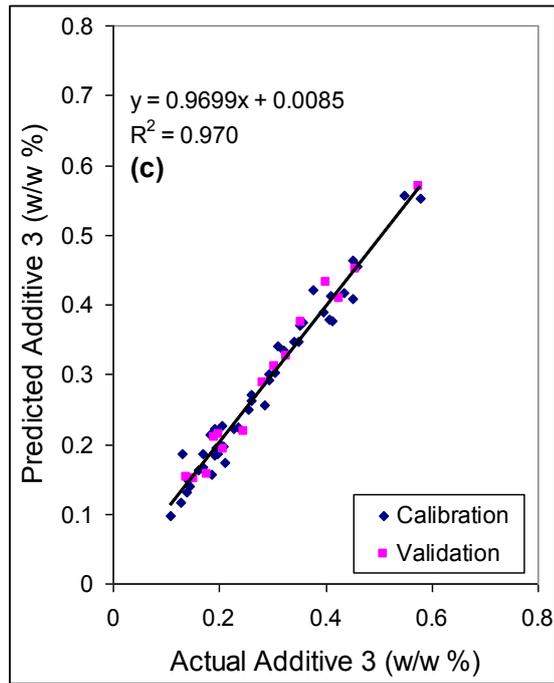


Figure 5.9. (cont.)

For the last component, Additive 4 actual versus predicted concentrations were shown in Figure 5.10. Using by GILS, the standard errors of calibration with cross validation (SECV) were found as 0.306 g/g (w/w%), 0.451 g/g (w/w%), 0.308 g/g (w/w%) and standard errors of prediction (SEP) were calculated as 0.251 g/g (w/w%), 0.311 g/g (w/w%) and 0.334 g/g (w/w%). The R^2 values of regression lines for Additive 4 were 0.988, 0.975 and 0.988 based on the resolution values. The results are good enough to make successful predictions of Additive 4 concentrations.

Additive 4 is one of the most important component used in engine oil formulations because total base number (TBN) of hydrocarbon lubricating oils are determined based on the Additive 4 content by spectroscopically. It represents the overall alkalinity contributed to an oil. It contains characteristic bands from strong C-H and C-C vibrations in the range of 1600-1500 cm^{-1} and 1100-800 cm^{-1} . The good results showed that this simple method is a successful alternative for the ASTM methods for the determination of the alkalinity of engine oils in routine oil condition monitoring and it can be used for the improvement over many methods presently available.

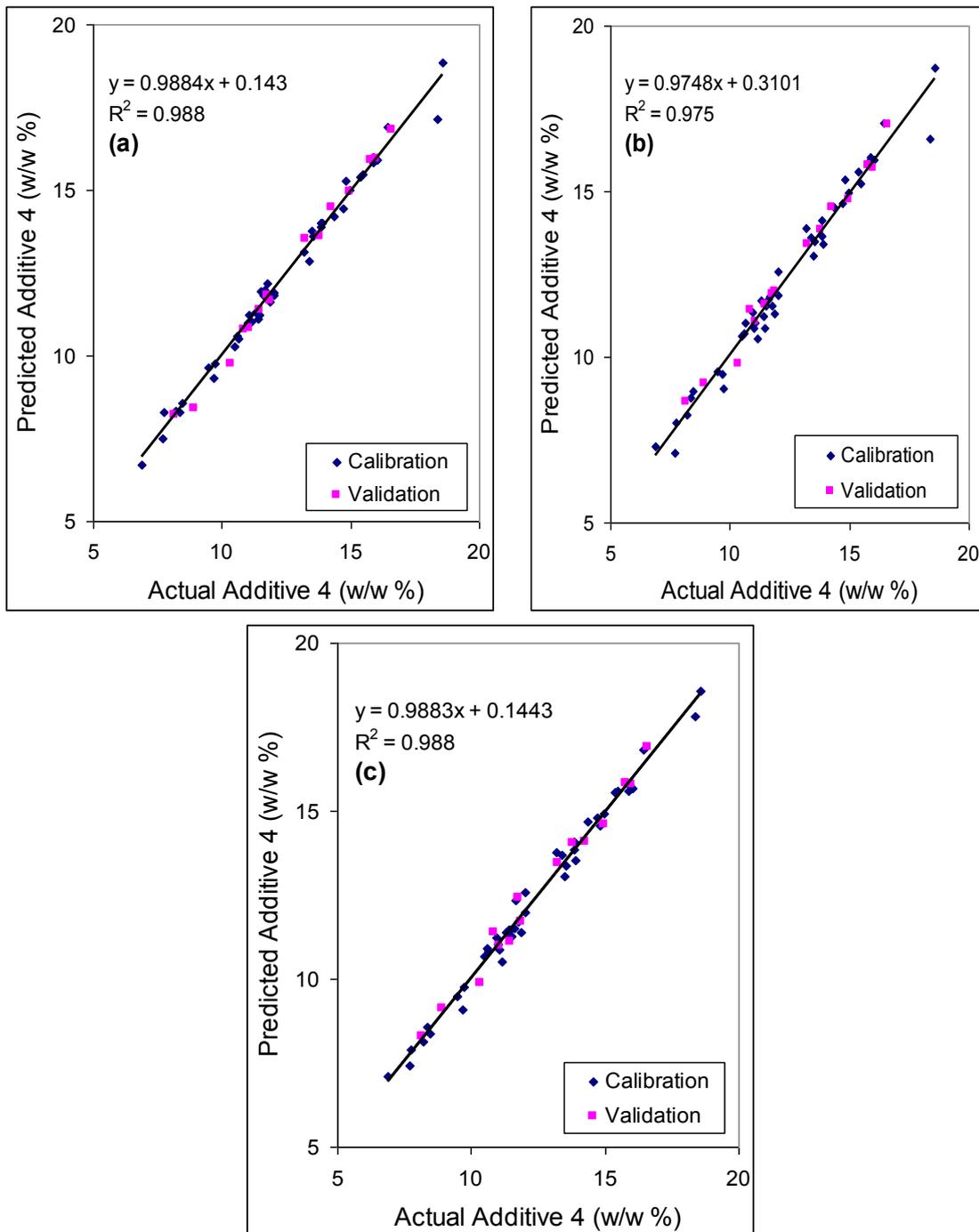


Figure 5.10. Reference vs. FTIR-ATR predicted Additive 4 contents for the data set of ‘Engine Oil 1’, when resolution was (a) 4 cm^{-1} (b) 16 cm^{-1} (c) 32 cm^{-1}

Figure 5.11 illustrates the frequency distribution of selected wavelengths in 100 runs with 30 genes and 50 iterations for the ‘Engine Oil 1’ data set when resolution was 4 cm^{-1} . As can be seen from Figure 5.11 there are a number of regions where selection frequencies are very high compared to the rest of the spectrum. The wavelength region around 600 cm^{-1} for Base Oil 1 content indicates a strong tendency for GILS method to

select while for Base Oil 2 content, around 2800 and 3000 cm^{-1} , for Additive 1 content 600 cm^{-1} is the most frequently selected region. However, for B1B2A1 content, selection was made in a wide range of wavenumbers. Also, the wavelength region around 1500 cm^{-1} for Additive 2 content indicates a strong tendency for GILS method to select while for Additive 3 and Additive 4 contents there is no strong tendency for selection.

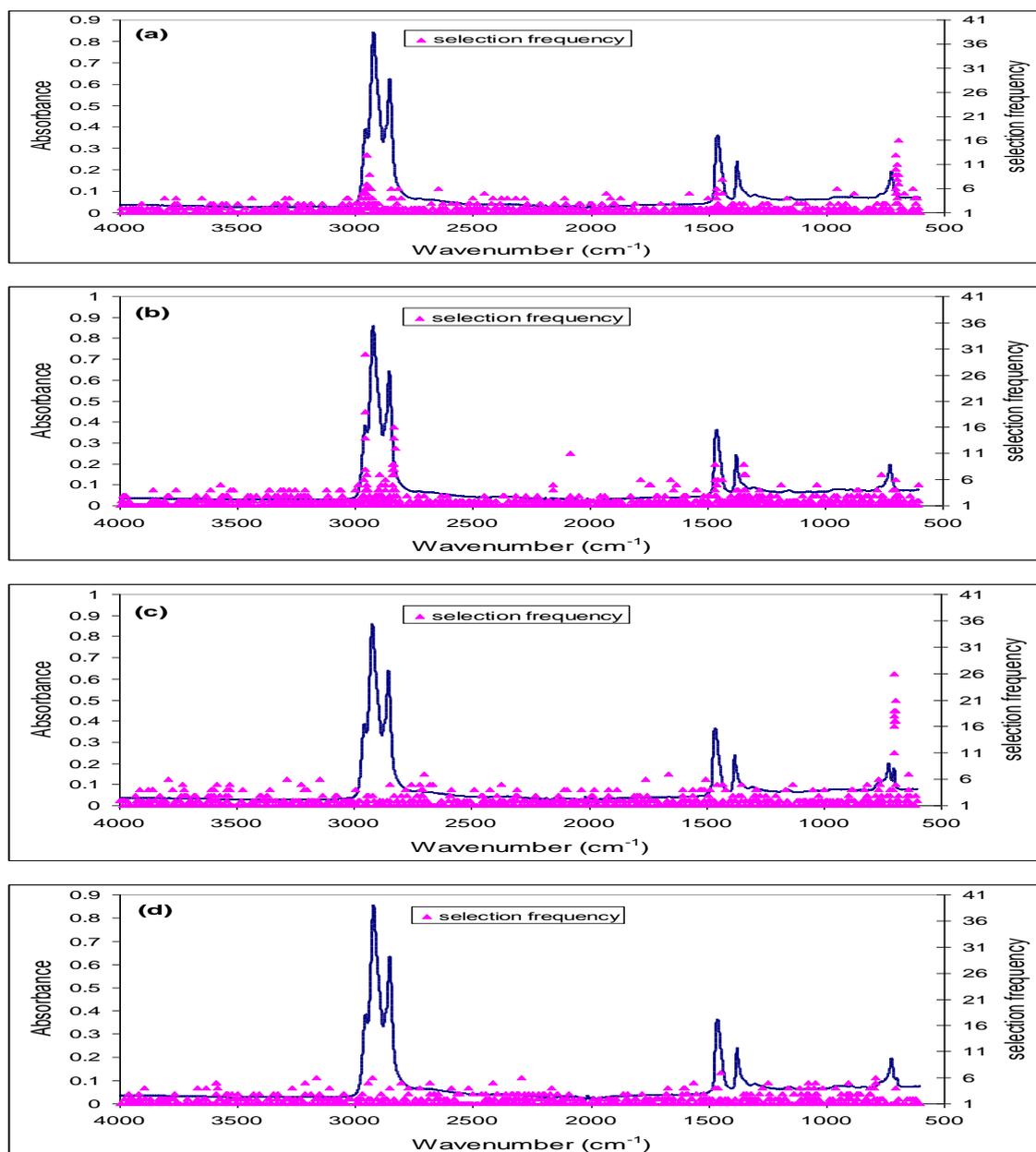


Figure 5.11. Frequency distribution of GILS selected FTIR-ATR wavelengths for (a) Base Oil 1, (b) Base Oil 2, (c) Additive 1, (d) B1B2A1, (e) Additive 2, (f) Additive 3, (g) Additive 4 contents of 'Engine Oil 1' samples when the resolution was 4 cm^{-1}

(cont. on next page)

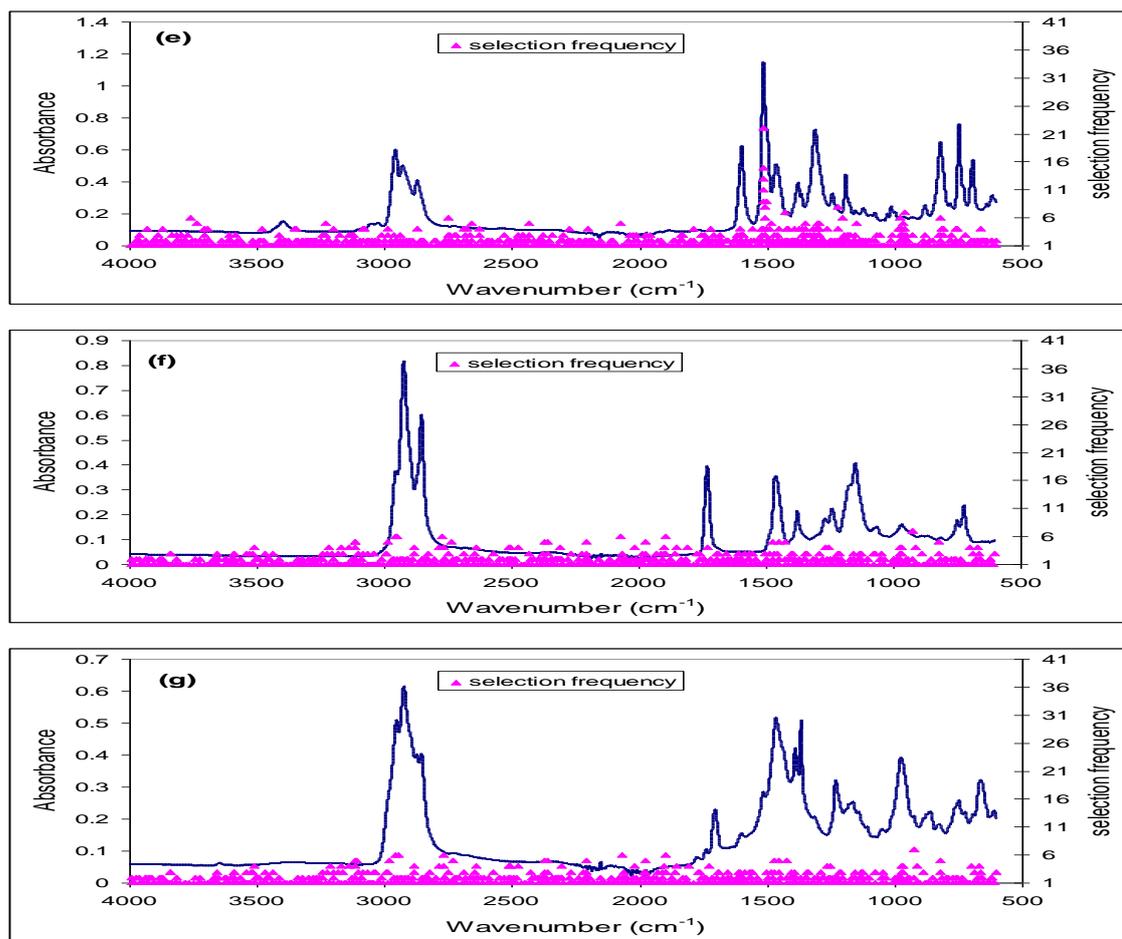


Figure 5.11. (cont.)

Figure 5.12 illustrates the frequency distribution of selected wavelengths in 100 runs with 30 genes and 50 iterations for the ‘Engine Oil 1’ data set when resolution was 16 cm⁻¹. As can be seen from Figure 5.12 the regions where selection frequencies observed are very high except for Additive 4 and B1B2A1. The wavelength region around 3000 cm⁻¹, 1500 cm⁻¹ and 700 cm⁻¹ for Base Oil 1 content indicates a strong tendency for GILS method to select while for Base Oil 2 content, around 3000 and 1500 cm⁻¹, for Additive 1 content 700 cm⁻¹ is the most frequently selected region. However, for Additive 4 and B1B2A1 contents, selection was made in a wide range of wavenumbers. Also, the wavelength region around 1300 cm⁻¹ and 1000 cm⁻¹ for Additive 2 and lastly for Additive 4 content 1200 cm⁻¹ indicates a strong tendency for GILS method to select the regions where the selection frequencies observed for all the components showed differences from that in Figure 5.11 based on the difference in resolution.

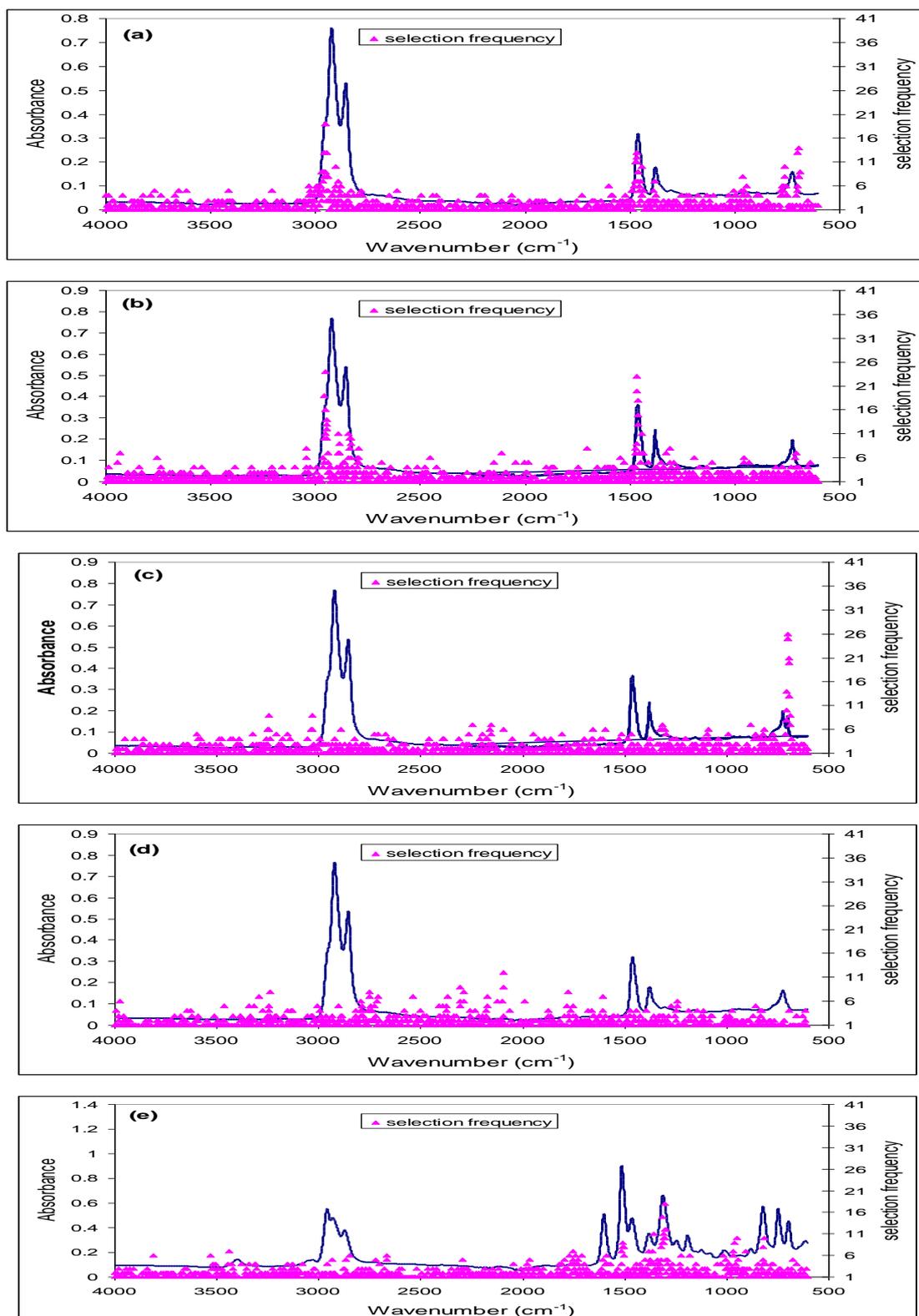


Figure 5.12. Frequency distribution of GILS selected FTIR-ATR wavelengths for (a) Base Oil 1, (b) Base Oil 2, (c) Additive 1, (d) B1B2A1, (e) Additive 2, (f) Additive 3, (g) Additive 4 contents of 'Engine Oil 1' samples when the resolution was 16 cm^{-1}

(cont. on next page)

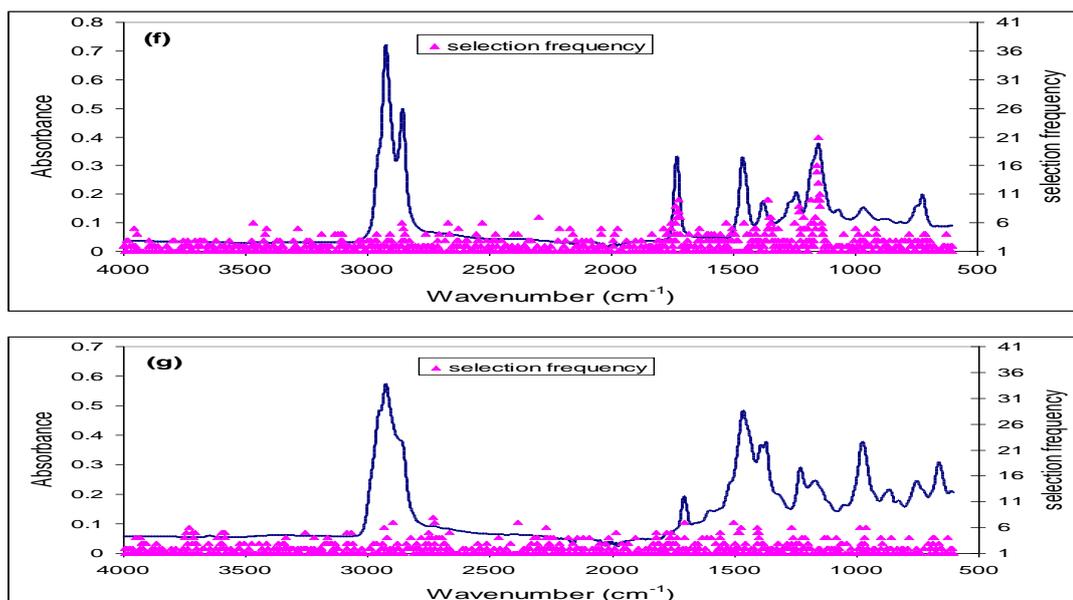


Figure 5.12. (cont.)

Figure 5.13 illustrates the frequency distribution of selected wavelengths in 100 runs with 30 genes and 50 iterations for the ‘Engine Oil 1’ data set when resolution was 32 cm^{-1} . According to the Figure 5.13, the regions of selection frequencies are more noticeable than that in the Figure 5.12. The wavelength region around 700 cm^{-1} for Base Oil 1 content indicates a strong tendency for GILS method to select while for Base Oil 2 content, around 3100 , 2000 and 700 cm^{-1} , for Additive 1 content 700 cm^{-1} is the most frequently selected region. Also, for B1B2A1 content, selection was made mostly around 1200 cm^{-1} . The wavelength region around 1300 cm^{-1} for Additive 2 content and 1200 cm^{-1} for Additive 3 indicates a strong tendency for GILS method to select while for Additive 4 contents there is no strong tendency for selection.

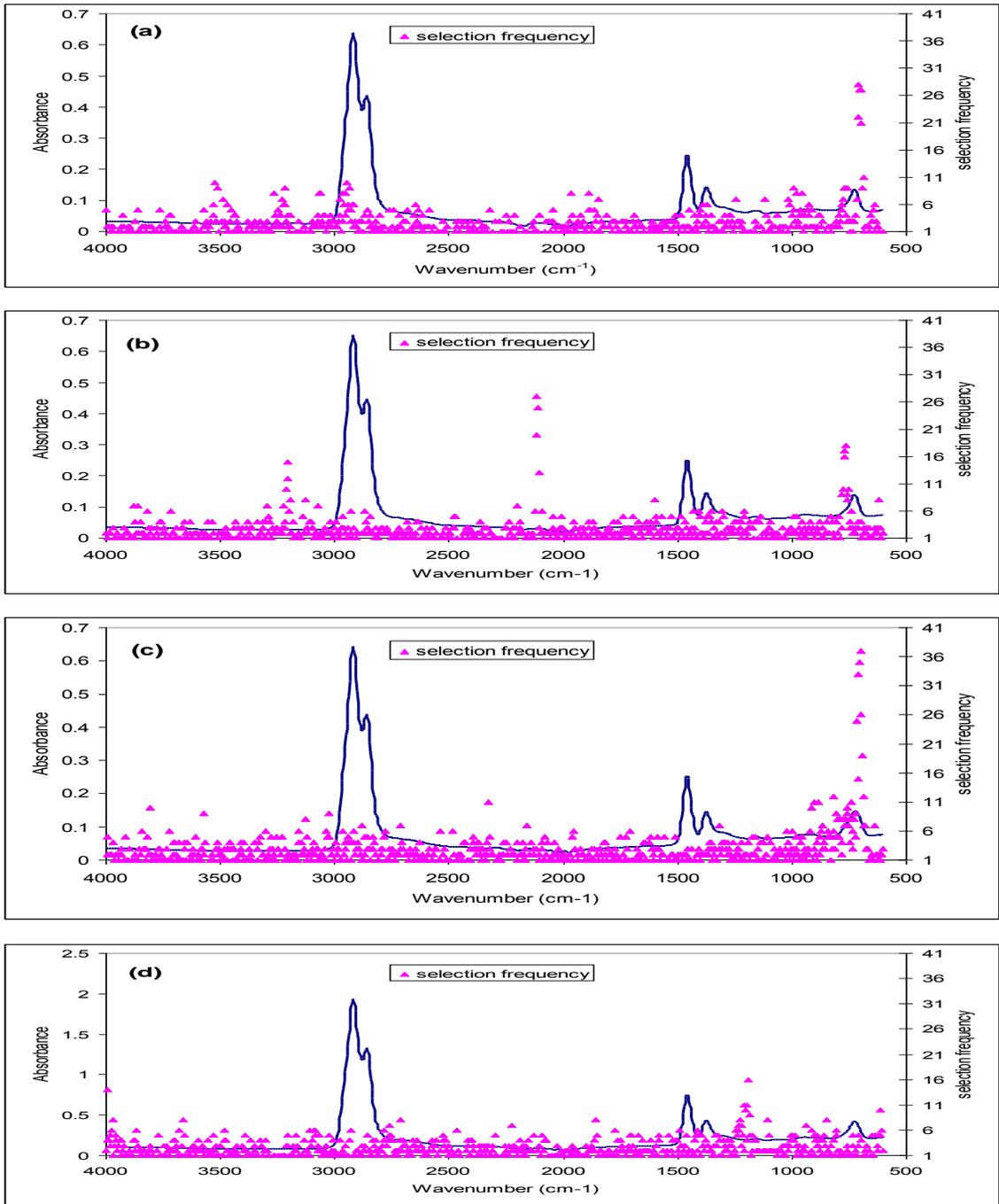


Figure 5.13. Frequency distribution of GILS selected FTIR-ATR wavelengths for (a) Base Oil 1, (b) Base Oil 2, (c) Additive 1, (d) B1B2A1, (e) Additive 2, (f) Additive 3, (g) Additive 4 contents of 'Engine Oil 1' samples when the resolution was 32 cm⁻¹

(cont. on next page)

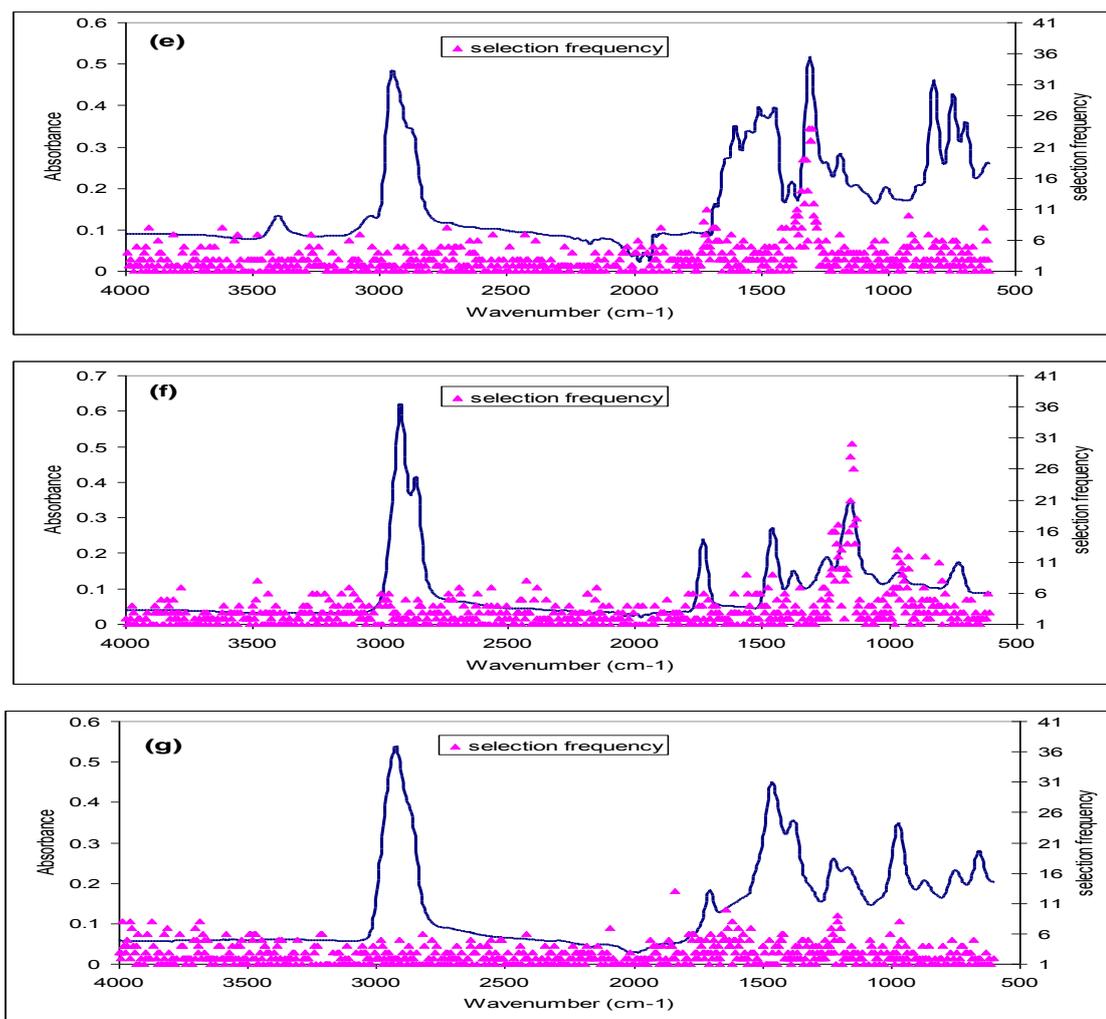


Figure 5.13. (cont.)

After GILS was performed for all the components at different resolutions, one-way ANOVA (Analysis of Variance) was used to make a comparison between the GILS models. ANOVA is a statistical test of whether or not the means of several groups are equal, and therefore generalizes t-test to more than two groups. According to ANOVA, there are two possible ways. The first one is the null hypothesis and it says that the means are equal: $H_0: \text{Mean}_1 = \text{Mean}_2 = \text{Mean}_3$. The other way is the alternative hypothesis which is that at least one of the means is different: H_a : At least one of the means is different. A test result (calculated from the null hypothesis and the sample) is called statistically significant if it is deemed unlikely to have occurred by chance, assuming the truth of the null hypothesis. A statistically significant result (when a probability (p-value) is less than a threshold (significance level)) justifies the rejection

of the null hypothesis. Table 5.1 shows the ANOVA results for the data set of ‘Engine Oil 1’.

Table 5.1. ANOVA Results for the data set of ‘Engine Oil 1’ ($\alpha = 0.05$)

Name of the Component	F	P-value	Fcrit
Base Oil 1	0.05	0.99	2.77
Base Oil 2	0.10	0.96	2.77
Additive 1	0.02	1.00	2.77
B1B2A1	0.01	1.00	2.77
Additive 2	0.01	1.00	2.77
Additive 3	0.04	0.99	2.77
Additive 4	0.01	1.00	2.77

According to ANOVA results, $F < F_{crit}$ and $P\text{-value} > \alpha$ for the components so it is unable to reject the Null Hypothesis and the means are same. This means that there is not a significant difference between the data for any component at different resolutions.

5.1.2. PLS Results

PLS is a kind of factor based methods that combines ILS approach using factor data. PLS assumes that the error can be derived from both absorbance readings and from the measurement of component concentration. Like ILS, PLS does not require the user to prepare a calibration standard where all of the interfering species are known. Also, it tries to find the factors which have the greatest relevance for prediction.

In this study, PLS was performed for both sets of ‘Engine Oil 1’ and ‘Engine Oil 2’ using the data in the range of 4000-600 cm^{-1} , 1500-600 cm^{-1} and 1300-800 cm^{-1} . The aim for selecting these regions is to determine the region that provides the most successful model for the prediction of concentrations. Figure 5.14 illustrates PLS results for the set of ‘Engine Oil 1’ in the range of 4000-600 cm^{-1} when resolution was 4 that was shown below. According to Figure 5.14, the models for B1B2A1 and Additive 4 have high prediction ability of concentrations with high regression coefficients. However, for other components the models are suffering from high collinearity and

cannot be trusted for producing a high calibration quality and for making prediction of concentrations in the independent validation samples.

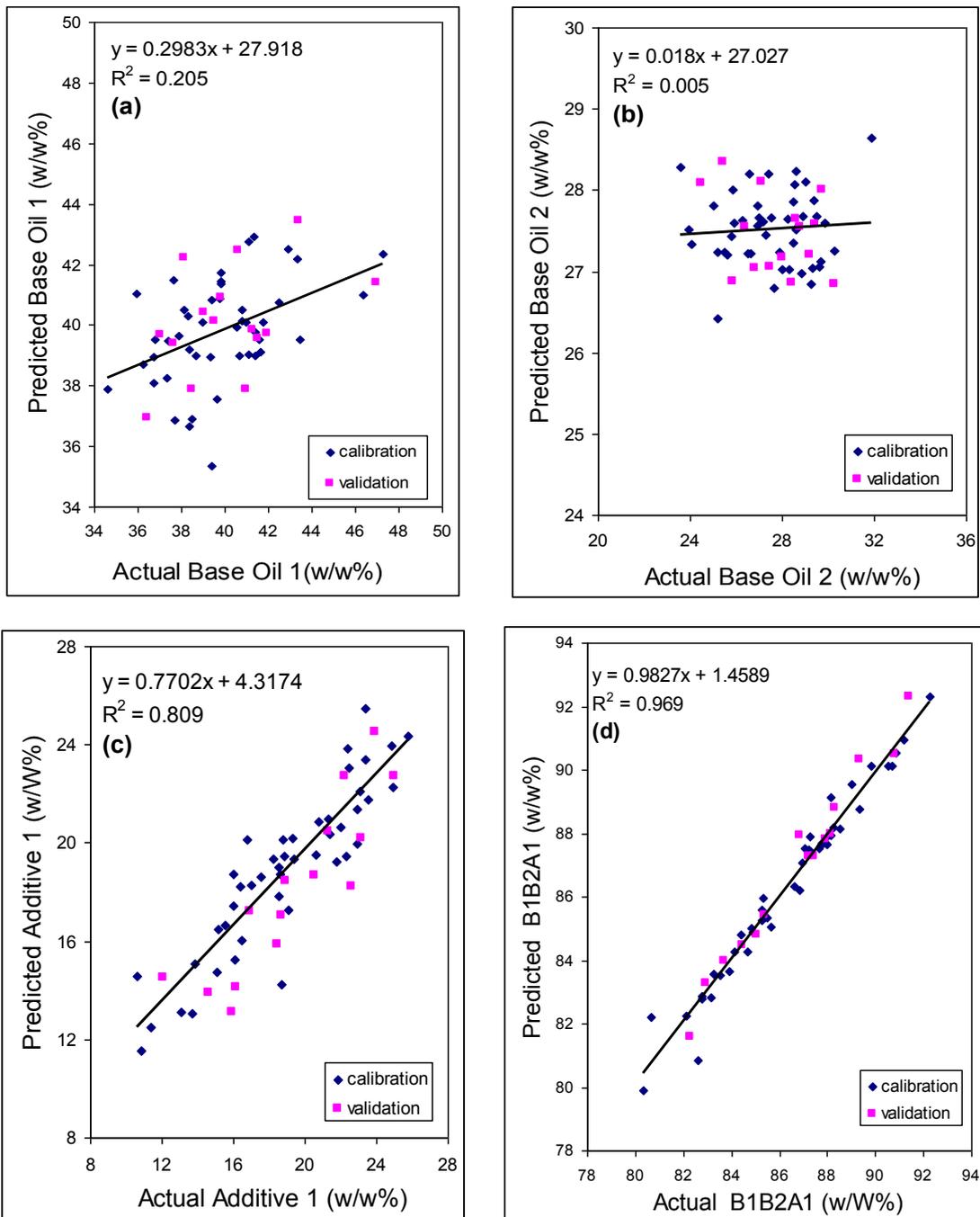


Figure 5.14. PLS Results for the set of 'Engine Oil 1' in the range of 4000-600 cm^{-1} (Res=4)

(cont. on next page)

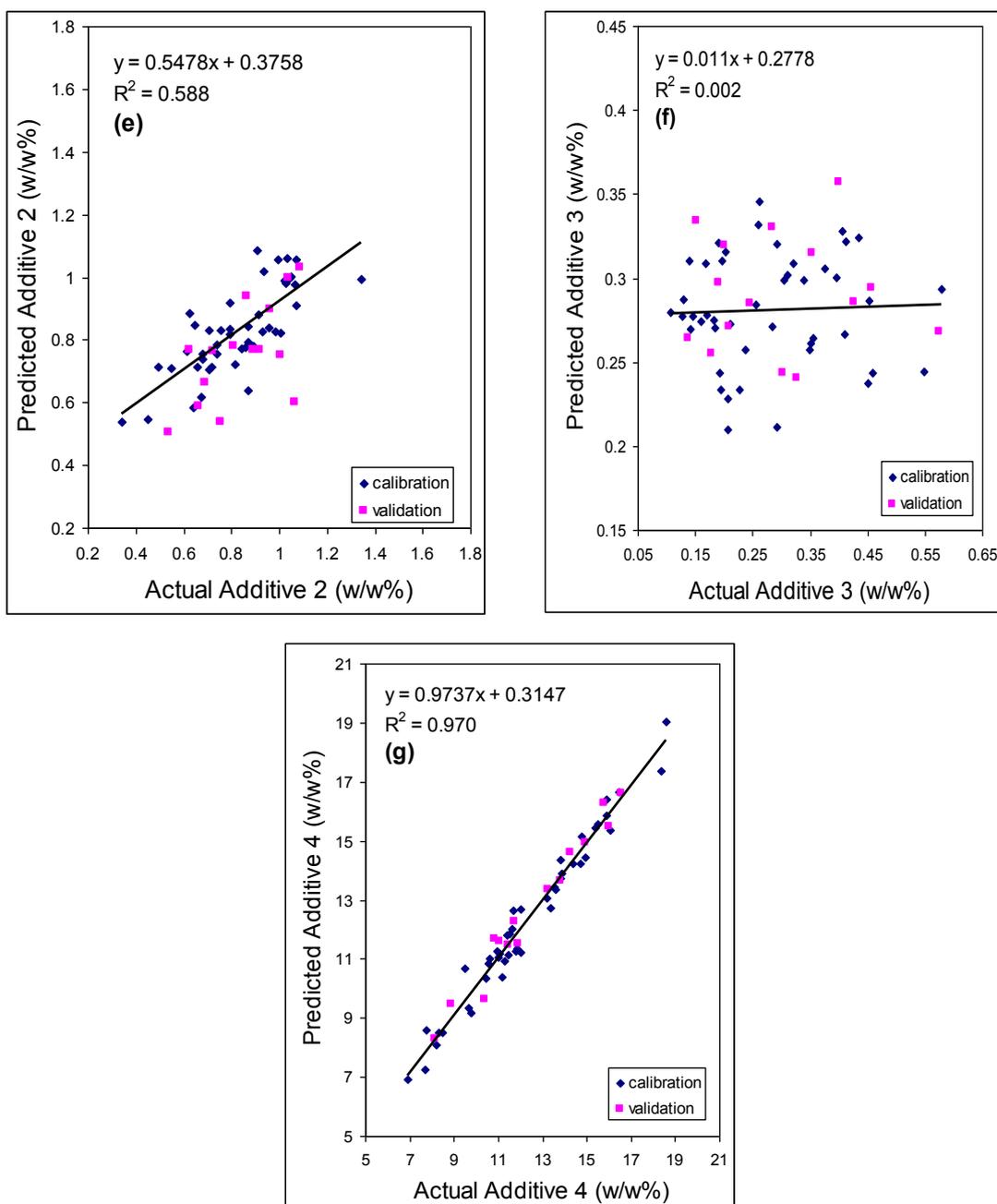


Figure 5.14. (cont.)

In theory, PLS should also be able to calibrate on the unsubtracted spectra, rather than the differential spectra, taking out the spectral contributions of the base oil and its respective additive packages without significant deterioration in quantitation. In practice, however, this was not found to work, apparently because our data set was too small to model the calibration adequately, most of the factors being used by PLS to strip out spectral variability of the oils and additive packages, leaving little information on which to model components individually. Hence it would appear that spectral

subtraction is required to eliminate the collinearity and unwanted information that would prevent the useful calibration. For this reason, the models shown below were obtained using the subtracted spectra in the range of 1500-600 cm^{-1} for the set of the 'Engine Oil 1'.

In the mid infrared region, the wavenumber range of 1500-600 cm^{-1} consists of many characteristic bands of lubricating oils especially belongs to the additives such as calcium or magnesium sulphonates, phenates and salicylates. In this spectral region, PLS was used to examine the prediction ability of the models. Figure 5.15 shows the models developed for the components in 'Engine Oil 1' using the subtracted data in mid infrared region .

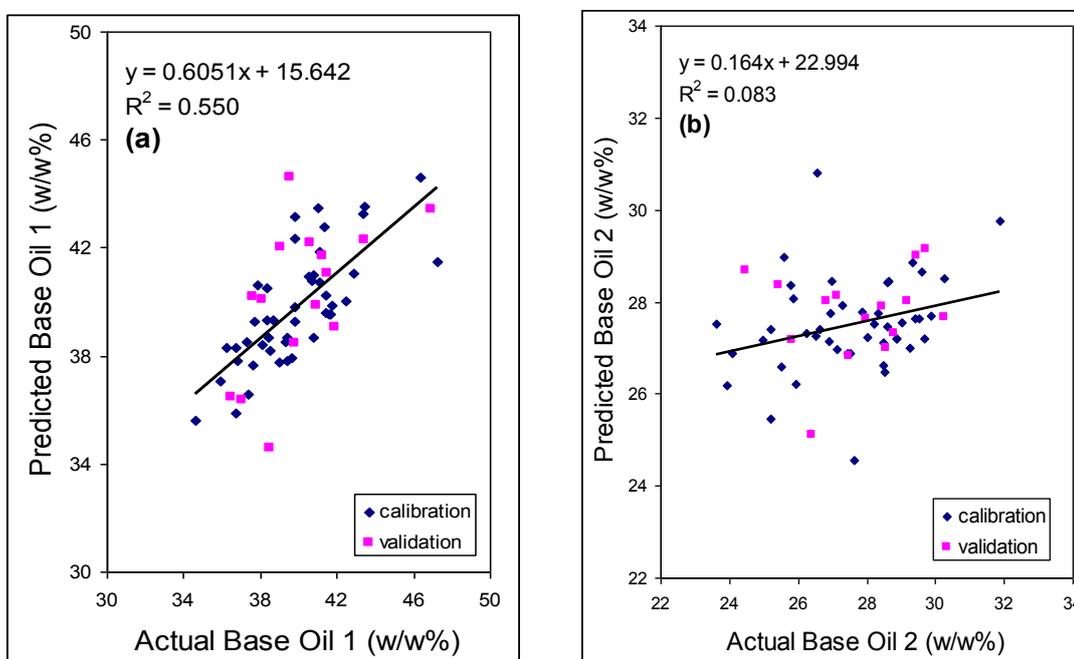


Figure 5.15. PLS Results for the set of 'Engine Oil 1' in the range of 1500-600 cm^{-1} (Res=4)

(cont. on next page)

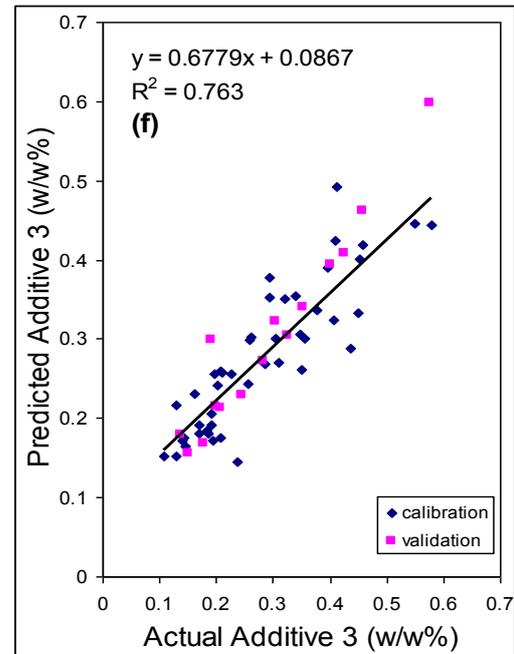
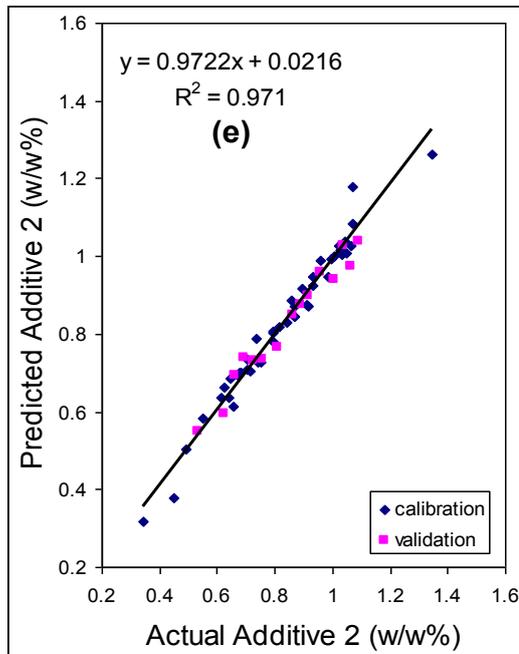
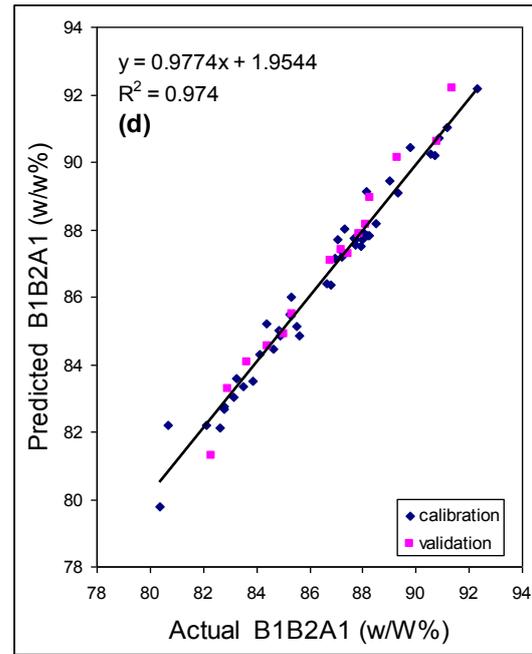
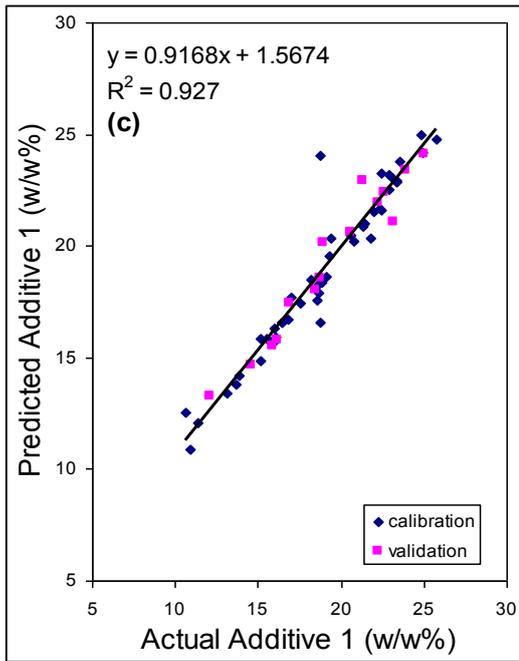


Figure 5.15. (cont.)

(cont. on next page)

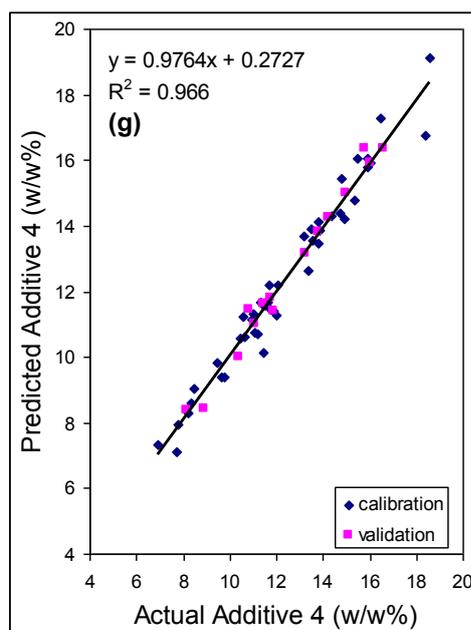


Figure 5.15. (cont.)

Restricting PLS to only 1500-600 cm^{-1} spectral region worked well for all the components except Base Oil 1 and Base Oil 2 because at this fingerprint region, the spectra of base oils overlap with each other. Like in GILS, PLS showed the unsuccessful models for base oils individually with high collinearity because of overlapping of spectral bands. However, the models developed for the additives have high prediction ability of concentration because additives have differentiated peaks in the this spectral region of the infrared spectrum based on their different chemical structures. At this point, it may be concluded that PLS requires information about all six component spectral contributions to be able to eliminate the confounding effects of the additives and component interactions.

In the third part of the study, PLS was used to develop models for the components in 'Engine Oil 1' samples in a different wavenumber region, 1300-800 cm^{-1} . Figure 5.16 illustrates the PLS results for each component in this spectral region that was shown below. Like the range of 1500-600 cm^{-1} , the spectral region of 1300-800 cm^{-1} also consists of different characteristic bands belong to the additives such as sharp peaks of Additive 3 in the region of 1000 to 950 cm^{-1} so models can be built using these observable peaks. Figure 5.16 shows the models developed for the components in 'Engine Oil 1' using the wavenumber range of 1300-800 cm^{-1} in mid infrared region .

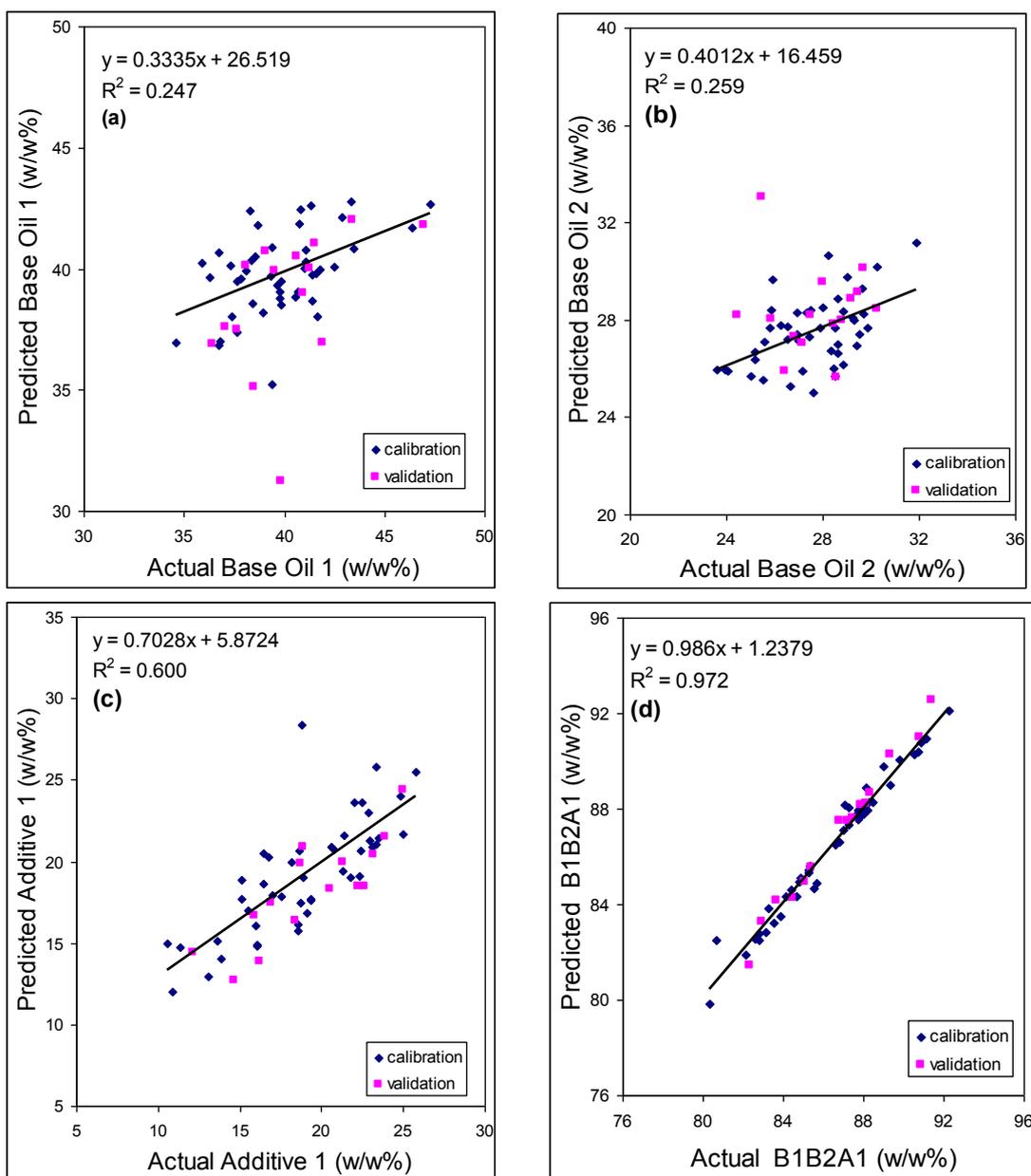


Figure 5.16. PLS Results for the set of 'Engine Oil 1' in the range of 1300-800 cm^{-1} (Res=4)

(cont. on next page)

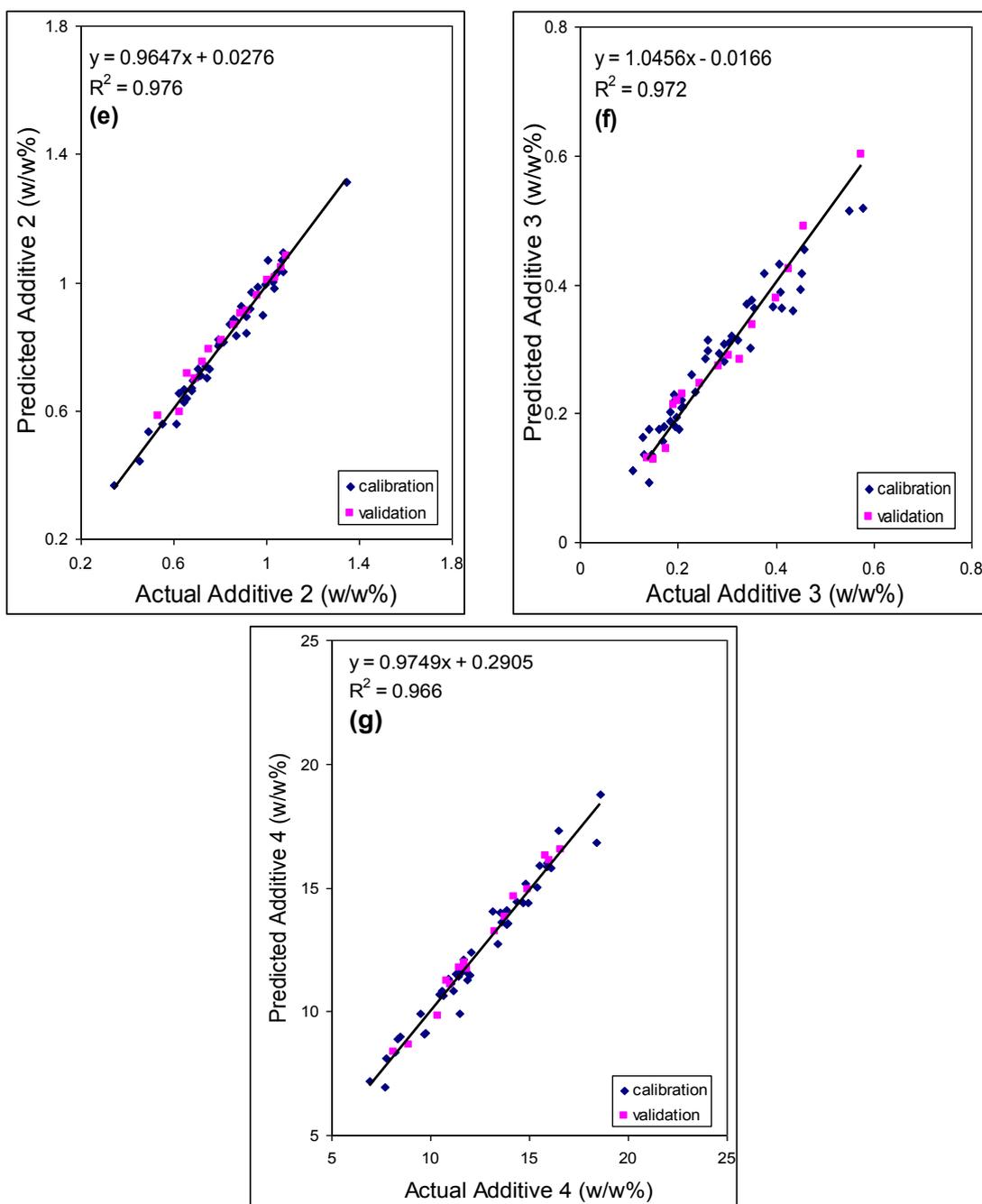


Figure 5.16. (cont.)

As it was stated about the results in the region of $1500\text{-}600\text{ cm}^{-1}$, the developed models for the Base Oil 1, Base Oil 2 and Additive 1 do not have high prediction ability of concentrations in the range of $1300\text{-}800\text{ cm}^{-1}$ because their spectra overlap with each other at that spectral region. However, the model for the combination of the Base Oil 1, Base Oil 2 and Additive 1 indicates that the optimized PLS calibration is capable of predicting the content of three components in 'Engine Oil 1' samples. Also, for all the additives it can be seen that subtracted data was sufficient to ensure accurate results

with high correlation coefficient, R^2 . This shows that although the wavenumber region got narrower, the characteristic peaks for developing models for the additives did not get lost. The region of $1300\text{-}800\text{ cm}^{-1}$ is also capable of predicting concentration of Additive 2, Additive 3 and Additive 4 contents.

For making a comparison between PLS and GILS, SECV, SEP and R^2 values are given for the data set of 'Engine Oil 1' in Table 5.2 below.

Table 5.2. Calibration summary for the data set of 'Engine Oil 1' (Res=4)

Engine Oil 1		PLS			GILS		
		SECV g/g (w/w%)	SEP g/g (w/w%)	R^2	SECV g/g (w/w%)	SEP g/g (w/w%)	R^2
Name of the Component	Base Oil 1	2.006	2.388	0.205	0.616	1.038	0.943
	Base Oil 2	1.587	1.876	0.005	0.478	1.364	0.933
	Additive 1	2.076	2.107	0.809	0.613	0.699	0.992
	B1B2A1	0.612	0.528	0.969	0.303	0.424	0.989
	Additive 2	0.153	0.165	0.588	0.013	0.015	0.996
	Additive 3	0.148	0.127	0.002	0.022	0.035	0.967
	Additive 4	0.590	0.445	0.970	0.306	0.251	0.988

For the data set of 'Engine Oil 1', the results for PLS and GILS are very different from each other. Although PLS results are suffering from lower R^2 and higher SECV and SEP values for all the components, GILS provided more successful models with accurate results to predict the concentrations of the components in 'Engine Oil 1' samples. Also, PLS is more time-consuming method than GILS because of its computational requirements. Thus, when a comparison is made, GILS is capable of determining the engine oil compositions quantitatively.

5.2. 'Engine Oil 2' Results

In this study, second set contains samples of the second engine oil coded as 'Engine Oil 2'. This is the secondly preferred engine oil which is sold by OpetFuchs Company. They were analysed by FTIR-ATR when the resolution was only 4 cm^{-1} and then for each component models were developed by GILS. They were shown in Figure 5.18 below.

5.2.1. GILS Results

According to the Figure 5.17, SECV values were found as 0.623 g/g (w/w%) for the first component, Base Oil 1. For the Base Oil 2, it was calculated as 0.674 g/g (w/w%) , 0.4029 g/g (w/w%) for the Additive 1, 0.206 g/g (w/w%) for the Additive 2 and 0.023 g/g (w/w%) for the last component, Additive 3. Also, after making prediction of concentration for the independent validation samples, standard error of prediction (SEP) values were calculated. For Base Oil 1, SEP was 0.768 g/g (w/w%). It was found as 0.579 g/g (w/w%) for Base Oil 2, 0.378 g/g (w/w%) for Additive 1, 0.519 g/g (w/w%) for Additive 2 and 0.022 g/g (w/w%) for Additive 3. Lastly, regression coefficients were found as 0.972, 0.982, 0.989, 0.996 and 0.943 for Base Oil 1, Base Oil 2, Additive 1, Additive 2 and Additive 3 respectively.

When these results were compared to the results of the ‘Engine Oil 1’ samples at the same resolution, 4 cm^{-1} , it is clearly seen that a better prediction was made for Base Oil 2, Additive 1 and Additive 3 in the set of ‘Engine Oil 2’. For other components of ‘Engine Oil 2’ samples, Base Oil 1 and Additive 2 predictions are not good enough as in ‘Engine Oil 1’. However, all these GILS results for all the components of ‘Engine Oil 2’ and ‘Engine Oil 1’ sets showed that the GILS method is able to model engine oil additives and base oil concentrations successfully using FTIR spectra of the process samples. Multivariate calibration models that are generated with GILS were component specific as observed from selection frequency plots indicating that with all the overlapping and complex nature of the FTIR spectra of the multicomponent mixtures, the GILS algorithm only focuses on the regions where the most concentration related information is contained. Also, it was very effective to extract necessary information while constructing multivariate calibration models resulting in a robust component specific modeling despite all the overlapping features in the spectra.

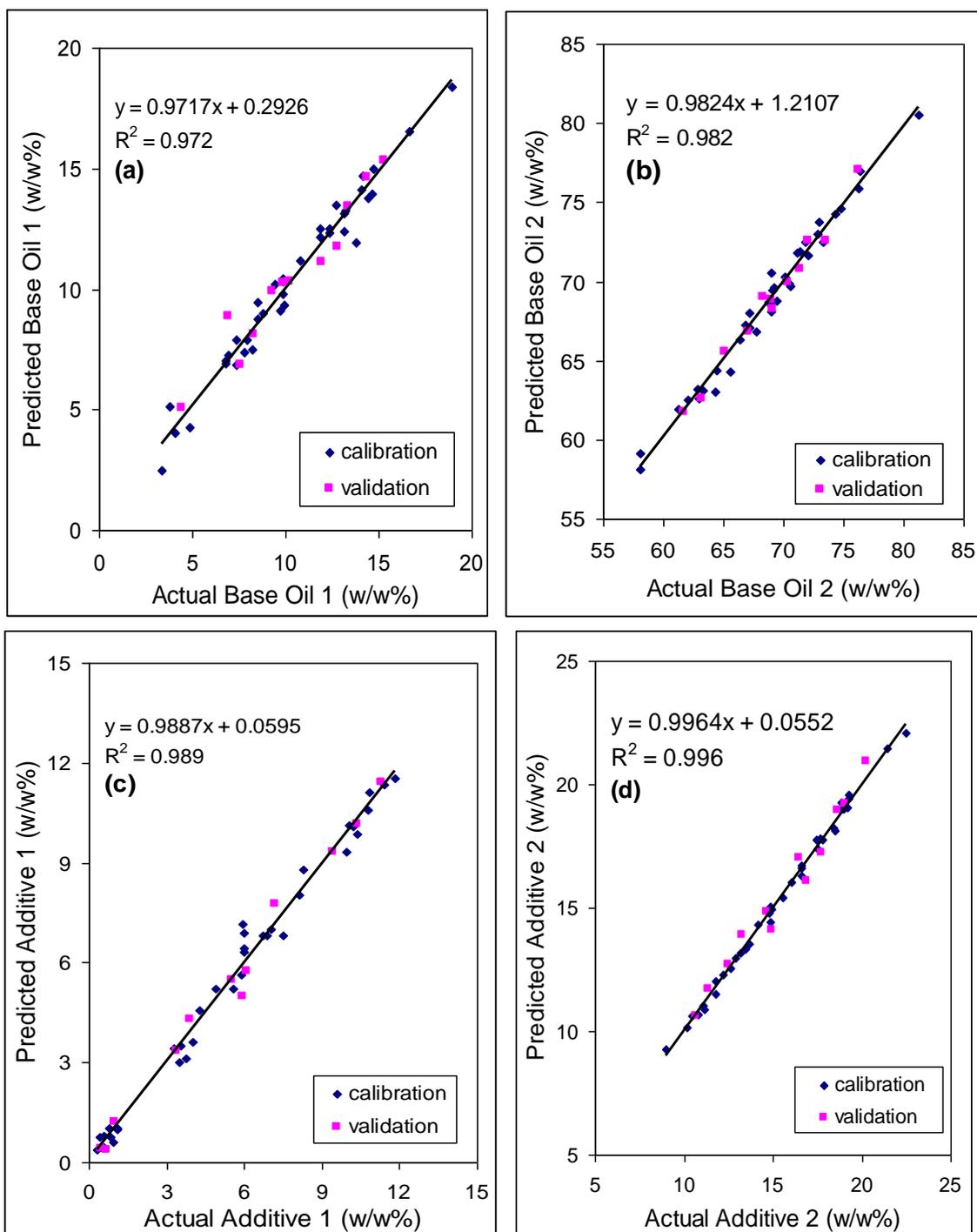


Figure 5.17. Reference vs. FTIR-ATR predicted (a) Base Oil 1, (b) Base Oil 2, (c) Additive 1, (d) Additive 2, (e) Additive 3 contents for the data set of 'Engine Oil 1', when resolution was 4 cm^{-1}

(cont. on next page)

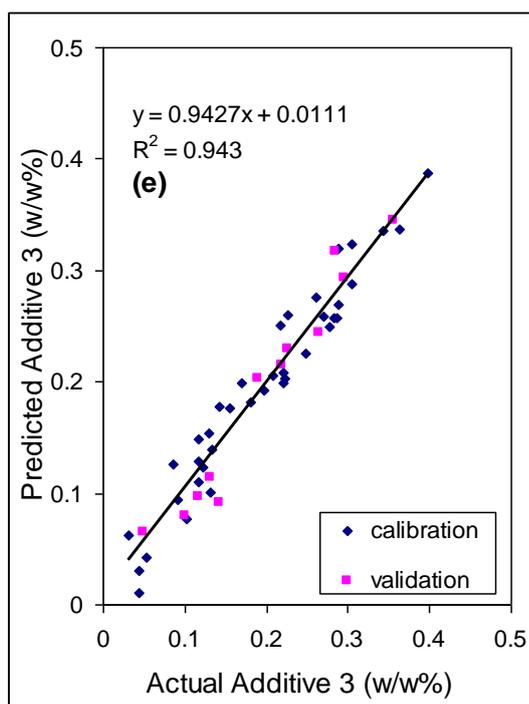


Figure 5.17. (cont.)

Figure 5.18 illustrates the frequency distribution of selected wavelengths in 100 runs with 30 genes and 50 iterations for the 'Engine Oil 2' data set. As can be seen from Figure 5.18 there are a number of regions where selection frequencies are very high compared to the rest of the spectrum. The wavelength region around 2900 cm^{-1} for Base Oil 1 content indicates a strong tendency for GILS method to select while for Base Oil 2 content around 300 cm^{-1} , for Additive 2 content around 1200 cm^{-1} , for Additive 1 content around 700 cm^{-1} and for Additive 3 content around $1000\text{-}1200\text{ cm}^{-1}$ are the most frequently selected regions.

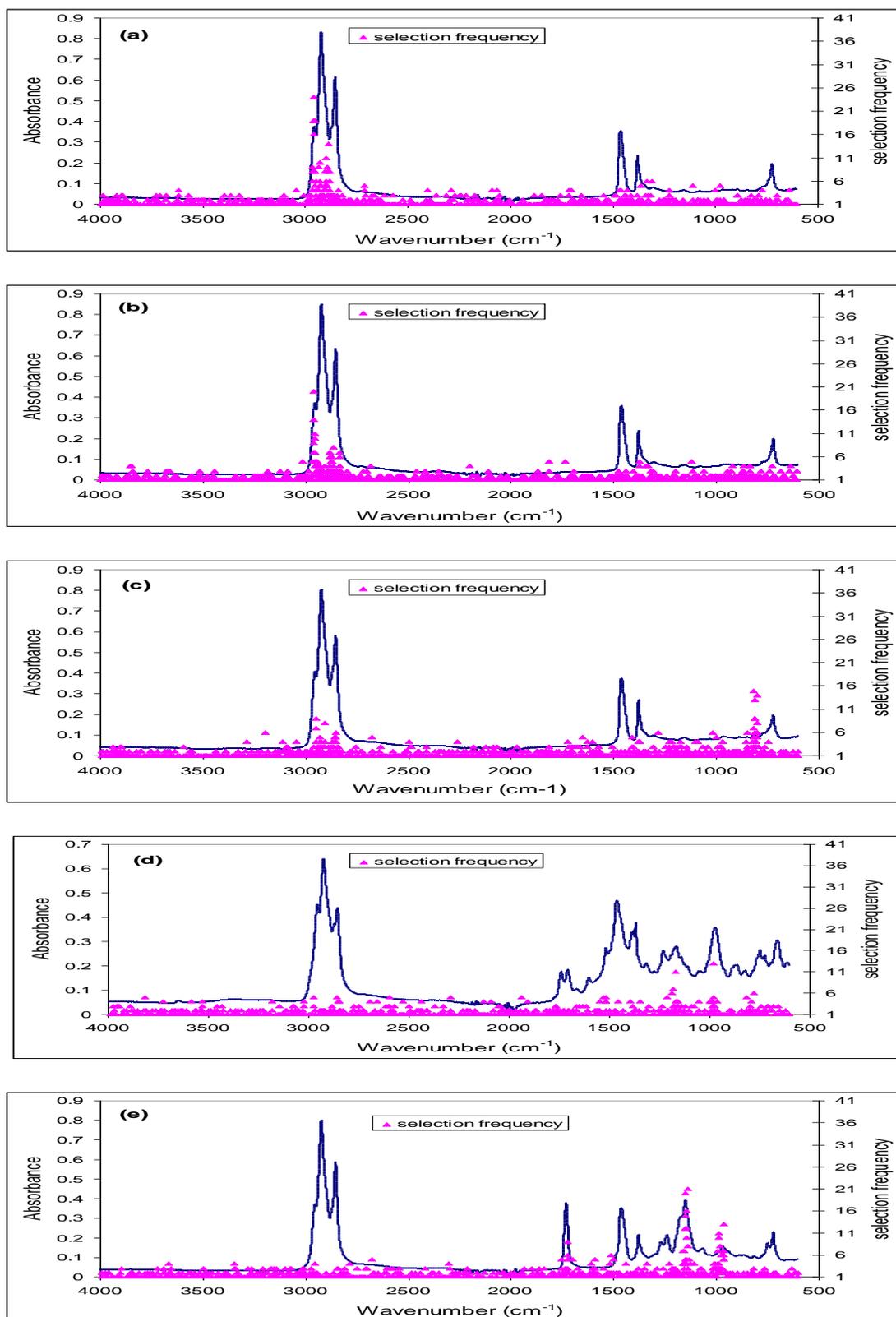


Figure 5.18. Frequency distribution of GILS selected FTIR-ATR wavelengths for (a) Base Oil 1, (b) Base Oil 2, (c) Additive 1, (d) Additive 2, (e) Additive 3, contents of 'Engine Oil 2' samples, when the resolution was 4 cm⁻¹

5.2.2. PLS Results

PLS was performed for the set of 'Engine Oil 2' using the data in the range of 4000-600 cm^{-1} , 1500-600 cm^{-1} and 1300-800 cm^{-1} . Figure 5.19 illustrates PLS results for the set of 'Engine Oil 2' in the range of 4000-600 cm^{-1} when resolution was 4 that was shown below. According to Figure 5.19, only one model for Additive 2 have high prediction ability of concentrations with high regression coefficients. However, for other components the models are suffering from high collinearity and cannot be trusted for producing a high calibration quality and for making prediction of concentrations in the independent validation samples.

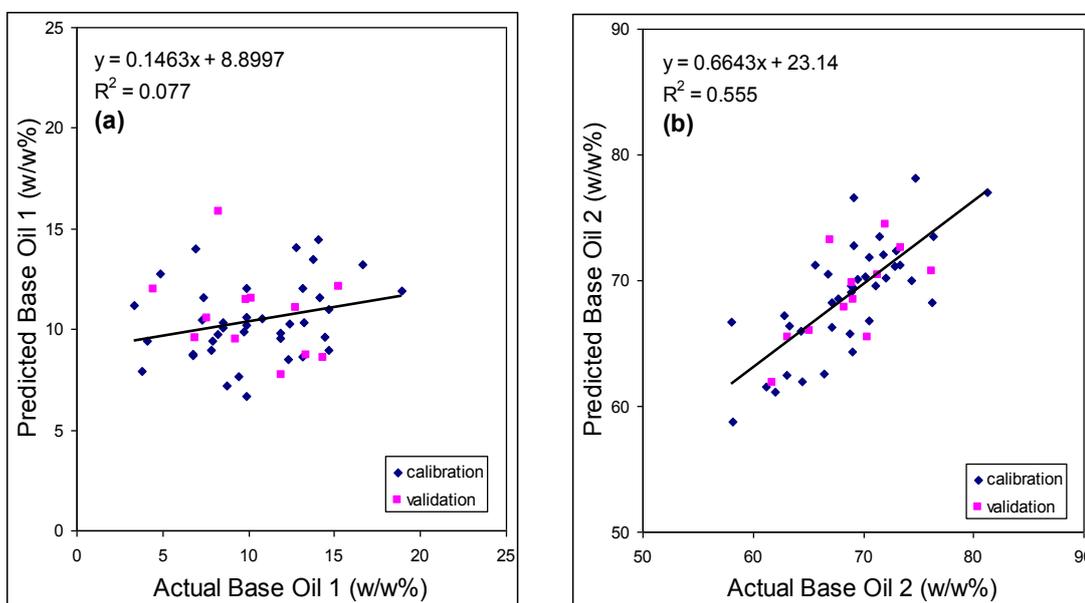


Figure 5.19. PLS Results for the set of 'Engine Oil 2' in the range of 4000-600 cm^{-1} (Res=4)

(cont. on next page)

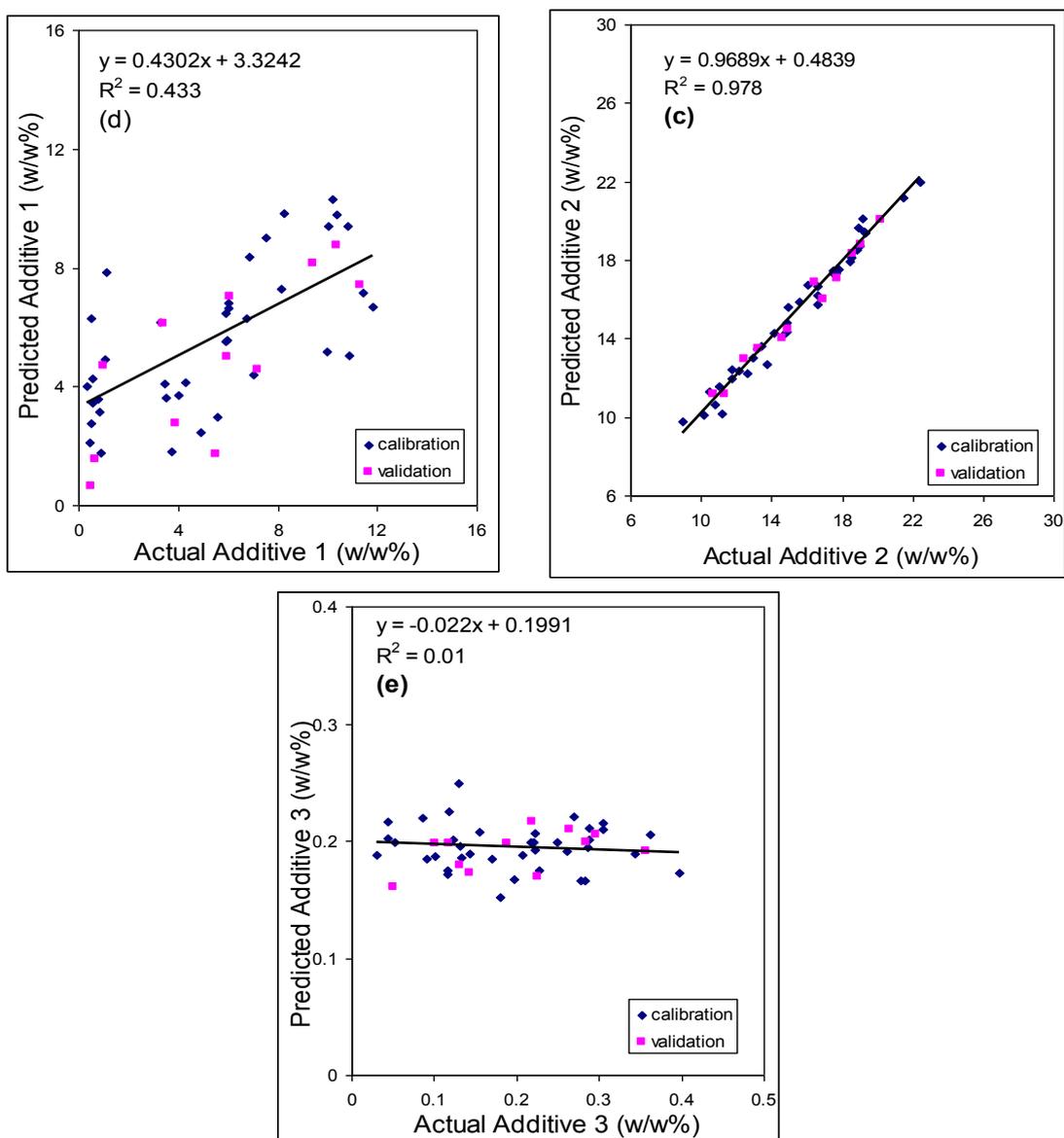


Figure 5.19. (cont.)

As it was stated in previous paragraphs for the set of 'Engine Oil 1', in the mid infrared region, the wavenumber range of $1500\text{-}600\text{ cm}^{-1}$ consists of many characteristic bands of lubricating oils especially belongs to the additives such as calcium or magnesium sulphonates, phenates and salicylates. In this spectral region, PLS was used to examine the prediction ability of the models. Figure 5.20 shows the models developed for the components in 'Engine Oil 2' using the subtracted data in mid infrared region.

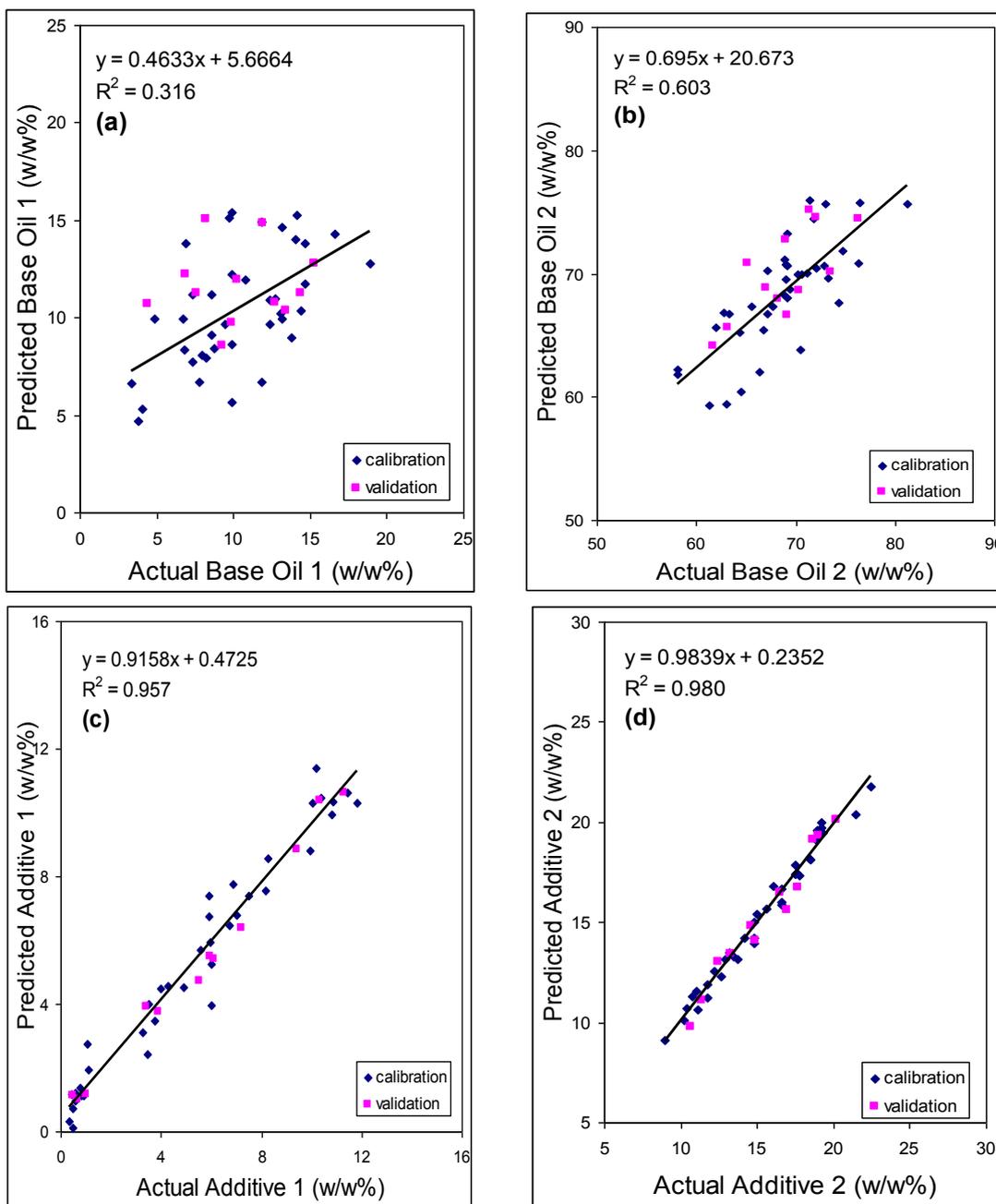


Figure 5.20. PLS Results for the set of 'Engine Oil 2' in the range of 1500-600 cm^{-1} (Res=4)

(cont. on next page)

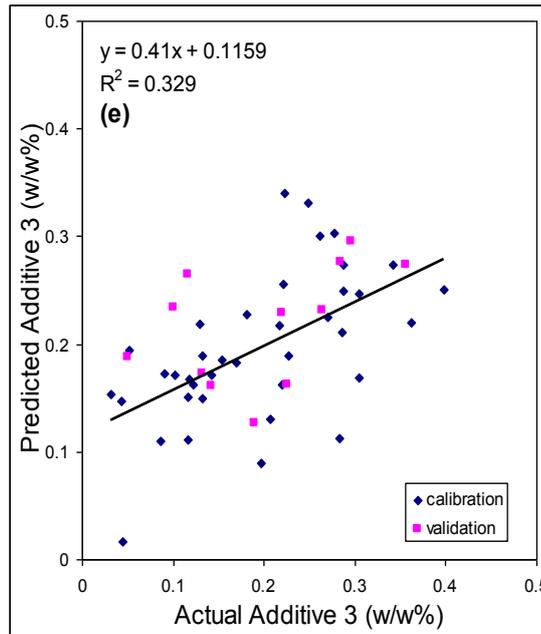


Figure5.20. (cont.)

Performing PLS to only $1500\text{-}600\text{ cm}^{-1}$ spectral region gave better results for all the components than the results for all the spectrum but still the models are suffering from collinearity except the model for Additive 2.

In the last part of the study, PLS was used to develop models for the components in 'Engine Oil 2' samples in a different wavenumber region, $1300\text{-}800\text{ cm}^{-1}$. Figure 5.21 illustrates the PLS results for each component in this spectral region that was shown below. Like the range of $1500\text{-}600\text{ cm}^{-1}$, the spectral region of $1300\text{-}800\text{ cm}^{-1}$ also consists of different characteristic bands belong to the additives so it was expected to develop more successful models. Figure 5.21 shows the models developed for the components in 'Engine Oil 2' using the wavenumber range of $1300\text{-}800\text{ cm}^{-1}$ in mid infrared region. For the set of 'Engine Oil 2' the best models were obtained using the data in this spectral range with highest R^2 .

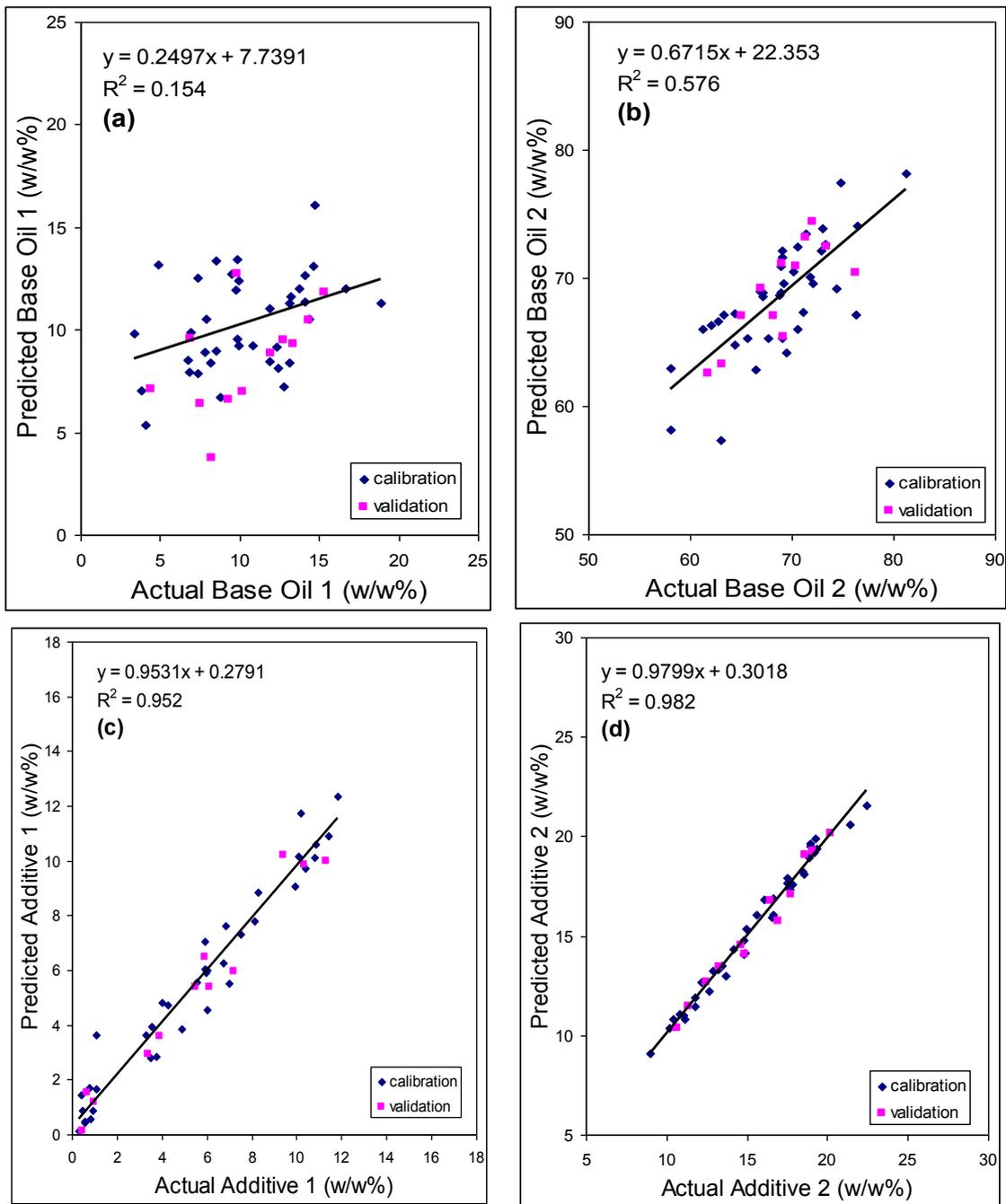


Figure 5.21. PLS Results for the set of 'Engine Oil 2' in the range of 1300-800 cm^{-1} (Res=4)

(cont. on next page)

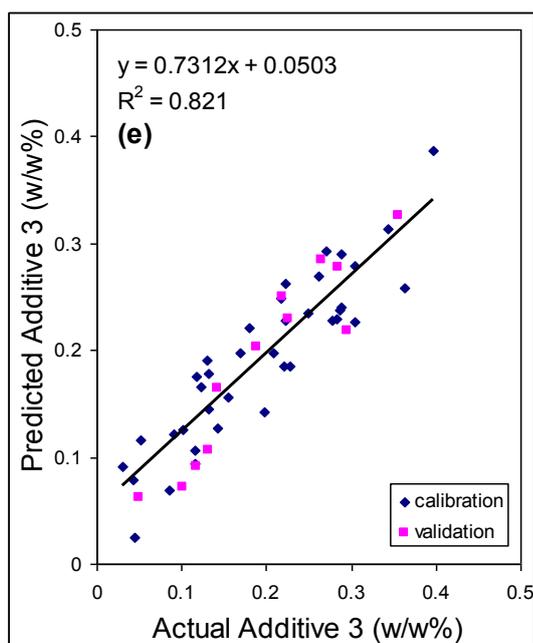


Figure 5.21. (cont.)

For making a comparison between the results of two statistical methods, PLS and GILS; SECV, SEP and R^2 values are given for the data set of ‘Engine Oil 2’ in Table 5.3 below.

Table 5.3. Calibration summary for the data set of ‘Engine Oil 2’ (Res=4)

Engine Oil 2		PLS			GILS		
		SECV g/g (w/w%)	SEP g/g (w/w%)	R^2	SECV g/g (w/w%)	SEP g/g (w/w%)	R^2
Name of the Component	Base Oil 1	4.595	3.834	0.077	0.623	0.768	0.972
	Base Oil 2	4.337	2.684	0.555	0.674	0.579	0.982
	Additive 1	3.595	2.093	0.433	0.403	0.378	0.989
	Additive 2	0.637	0.427	0.978	0.206	0.519	0.996
	Additive 3	0.126	0.073	0.010	0.023	0.022	0.943

Like for the set of ‘Engine Oil 1’, for the data set of ‘Engine Oil 2’, the results of PLS and GILS are very different from each other. Although PLS results are suffering from lower R^2 and higher SECV and SEP values for all the components, GILS provided more successful models with accurate results to predict the concentrations of the components in ‘Engine Oil 2’ samples. Also, PLS is more time-consuming method than GILS because of its computational requirements. Thus, when a comparison is made, GILS is capable of determining the engine oil compositions quantitatively.

CHAPTER 6

CONCLUSIONS

In this study, some calibration models were developed for two engine oils by the combination of FTIR-ATR spectroscopy and multivariate calibration methods. Samples were analyzed in mid-infrared region by the technique of FTIR-ATR spectroscopy at different resolution values, 4, 16 and 32 cm^{-1} . GILS was used as the multivariate calibration method when the calibration models were developed for each component for both engine oil sets. Reliability of the calibration models were determined by SECV and SEP values as well as with the R2 values from the reference vs. predicted content plots. Then the models for each components at different resolutions were compared by a statistical technique of ANOVA. According to ANOVA, there is no significant difference between the results obtained at different resolutions for each components. From the results, it is seen that successful calibration models can be constructed by using the methods mentioned to provide fast and nondestructive determination of base oils and additives in lubricating oils. This might provide an alternative method in lubricants industry instead of time-consuming and destructive ASTM methods.

Also, for making a chemometric comparison between the results, PLS was used. It was performed for all the components using the whole spectra, then only for the range of 1500-600 cm^{-1} and lastly for the range of 1300-800 cm^{-1} . According to this technique, the results showed that subtracted data gave more successful models. Unlike ILS, there is no variable selection necessary, and there is less of a tendency to overfit the data due to noise.

At the end of the chemometric study, the noticeable differences can be summarized between the results for PLS and GILS. PLS has a significant advantage that it models data in a more compact form and, because it organizes the data based on the similarity of information content, it can also aid in chemical interpretation of the system. However, multivariate calibration models that are generated with GILS use selection frequency plots indicating that with all the overlapping and complex nature of the FTIR spectra of the multicomponent mixtures because GILS algorithm only focuses on the regions where the most concentration related information is contained. This

provides to increase the prediction ability of the models for each components in complex mixtures. When we compare GILS models with PLS models developed in this study, it was clearly showed that GILS models had better prediction performance than PLS models for all the components studied. Also, quantitative determination of the components based on FTIR spectroscopy coupled with multivariate calibration offers a much faster analysis that could allow continuous monitoring of quality control process of the lubricants.

REFERENCES

1. Mortier, R. M.; Fox, M.F; Orszulik, S.F; *Chemistry and Technology of Lubricants*; Springer Dordrecht Heidelberg London New York; 3rd Edition; 2010, 3-10, 77-90.
2. United States Environmental Protection Agency (EPA) ; *Managing Used Oil Advice for Small Businesses*; Solid Waste and Emergency Response (5305W); EPA530-F-96-004; 1996.
3. Dong J.; Van De Voort F. R.; Yaylayan V.; Ismail A.A.; *Determination of total base number (TBN) in lubricating oils by mid-FTIR spectroscopy*; *Lubrication Engineering*; 57, 11; 2001; 24-30.
4. Felkel Y.; Dorr N.; Glatz F.; Varmuza K.; *Determination of the total acid number (TAN) of used gas engine oils by IR and chemometrics applying a combined strategy for variable selection*; *Chemometrics and Intelligent Laboratory Systems*; 101, 1; 2010; 14-22.
5. Al-Ghouti M. A.; Al-Degs Y. S.; Amer M.; *Application of chemometrics and FTIR for determination of viscosity index and base number of motor oils*; *Talanta*; 81, 3; 2010, 1096-1101.
6. Borin, A., Poppi, R., *Multivariate Quality Control of Lubricating Oils Using Fourier Transform Infrared Spectroscopy* , *J. Braz. Chem. Soc.*, 15, 4, 2004; 570-576.
7. Blanco, M.; Coello, J.; Iturriaga H.; Maspoch, S.; Gonzales, R.; *Determination of water in lubricating oils by mid- and near-infrared spectroscopy*; *Microchimica Acta*; 128, 3; 1998; 235-239.
8. Setle, F.A; *Handbook of Instrumental Techniques for Analytical Chemistry*; Prentice Hall PTR; 543, 7; 1st Edition; 1997; 253-260.
9. Derrick, M.R; Stulik, D; Landry, J.M; *Infrared Spectroscopy in Conservation Science*; CIP; 99-37860; 1999; 5-18.
10. Kenkel, J.; *Analytical Chemistry for Technicians*; CRC Press; 3rd Edition; 2003; 161.
11. Skoog, D. A.; Holler, F. J.; Crouch, S. R.; *Principles of Instrumental Analysis*; Thomson Brooks/Cole Publishing; 6th Edition; 2005, 469-471.
12. Smith, B. C; *Fundamentals of Fourier Transform Infrared Spectroscopy*, CRC press, 2nd Edition, 1996,.41-47.
13. Stuart, B; *Infrared Spectroscopy: Fundamentals and Applications*, Wiley Publishing, 1st Edition; 2004; 46-49.

14. Gayete, J.F; de la Guaerdia, M; Garrigues, S; *Attenuated total reflectance infrared determination of sodium nitrilotriacetate in alkaline liquid detergents*; 70(4):870-5; 2006.
15. Carolei, L.; Gutz, I.G.R; *Simultaneous determination of three surfactants an water in shampoo and liquid soap by ATR-FTIR*. Talanta; 66; 2005; 118-124.
16. Hayden, Y.V; Boqué, R; Cuesta, M.J.R; *Calibration*; Lcgc Europe; 20, 6, 2007; 349-357.
17. Özdemir, D; Karaman, İ; *Prediction of extractives and lignin contents of anatolian black pine (Pinus nigra Arnold. var pallasiana) and Turkish pine (Pnus brutia Ten.) trees using infrared spectroscopy and multivariate calibration*; 2008, 18-23.
18. Özdemir D.; Öztürk B., *Genetic Multivariate Calibration Methods for Near Infrared (NIR) Spectroscopic Determination of Complex Mixtures*. Turk. J. Chem. 28, 2004; 497-514.
19. Haaland, D.M; Thomas, E.W; *Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information*; Anal. Chem.; 60, 1988, 1193-1202.
20. Wang, Y., Veltkamp, D. J., and Kowalski, B. R, *Multivariate instrument standardization.*; Analytical Chemistry, 63, 1991, 2750-2756.
21. McClean, P; *Population and evolutionary genetics*; 1997.
22. Mitchell M.; *An Introduction to Genetic Algorithms*; The MIT Pres; 5th Edition; 1999; 2-5.
23. Özdemir, D.; Yalçın, A.; Ergün, D.; Uçar, O.; *Determination of aluminum rolling oil additives and contaminants using infrared spectroscopy coupled with genetic algorithm based multivariate calibration*; Vibrational Spectroscopy; 54, 1; 2010, 10-20.