SPECTROSCOPIC DETERMINATION OF VEGETABLE OIL AND BIODIESEL IN PETROLEUM DIESEL USING MULTIVARIATE CALIBRATION

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

MASTER OF SCIENCE

in Chemistry

by
Aysun ARIKAN

July 2008
İZMİR

We approve the thesis of **Aysun ARIKAN**

Doç. Dr. Durmuş ÖZDEMİR
Supervisor

Prof. Dr. Serdar ÖZÇELİK
Committee Member

Yrd. Doç. Erol ŞEKER
Committee Member

Doç. Dr. Şerife YALÇIN
Committee Member

Doç. Dr. Figen KOREL
Committee Member

14 July 2008

Prof. Dr. Levent ARTOK
Head of the Chemistry Department

Prof. Dr. Hasan BÖKE
Dean of the Graduate School of
Engineering and Science

ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to my supervisor, Assoc. Prof. Durmuş Özdemir for his support, advise and patience during my thesis. Without his generous donation of his time and encouragement this thesis would not have been completed.

I wish to thank my co-workers Betül Öztürk and İbrahim Karaman for their support and helps during my thesis project.

I am pleased to acknowledge TUBITAK for the financial support.

I want to express my thankfulness to the Department of Chemistry to accepting me to the master programme and give me the opportunity.

I am also thankful to my friends in IYTE for their company and for sharing hard and happy moments. Most importantly, I would like to thank Mehmet Emin Uslu for his suggestions, and continues support during my thesis project. I would not be able to finish this thesis without his helps and encouragement.

Lastly numerous thanks to my family for their support and encouragement all these years, especially to my brother Ayberk for his helps and contributions.

ABSTRACT

SPECTROSCOPIC DETERMINATION OF VEGETABLE OIL AND BIODIESEL IN PETROLEUM DIESEL USING MULTIVARIATE CALIBRATION

Due to the limited petroleum reserves and pollutant effect of petroleum fuels, the use of alternative fuels has became important in recent years. Diesel is one of the most used petroleum fuel, whose exhaust emissions composed of harmful particles, that pollutes the environment. In this sense, vegetable oils and their esters (biodiesel) are considered environmentally friendly fuels, which reduce hazardous impact of diesel emissions. However, using vegetable oils directly in diesel engines may cause some engine problems due to their high viscosity. The most commonly used way to reduce their viscosity is the converting into biodiesel. Because biodiesel production is expensive and time consuming, diesel may be illegally adulterated with vegetable oils before converting into biodiesel.

Diesel may also adulterated with kerosene due to the large price differences. The main impact of this adulteration is increased emissions, which damage the environment. On the other hand, the addition of kerosene may also damage the engine. Because of these reasons, it is important to determine these adulterants illegally present in petroleum diesel. In this study, we have determined the adulteration of diesel with sunflower, canola oil, used frying oil, kerosene, and biodiesel by different molecular spectroscopic techniques combined to genetic inverse least squares (GILS). The results showed that the GILS method is suitable in the fast determination of diesel adulteration with vegetable oils, used frying oil, kerosene, and biodiesel when combined to NIR, FTIR-ATR, and molecular spectroscopic techniques.

ÖZET

PETROL TÜREVİ DİZEL İÇERİSİNDE BİTKİSEL YAĞ VE BİYODİZELİN ÇOK DEĞİŞKENLİ KALİBRASYON KULLANILARAK SPEKTROSKOPİK TAYİNİ

Petrol rezervlerinin sınırlı sayıda olması ve petrol yakıtlarının kirletici etkisinden dolayı son yıllarda alternatif yakıtların kullanımı büyük önem kazanmıştır. Dizel yakıtı en çok kullanılan petrol yakıtlarından biridir. Ancak yaydığı egzoz gazları insan sağlığı ve çevre açısından son derece zararlıdır. Bu nedenle bitkisel yağlar ve biyodizel olarak tanımlanan alkil esterleri çevre ile dost alternatif yakıtlar olarak kullanılmaktadırlar. Yüksek viskozitelerinden dolayı bitkisel yağların dizel araçlarında direk olarak kullanımı motorda bazı problemlere neden olabilmektedir. Bitkisel yağların belli miktarlarda dizel ile karıştırılması bu sorunu çözse de en etkili yöntem biyodizele dönüştürülmeleridir. Ancak biyodizel üretimi belli bir maliyet gerektirdiğinden ve zaman aldığından bitkisel yağlar biyodizele çevrilmeden direk olarak dizel içerisine katılabilmektedirler.

Dizel yakıtı aynı zamanda yüksek fiyat farkı olan kerosen ile de karıştırılabilmektedir. Kerosen de bitkisel yağlar karıştırıldığında olduğu gibi motorda problemlere neden olmaktadir. Bu nedenle dizel içerisinde ihtiva eden yasal olmadan karıştırılmış maddelerin belirlenmesi önem taşımaktadır. Bu çalışmada çeşitli moleküler spektroskopik yöntemler ile birlikte genetik çok değişkenli kalibrasyon metodu kullanılarak dizel içerisinde bulunan bitkisel yağ, kullanılmış kızartma yağı, biyodizel ve kerosen gibi maddelerin tayini yapılmıştır. Deney sonuçları çok değişkenli kalibrasyon metodunun spektroskopik yöntemlerle kullanıldığında dizel içerisine karıştırılmış olan maddeleri hızlı ve kolay bir şekilde tayin ettiğini göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

Introduction

The fact that one day petroleum will be finished due to the limited reserves has directed the scientist to search for renewable alternative fuels instead of petroleum fuels. Another important reason for the researches on alternative fuels is the pollutant and hazardous effects of petroleum fuels that damage public's health and cause global warming, which is an important problem of today's world.

Vegetable oils and their monoalkyl esters (biodiesel) are considered one of the renewable alternative fuels of the future due to their environmentally benefits. However, they haven't entirely displaced petroleum as an engine fuel in terms of technical, economical and political considerations.

Diesel is one of the most prevalent petroleum based fuel used in many areas especially in transportation. Therefore, this fuel plays an important role in the economy of a country.

1.1. Diesel Fuel

Petroleum diesel or simply referred as diesel is a complex mixture of aliphatic, aromatic and olefinic hydrocarbons. It also contains in a small amounts of sulfur, nitrogen, oxygen, metals, etc. It is the main energy source for compression ignition (diesel) engines found in trucks, ships, locomotives, and passenger cars. Diesel fuels are composed of molecules with 8–40 carbon atoms (Vieira, et al. 2006).

The exhaust of diesel fuel composed of harmful particles which pollute the environment and damage the human health due to its cancerogenic effect. The cancerogenic particulate of diesel exhaust has been proved from many studies in humans and in animals. The main hazardous pollutants emitted from the exhaust of a typical diesel engine include carbon monoxide (CO), carbon dioxide (CO_2), sulfur oxides (SO_x), nitrogen oxides (NO_x), polycyclic aromatic hydrocarbons (PAHs) and particulate matter (PM) (Altun, et al. 2007). In this sense, it became very important to find environmentally friendly alternative fuels that reduce the hazardous impact of diesel fuel emissions.

Vegetable oils and their monoalkyl esters (biodiesel) have been pointed out as environmentally friendly alternative fuels for diesel engines since recent studies show that they exhibit decrease in CO_2, CO, SO_x, aromatic hydrocarbons, and PM emissions when compared with diesel fuel (Rakopoulos, et al. 2006). These alternative fuels can be extensively used with or without modification in the compression ignition (IC) engines as a diesel fuel (Agarwal 2007). Whether they haven't entirely displaced diesel fuel it is evident that they will be the main widely used fuels for diesel engines in the near future.

1.2. Vegetable Oils as Fuel

Vegetable oils are the mixtures of triglycerides (TGs), which are the esters of a trialcohol (glycerol) with three different fatty acids. One of the biological function of triglycerides is its use as fuel (Stavarache, et al. 2006). Hence, in various countries many studies have been performed on different vegetable oils including soybean oil, sunflower oil, cotton seed oil, corn oil, canola oil, in addition to waste (used or frying) vegetable oil (WVO) for their feasibility as diesel fuel. In fact, the works carried out on the use of vegetable oils depend on 1900s. Primarily Rudolph Diesel, the inventor of the diesel engine used the peanut as a fuel in diesel engines in 1900. Then the works on vegetable oils as an engine fuel was continued in 1930s and 1940s and after the fuel crises early 1980's gain more attention. The results of these studies have shown that vegetable oils in crude (raw) form can be used as a diesel engine fuel with small amounts of power loss when compared to diesel fuel (Knothe 2001). However, using vegetable oils directly in diesel engines may cause some engine problems due to their high viscosity, which is 10–20 times higher than petroleum diesel (Stavarache, et al. 2006).

There are many ways to reduce the viscosity of vegetable oils in order to use them directly in diesel engines. One way is modifying the engine, so it preheats the oil to produce an acceptable viscosity, another way is blending vegetable oils in a small ratios with diesel fuel, and the most common way is converting vegetable oils into biodiesel which doesn't require any modification in diesel engines and has a high energetic yield (Rakopoulos, et al. 2006, Wang, et al. 2006, Pugazhvadivu, et al. 2005, Huzayyin, et al. 2004).

1.3. Biodiesel as Fuel

Among the alternative fuels biodiesel has gained popularity in recent years as a substitute for diesel fuel. Almost all researches on biodiesel show that they can be performed well in diesel engines, which is reported better than petroleum diesel in several aspects. It contains no more petroleum, "bio" describes its renewability and "diesel" describes its use as diesel fuel. It is a mixture of fatty acid alkyl esters derived from renewable vegetable oils, animal fats, and used frying oils. Biodiesel is chemically simple since biodiesel mixture doesn't contain more than six or seven fatty acid esters (Stavarache, et al. 2006).

Biodiesel is made through a chemical process called transesterification shown in the Figure 1.1. The transesterification of vegetable oils comprises the reaction of triglycerides with an alcohol in the presence of a strong acid or base catalyst producing a mixture of fatty acid alkyl esters (biodiesel) and glycerol (Demirbaş 2007, Wang, et al. 2006).

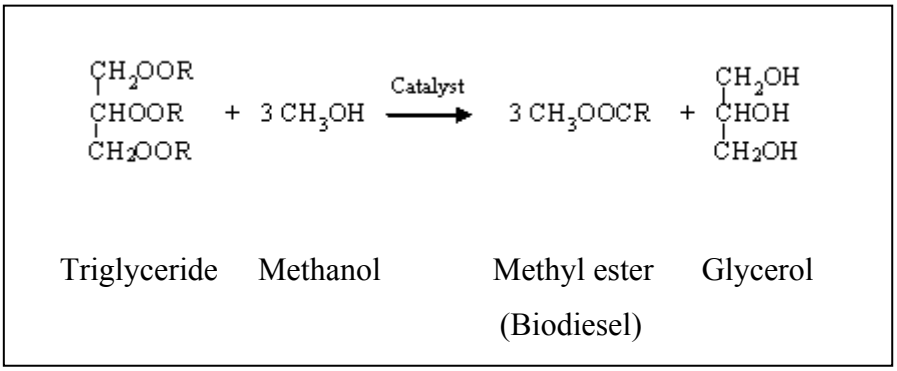

Figure 1.1. Principle of the transesterification reaction

NaOH and KOH are the most studied catalysts for the production of biodiesel. The overall process consists roughly of three consecutive, reversible reactions in which monoglycerides and diglycerides are formed as intermediates. However, the real situation is much more complicated. Base catalyzed transesterification reactions are faster than acid-catalyzed processes (Kwanchareon, et al. 2007).

There are several choices for vegetable oil sources in the production of biodiesel. Transesterification reactions have been studied for many vegetable oils such as soybean,

rapeseed, sunflower, palm, corn, cotton, castor, and canola oil. In Turkey, canola oil is the most preferred vegetable oil for the production of biodiesel.

Using biodiesel has many advantages over regular diesel fuel. The most important advantages of biodiesel are its renewability, and biodegradability which means that the product can be recycled by nature, or can be broken down into its smallest parts through the actions of microorganisms so that it will not pollute the environment. Moreover, the reduction of exhaust emissions and greenhouse gas effect due to its minimal sulfur, carbon dioxide and aromatic content make biodiesel an environmentally friendly fuel (Knothe, et al. 2003, Hernando, et al. 2007).

Biodiesel can be used in diesel engines by blending with petroleum diesel in any proportion without engine modifications. Therefore biodiesel is widely used by blending with petroleum diesel. The proportion of the blend is defined as BX, where X is the amount of biodiesel in the blend. For example, diesel engines can run using blends of diesel up to 20% which is called B20. In addition, pure biodiesel fuel contains 100% esters of fatty acids called as B100. B5 used in Europe contains 5% biodiesel in diesel (Pinto, et al. 2005). In Turkey, the use of biodiesel allowed up to a volume of 2% in diesel (4.12.2003, 5015 sayılı Petrol Piyasası Kanunu, Madde 16). This blend is referred as B2. There are many studies about the blends of various vegetable oils with diesel fuel. From these studies the blend of 20% oil and 80% diesel was found successful in diesel engines (Agarwal 2007).

1.4. The Adulteration of Diesel Fuel

The term "adulteration" means the illegal addition of any chemical foreign substance, which is called as adulterant, into another substance, generally to reduce manufacturing costs. The adulterants also may be accidentally or unknowingly introduced into substances. Because biodiesel production process is expensive and time consuming diesel fuel may be illegally adulterated with raw or used frying vegetable oils before converting into biodiesel. Diesel fuel may also involve much cheaper kerosene.

Kerosene is a petroleum product which is obtained similar to diesel. It is widely used as a heating, lighting and rocket fuel. The fact that it is miscible with diesel, some amount of mixing with almost no change in the properties of automotive fuel is possible. In addition, large price differences between kerosene and diesel, make this

unhealthy and unethical adulteration available (Roy 1999, Divya and Mishra 2007). Adulteration of diesel with low-taxed or subsidized kerosene is a common problem in many countries. The main impact of this adulteration is increased emissions. In the case of the adulteration of diesel with kerosene, the exhaust have harmful, cancer causing hydrocarbons, nitrous oxides, and carbon monoxide so that it damages the environment and human health (Taksande and Hariharan 2006). The adulteration of diesel by kerosene not only causes serious environmental problems also damaging automotive engines. It reduces diesel's lubricating function that causes faster wear of the pistons so that the performance and life of the engine decreases. In this sense, it is important to determine or verify the illegal adulteration of petroleum diesel. For this purpose a number of reliable and fast analytical methods have been developed for checking adulteration of petroleum diesel where they are practiced in the illegal market. Moving beyond that the aim of this work is to determine the amount of illegal adulteration contents present in petroleum diesel by developing calibration models based on molecular spectroscopic techniques combined to multivariate calibration.

CHAPTER 2

MOLECULAR SPECTROSCOPY

Molecular spectroscopy is important for analytical chemistry in the detailed investigation of molecules. One of the most spectacular applications of molecular spectroscopy is identification the structure of molecules based upon its electromagnetic characteristics. The structural information from this kind of spectroscopy is very detailed and precise. The investigation of chemical reactions plays also important role in molecular spectroscopy. The possible variety in molecular spectroscopy is large due to the numerous different kinds of molecules compared to the approximately 100 different types of atoms. The molecules can be determined even in gas, liquid or solid phases. The spectra in molecular spectroscopy is complex because of the greate number of excited states. The determination of mixtures of molecules is difficult due to the width of molecular electronic bands. The spectral bands in molecular spectroscopy result from the absorption, emission, reflection, and scattering of electromagnetic radiation when the energy of a molecule changes. In molecular spectroscopy, bonding of atoms lead to rotational and vibrational transitions in addition to electronic transitions. These added transitions result from very broad peaks in uv-visible region to the microwave region. Usually, transitions within the rotational energy levels are observed in far infrared and microwave regions, transitions within the vibrational energy levels are observed in infrared region, and transitions between the electronic energy levels are observed in ultraviolet, visible and near IR regions of the spectrum. In general, the vibrational transitions result in changes in the rotational mode and the electronic transitions result in changes in the rotational and vibrational modes as well (Banwell and McCash 1994, Hollas 2004, Ingle and Crouch 1988).

There are many specific molecular spectroscopic techniques such as infrared spectroscopy and molecular fluorescence spectroscopy that are used in this study.

2.1. Infrared (IR) Spectroscopy

Infrared spectroscopy is a technique based on the vibrations of the bonds of a molecule. It is the measurement of transmittance or absorption intensity of a sample in the IR region at different IR frequencies. An infrared spectrum is commonly obtained by passing infrared radiation through a sample and determined what fraction of the incident radiation is absorbed at a particular energy. IR spectroscopy is one of the most commonly used spectroscopic technique to determine the chemical functional groups in the sample since different functional groups absorb characteristic frequencies of IR radiation. In addition, infrared spectroscopy can be used for the identification the amount of a particular compound presents in a mixture. It is an important and popular tool in terms of the wide range of sample types as gase, liquid, and solid phases. One of the great advantages of infrared spectroscopy is that almost any sample can be studied in any state. Liquids, solutions, pastes, powders, films, fibres, gases and surfaces can all be examined by IR spectroscopy (McKelvy, et al. 1996, Ingle and Crouch 1988).

Infrared energy is the electromagnetic energy of molecular vibrations consist of stretching and bending as illustrated in Figure 2.1. While streching vibrations occur at higher energy levels, bending vibrations occur at higher energy levels.

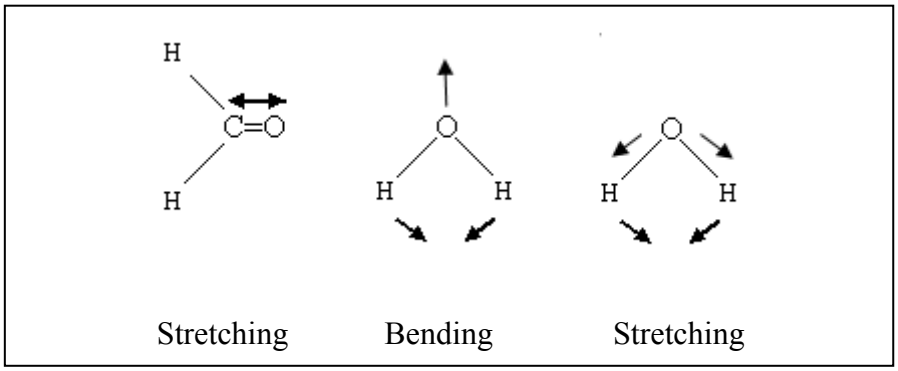

Figure 2.1. Streching and bending vibrations

The absorption of the IR radiation is observed, when the frequency of a specific vibration is equal to the frequency of the IR radiation. In addition, minimum one vibrational motion must change the net dipole moment of the molecule so that the absorption of the infrared radiation occurs. A heteronuclear diatomic molecule can be given as an example of an infrared-active molecule in which the dipole moment of the

molecule changes. On the other hand, infrared-inactive molecules are the homonuclear diatomic molecules such as H_2, Br_2, and I_2 as their dipole moments are zero (Skoog and Leary 1992, Hsu 1997).

Infrared radiation, commonly defined as electromagnetic radiation have wavenumbers from approximately 13,000 to 10 cm^{-1}, or wavelengths from 0.78 to 1,000 μm. It is delimited from the red end of the visible region at high frequencies to the microwave region at low frequencies. Because the IR region covers a wide range of the electromagnetic spectrum, it is divided into three areas in order to provide the requirements of instruments for their different applications (Skoog and Leary 1992). The areas of IR region defined as near infrared (NIR), middle infrared (MIR), and far infrared (FIR). Table 2.1. shows the wavelength and wavenumber ranges of these IR regions.

Table 2.1. The corresponding wavenumber and wavelength ranges of the IR region

	Near IR	Mid IR	Far IR
Wavenumber	12,800-4000 cm^{-1}	4000-200 cm^{-1}	200-10 cm^{-1}
Wavelength	780-2500 nm	2500-50,000 nm	50,000-1x10^6 nm

The near infrared (NIR) spectroscopy is used to detect solid and liquid samples with no prior manipulation. Moreover, this technique is useful for the quantitative determination of complex mixtures by the help of computers and chemometrics. Recently, near IR spectroscopy has gained increased interest, especially in process control applications (Xixiong, et al. 2007). The middle infrared (MIR) spectroscopy is mostly used to identify the structure of organic molecules as each functional group in the molecule has sharp absorption bands in this region. It is also used for both qualitative and quantitative analysis in analytical chemistry. The mid IR region covers the fingerprint region (1300 to 910 cm^{-1}) where the absorptions include the contributions from complex interacting vibrations that give each compound a unique information. A good compliance between the IR spectra of two compounds in all frequency ranges, particularly in the fingerprint region means that they have the same molecular structures. The far infrared (FIR) spectroscopy is used for qualitative analysis

of organic, inorganic, and organometallic compounds which involve heavy atoms (mass number over 19). They can be analyzed due to their metallic bands in this region. The far IR spectroscopy requires the use of specialized optical materials and sources. It provides useful information to structural studies such as conformation and lattice dynamics of samples (Skoog and Leary 1992, Hsu 1997).

2.2. Near Infrared (NIR) Spectroscopy

Generally, in the scientific literature, it is defined that the NIR region of the electromagnetic spectrum bounded from 780 to 2500 nanometers (12,800-4000 cm^{-1}). In this region the absorption bands observed due to overtones of hydrogen-stretching vibrations or due to combinations of stretching and bending modes of fundamental vibrations found in the mid infrared region. The most observed fundamental vibrations are the harmonic C-H stretching vibrations and their corresponding combinations including methyl C-H stretching, methylene C-H stretching and aromatic C-H stretching vibrations, also O-H stretching and N-H stretching vibrations which occur from approximately 690 to 3000 nm. Therefore especially C-H, O-H, and N-H functional groups can quantitatively measured by this technique. The basic uses of near infrared spectroscopy are, process control, quality assessment, identification of raw materials and process byproducts, and quantitative analysis of complex mixtures (Stuart 2004, Skoog and Leary 1992).

2.2.1. Theory and Principles of NIR Spectroscopy

The near infrared spectral region is energetic enough to excite the molecule to the vibrational or rotational states. The absorption bands are obtained due to the overtones (780-1800 nm) and combinations of fundamental vibrations (1800-2500 nm) at mid IR region. Fundamental vibrations in infrared spectroscopy can be explained by diatomic harmonic oscillator related to Hooke's law. The Hooke's law describes the frequency (υ) of a vibration for a simple two body diatomic harmonic oscillator (Barton 2002, Burns and Ciurczak 2001), which can be calculated from the equation below:

$$\upsilon = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{2.1}$$

where k is the force constant, which is related to the bond strength and μ is the reduced mass stated as:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \tag{2.2}$$

where m_1 and m_2 are the masses of atoms of the diatomic molecule involved in the vibration. If the diatomic molecule demonstrates ideal harmonic oscillator, its potential energy is calculated as:

$$V = \frac{1}{2}kx^2 \tag{2.3}$$

where x is the displacement of the internuclear distance. Because the quantized nature of the molecular vibration energies are not described by this equation, the equation is treated and developed to give the potential energy (E) as:

$$E = \left(v + \frac{1}{2}\right)\frac{h}{2\pi}\sqrt{\frac{k}{\mu}} \tag{2.4}$$

where h is the Planck's constantand and υ is the vibrational quantum number, which only takes positive integer values including zero. The potential energy equation can be rearranged using the equation 2.1 as:

$$E = \left(v + \frac{1}{2}\right)h\nu \tag{2.5}$$

In this way only distinct energy levels are allowed. The separation between two vibrational levels would be:

$$\Delta E = h\upsilon \tag{2.6}$$

The ideal harmonic oscillator can be shown in Figure 2.4 by a plot demonstration. The dotted line labeled with (a) demonstrates the "ideal" case that the spring stretches and reaches a point where it loses its shape and decreases to zero. In molecules, the charges on the nuclei limit the approach of the nuclei during the compression step, creating an energy barrier. At the extension of the stretch the bond eventually breaks when the vibrational energy level reaches the dissociation energy. This effect is illustrated with asymmetric potential curve in the Figure 2.2 (Burns and Ciurczak 2001).

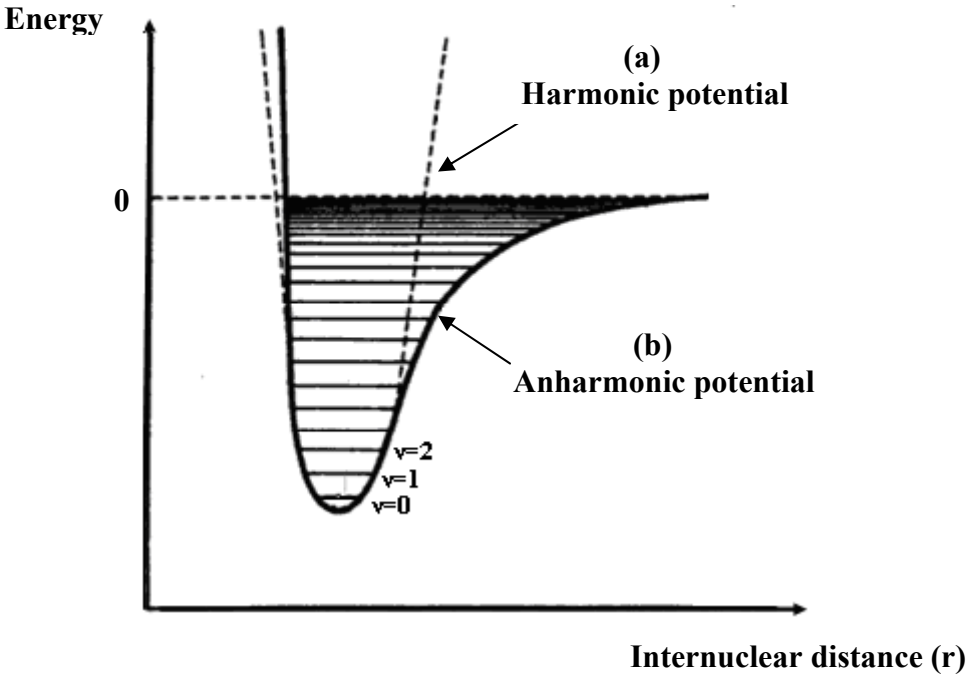

Figure 2.2. Energy diagram of vibrational modes

In fact, molecules are not ideal oscillators and transitions between more than one vibrational state are forbidden for such an ideal harmonic oscillator by quantum mechanical selection rules which allow anharmonic oscillator. This anharmonic behaviour is illustrated in Figure 2.4 at higher vibrational states, where the departures from harmonic behaviour occur since at high level of potential energy a molecule tends to dissociate and can no longer return to its equilibrium position. For an anharmonic oscillator, only the transitions with $\Delta\upsilon>1$, between vibrational states of $\Delta\upsilon=2$ or $\Delta\upsilon=3$ are possible. Near infrared overtone bands arise from these multi level transitions that occur at multiples of the fundamental vibrational frequency. A transition from $\upsilon=0$ to 2

is called first overtone and a transition from υ=0 to 3 is called second overtone. The wavenumbers of overtones can be estimated from their fundamental vibrations with an anharmonicity constant χ of 0.01–0.05 by the following equation:

$$\upsilon_x = \Delta\upsilon\ \upsilon_0 (1 - \Delta\upsilon_\chi) \qquad (2.7)$$

where υ_x is the wavenumber of x^{th} overtone, υ_0 is the wavenumber of fundamental vibration and χ is the anharmonicity constant (Burns and Ciurczak 2001).

Combination bands compose around 1900 nm to 2500 nm, occur when a single photon simultaneously excites two or more different molecular vibrations. The sum of energies of these different molecular vibrations is nearly equal to the transition energies of combination bands. Because near infrared absorption bands are wide, overlapping and 10-100 times weaker than their corresponding fundamental middle infrared absorption bands, the sensitivity of near infrared spectrometer is limited and near infrared spectra are very complex so that spectral interpretation is very difficult. Thus, the use multivariate calibration methods to process the recorded signals and extract the relevant information for qualitative or quantitative analysis are required (Abajo, et al. 2006, Reich 2005).

2.3.2. Instrumentation of NIR Spectroscopy

A NIR spectrometer is generally composed of a light source, a monochromator, a sample holder or a sample presentation interface, and a detector, allowing for transmittance or reflectance measurements. The light source is usually a tungsten halogen lamp, which is small and rough. Detectors used in NIR spectrometer include lead sulfide (PbS), silicon, and indium gallium arsenide (InGaAs). Sample holders can be glass or quartz and typical solvents are CCl_4 and CS_2. To measure good NIR spectra, the proper sample presentation is important so that several types of sample cells, such as quartz cuvettes with defined optical path length for liquids, specifically designed sample cells with quartz windows for semi-solids and powders, and adjusted sample holders for tablets and capsules have been developed. According to their properties near infrared instruments are divided into two classes as dispersive and fourier transform instruments. Dispersive instruments such as UV-visible-NIR spectrometers depend on the

monochromator types. The appropriate NIR measuring mode will be dictated by the optical properties of the samples (Skoog and Leary 1992, Ingle 1988).

2.3.3. Advantages of NIR Spectroscopy

NIR spectroscopy has many advantages over other vibrational spectroscopic techniques and classical methods. It can be used in many different fields such as agriculture, pharmaceutical, biotechnology and food industry because of its advantages. It offers on-line process control, non-invasive analysis. It is a non-destructive analytical technique with high-speed quantitative analysis without consumption or manipulation of the sample. In addition, NIR spectroscopy provides a low cost instrumentation with high signal-to-noise (Özdemir and Öztürk 2007, Font, et al. 2006).

2.4. Fourier Transform Infrared (FTIR) Spectroscopy

Infrared spectrometers have been commercially used since the 1940s, at the time that instruments depend upon prisms and act as dispersive instruments. Around 1950s, diffraction gratings had been introduced into dispersive machines. As a result of the improvement in instrumentation, a variety of IR spectroscopic techniques have been developed in order to remove the intractable drawbacks of classic dispersive IR spectroscopy. Hence, the most significant advances in infrared spectroscopy have come with the introduction of Fourier transform infrared (FTIR) spectroscopy. When compared with classic dispersive IR spectroscopy, FT-IR spectroscopy no longer measures one wavelength after the other, but applies an interferometric modulation of radiation. This type of instrument employs an interferometer and exploits the well established mathematical process of Fourier-transformation). Fourier-transform infrared (FTIR) spectroscopy has improved the quality of infrared spectra and minimized the time required to obtain data. FTIR spectrometers have been much more common than the traditional dispersive instruments. The working mechanism of a Michelson interferometer in an FTIR spectrometer is shown schematically below in Figure 2.3 (Ingle 1988, Smith 1996).

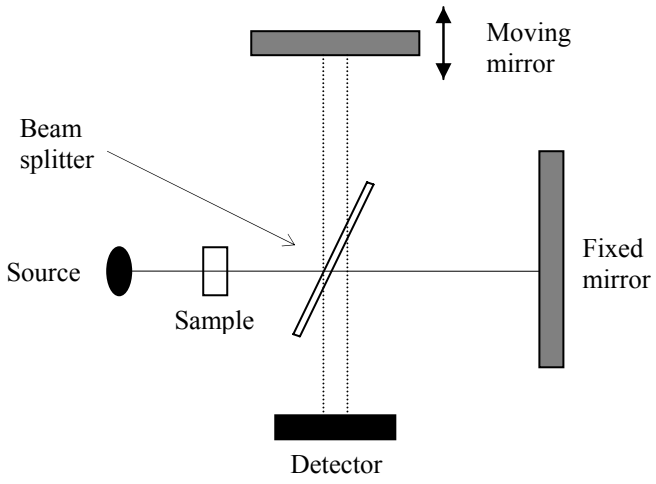

Figure 2.3. A schematic diagram of a FTIR spectrometer

The Michelson interferometer is a device designed by Michelson similar to an interferometer that is used for modulating optical radiation. The Michelson interferometer splits a beam of radiation into two beams in nearly same power and then recombines them in such a way that intensity is kept as it is. As shown in the Figure 2.5, IR radiation from the source is split into two beams by a beam splitter, which is a plate whose surfaces are partly reflecting. The beam from the source which passes straight through (the solid line) reflects off a fixed mirror and then this beam is diverted towards the detector by the beam splitter. The beam which is reflected by the splitter (the dashed line) reflects off the moving mirror, passes through the splitter and then on to the detector. The key thing about this arrangement is that, by the time they arrive at the detector, the two beams have different traversed different path lengths. Furthermore, by moving one of the mirrors, we can alter the path length of one of the beams; in the diagram above, it is the dotted beam whose path length can be changed. In the FT-IR spectrometers, the interference patterns of the modulated signals from interferograms are amplified, digitised, electronically stored and finally transformed into a spectrum by the fast fourier transform (FFT) algorithm. Therefore, the Fourier transformation can be considered simply as a mathematical means of extracting the individual frequencies from the interferogram for final representation in an IR spectrum (Skoog and Leary 1992, Smith 1996).

2.4.1. Advantages of FTIR Spectroscopy

The FTIR spectrometers have much greater sensitivity than conventional dispersive instruments for three major reasons. The first is the absorption at all wavelengths is recorded at the same time, rather than scanning through the spectrum so that more efficient use is made of the experiment time. This is called the multiplex or Fellgett advantage. The second is the Throughput or Jacquinot advantage. Since there are no narrow slits in FTIR spectrometers, the throughput of light is much greater than in dispersive instruments. The third is the extremely high wavelength accuracy and precision of the FTIR spectrometers which makes signal averaging possible. The output of the detector can be passed through in integrator so that improves the signal-to-noise ratio. These advantages have made FTIR spectrometers substitute to conventional dispersive instruments, especially for routine use. Furthermore, the mechanical simplicity of the interferometer is also an attractive advantage of FTIR spectrometers (Smith 1996, Skoog and Leary 1992).

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

The attenuated total reflectance (ATR) technique is used to obtain the spectra of solids, liquids, semi solids, and thin films. Different experimental setups of ATR-FTIR spectroscopy have been designed, including fiber optics for the study of various samples. It is a fast technique which yields a strong signal with only a few micrograms of sample and recent ATR devices allow the recording of nanogram quantities (Gayete, et al. 2006, Smith 1996).

2.5.1. Principles of ATR -FTIR Spectroscopy

The ATR- FTIR spectroscopy is executed using an accessory which is mounted on the sample compartment of an FTIR. At the heart of the accessory there is a crystal of infrared transparent material of high refractive index and on the accessory there are mirrors which bring the IR radiation to a focus on the face of the crystal. The crystal behaves as a waveguide for the IR radiation that will follow the shape of the crystal. If

the crystal has the proper refractive index and the light has the proper angel of incidence, the infrared radiation undergoes and reflects off the crystal surface rather than leaving it that is called total internal reflection. This is demonstrated in the schematic diagram of an ATR accessory in Figure 2.4.

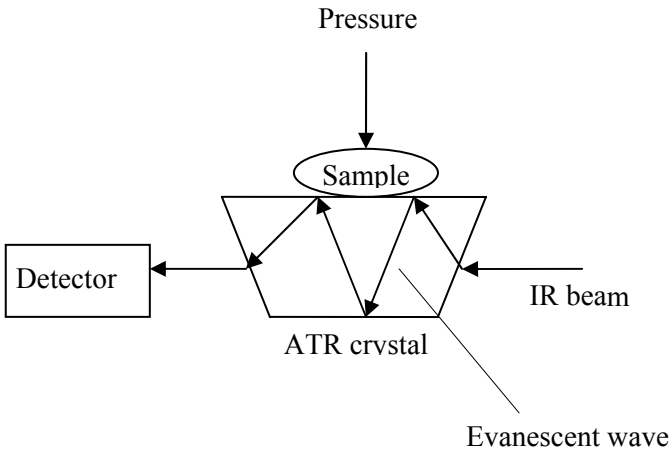

Figure 2.4. A schematic diagram of an ATR accessory

There are several internal total reflections occur within the crystal until the beam reaches the end. It can be shown from Maxwell's equations that superimposition of incoming and reflected waves yields a standing wave within the crystal established normal to the totally reflecting surface. The beam is completely reflected when it collides on the surface of the crystal. 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:

$$\theta_c = \sin^{-1} n_{21} \tag{2.8}$$

Furthermore, when the radiation is inside the crystal, a standing wave of radiation is set up which is called evanescent wave. The evanescent wave is attenuated by the sample's absorbance that the name attenuated total reflectance (ATR) came from. Evanescent wave is characterized by its amplitude which falls exponentially with the distance from the interface represented as:

$$E = E_0 e^{-z/dp} \tag{2.9}$$

where E_0 is 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 dp is the penetration depth of the evanescent field which is given by:

$$d_p = \frac{\lambda 1}{2\pi(\sin^2\theta - n^2{}_{21})^{1/2}} \tag{2.10}$$

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. From this equation it can be drawn that the sample has to be in close contact with the crystal. In addition, from Equation 2.10 the band intensity will depend on the wavelength since the penetration depth, and so the interaction with the sample, increases with λ (Carolei and Gutz 2004, Smith 1996).

2.5. Molecular Fluorescence Spectroscopy

Fluorescence spectroscopy has become a popular spectroscopic technique due to its high sensitivity and selectivity. Fluorescence spectroscopic measurements can be carried out from simple of steady-state emission intensity to quite sophisticated time-resolved measurements. Although, fluorescence measurements do not provide detailed structural information fluorescence spectroscopy is gaining interest in many areas of science for quantitative analysis of complex mixtures with the help of chemometrics. Fluorescence occurs in simple as well as in complex gaseous, liquid, and solid chemical systems. While fluorescence can be observed from almost all molecules with an excitation beam in adequate intensity only a small part of molecules demonstrate fluorescence characteristics which are desirable for the analytical purposes. Therefore, fluorescence spectroscopy is less universal than absorption techniques although it is more selective. However, in some applications in terms of its lower detection limits and greater selectivity, fluorescence spectrometry is a preferred technique than molecular absorption spectrometry (Valeur 2002, Ingle 1988).

2.5.1. Theory of Fluorescence Spectroscopy

Fluorescence is a radiational transition between electronic states of the same multiplicity and involves a singlet-singlet transition which is illustrated by a simple Jablonski diagram in Figure 2.5.

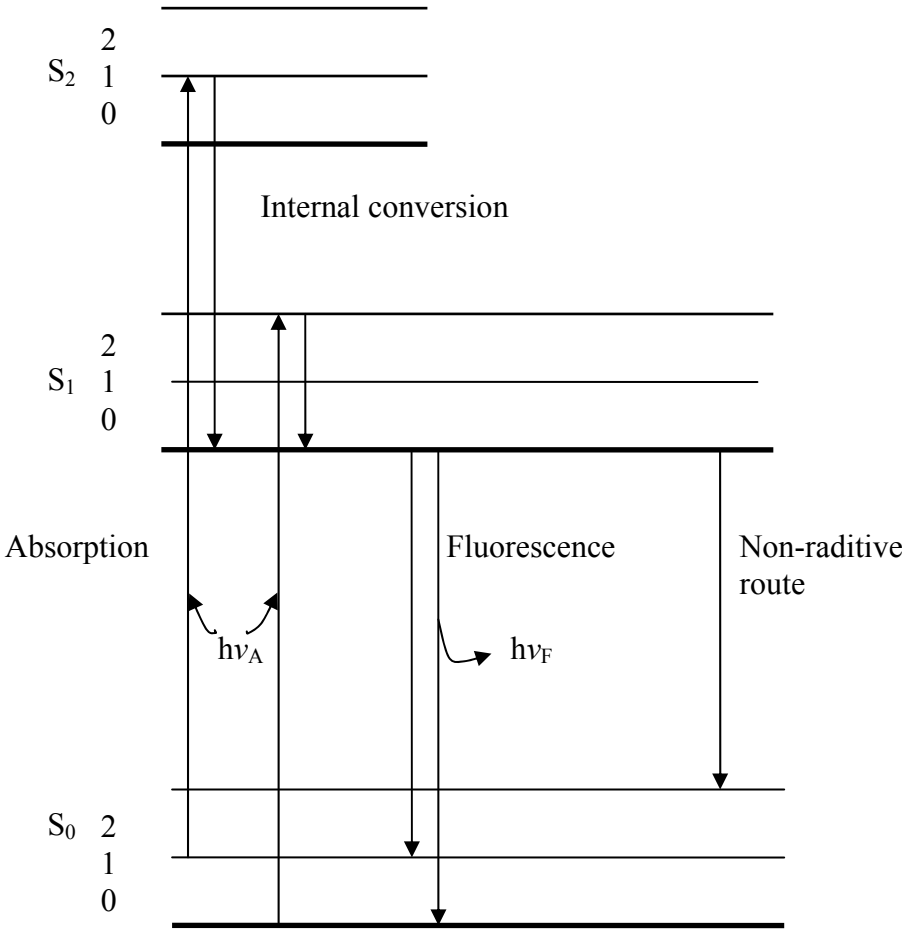

Figure 2.5. A simple Jablonski energy diagram for fluorescent molecules

In this diagram, the electronic singlet states S_0, S_1 and S_2 along with three vibrational energy levels are shown. $h\nu_A$ and $h\nu_F$ symbolizes absorption and fluorescence, respectively. Fluorescence occurs from the ground vibrational state of S_1 to various vibrational levels in S_0 or to higher vibrational levels in the S_1 level. It can also occur from S_2 to S_0 level. Fluorescence usually appears at longer wavelengths than absorption as absorption transitions are higher excited electronic states. The internal conversion resulting in fast transition when a molecule relaxes from the second

vibrational level of S_2 to the first excited singlet state of S_1. Emission occurs from the lowest vibrational level of the lowest excited singlet state because relaxation from the excited vibrational levels is much faster than emission. Therefore fluorescence spectrum is generally independent of the excitation wavelength. After emission the molecule returns to the ground state. This completes the simplest case of fluorescence: excitation, internal conversion, emission and relaxation (Skoog and Leary 1992, Ingle 1988).

2.5.2. Instrumentation

All fluorescence instruments contain three basic items including a source of light, a sample holder, excitation and emission monochromators, and a detector. To be of analytical use, the wavelength of incident radiation needs to be selectable and the detector must be capable of precise manipulation and presentation. The schematic block diagram of a typical fluorescence spectrometer is shown in Figure 2.6.

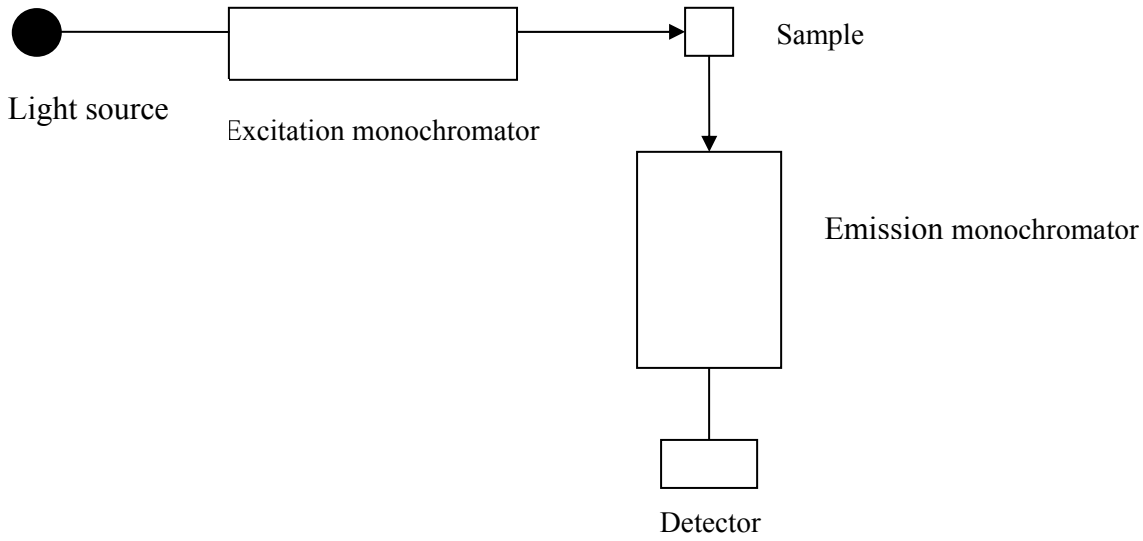

Figure 2.6. A schematic diagram of a fluorescence spectrometer

Simple fluorescence spectrometers have a means of analysing the spectral distribution of the light emitted from the sample, the fluorescence emission spectrum, which may be by means of either a continuously variable interference filter or a monochromator. In more sophisticated instruments, monochromators are provided for both the selection of exciting light and the analysis of sample emission. Such

instruments are also capable of measuring the variation of emission intensity with exciting wavelength, the fluorescence excitation spectrum. The greatest sensitivity can be achieved by the use of filters, which allow the total range of wavelengths emitted by the sample to be collected, together with the highest intensity source possible. Filter fluorimeters, the wavelengths of excited and emitted light are selected by filters which allow measurements to be made at any pair of fixed wavelengths. In practice, to realize the full potential of the technique, only a small band of emitted wavelengths is examined and the incident light intensity is not made excessive, to minimize the possible photodecomposition of the sample.

Different modes of excitation are illustrated in Figure 2.7. To the left, a dilute solution, to the middle, concentrated solution or solid sample and to the right transparent solid sample in a standard 1 x 1 cm cuvette are shown.

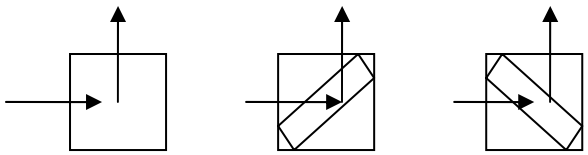

Figure 2.7. Three modes of excitation in a standart cuvette

Most measurements are done as the left way. Horizontal arrow illustrates excitation light and the vertical arrow fluorescence. In very concentrated solutions all the light gets absorbed right at the edge of the cuvette and the measured intensity drops very fast across the cuvette. In this case front-face illumination can be used, as shown in the middle. With the front-face illumination, there is a risk that the reflected light from the surface and back surface interfere with each other to give regular peaks. Moreover, large amounts of light may be reflected directly into the emission monochromator that may result in large amount of stray light.

2.5.3. Excitation Emission Matrix Fluorescence (EEMF) and Synchronous Fluorescence Spectrometry

For the analysis of multifluorophoric systems the widely used methods are the excitation emission matrix fluorescence (EEMF) spectroscopy and synchronous

fluorescence spectroscopy (SFS). EEMF spectroscopy is a rapid and inexpensive technique. The EEMF analysis provides a "fingerprint" consisting of a 3-D emission-excitation intensity diagram. This "fingerprint" along with multivariate calibration can be used for the qualitative and quantitative information about the multifluorophores present in the sample. The data generated by EEMF (fluorescence emission spectra measured at several excitation wavelengths for several samples) method, having at least three dimensions, can be considered as one of the most suitable types of data for *N*-dimensional analysis. EEMF provides a three-way data set, in which each sample gives rise to a data matrix (Divya and Mishra 2006).

Synchronous fluorescence spectroscopy is a highly sensitive and simple technique. In synchronous fluorescence (SYF) spectroscopy both the excitation and emission monochromators are simultaneously scanned at a constant wavelength interval between emission and excitation wavelengths ($\Delta\lambda$), so that spectral overlaps are reduced and the spectra is simplified (Poulli, et al. 2006).

CHAPTER 3

MULTIVARIATE CALIBRATION USED IN SPECTROSCOPY

In the past, performing chemical analysis was tiring and time consuming as analytical instruments were primitive and most analyses were carried out using wet chemistry. Nevertheless, over the years, enormous developments in spectroscopy, has allowed analytical chemists to obtain both quantitative and qualitative information about a sample of interest. The information of an analyte that is desired to know is usually its chemical quantity or concentration in quantitative analysis. However, this cannot be measured directly with the use of spectroscopic techniques. Because direct measurements are impossible with spectroscopic analysis, a calibration process is required. Thus, the concentration of the analyte, x, can be indirectly determined by the calibration function y = f (x), with the use of another physical quantity, y. The physical quantity is measured on a selected number of samples (standards) where the content of the analyte is known. Frequently the calibration function is linear and can be obtained by an mathematical model relating the measured quantities with the corresponding chemical quantities:

$$y = a + bx \tag{3.1}$$

where a and b are the regression coefficients, intercept and slope of the straight line. Finally, in order to obtain the value of the unknown sample concentration of the new sample, the inverse of Equation 3.1 is used:

$$x = (y-a)/b \tag{3.2}$$

In addition, calibration is an important step in chemical analysis since a good accuracy and precision can only be achieved with a good calibration model. There are many different calibration methods have been developed for spectrochemical analysis and generally they are divided into two types as univariate and multivariate calibration

methods. While univariate calibration is useful to determine the concentration of a single compound using one wavelength, multivariate calibration provides possibility to determine the concentration of a multi-component mixture using all or several of the wavelengths instead of one wavelength. Before understanding the development of multivariate calibration method, it is useful to investigate the univariate calibration methods (Massart, et al. 1988, Brereton 2003).

3.1. Univariate Calibration Methods

Univariate calibration which is described as zero-order calibration, involves the use of single measurements from an instrument that all observations or responses depend on a single variable, x. It is usually called as simple linear regression which is the simplest method for obtaining concentration information from the instrument data. This method is generally have been used for quantitative analysis in many spectroscopic techniques such as UV-Vis, IR and NIR spectroscopy, where the relationship between the concentration of an analyte and the instrumental response is expressed by Lambert Beer´s law. This calibration method also requires that the instrument's response only depend on the analyte of interest without interference. Univariate calibration process is divided into two steps called calibration and prediction. The first step includes a data set that contains measurements on a set of known samples which is called as training or calibration set. This step is the most time consuming step in terms of preparing this set of known samples named reference samples. The calibration set consists of an absorbance matrix containing instrumental spectra that are previously measured and a concentration matrix containing concentration values which have been determined by a reliable, independent method (Massart, et al. 1988).

The absorbance matrix consists of spectral data, can be illustrated with the use of row-wise organization where each spectrum is placed as a row vector:

$$\begin{matrix} A_{11} & A_{12} & A_{13} & \dots & A_{1x} \\ A_{21} & A_{22} & A_{23} & \dots & A_{2x} \\ \dots & \dots & \dots & \dots & \dots \\ A_{m1} & A_{m2} & A_{m3} & \dots & A_{mx} \end{matrix}$$

Figure 3.1. Absorbance matrix

where A_{mx} is the absorbance for sample m at the x^{th} wavelength. Similarly the concentration matrix consists of the concentration data, can be illustrated through a row vector:

$$\begin{matrix} C_{11} & C_{12} & C_{13} & \dots & C_{1n} \\ C_{21} & C_{22} & C_{23} & \dots & C_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ C_{m1} & C_{m2} & C_{m3} & \dots & C_{mn} \end{matrix}$$

Figure 3.2. Concentration matrix

where C_{an} is the concentration for sample m of the n^{th} component (Kramer 1998).

The calibration or training set is used to develop a calibration model in order to predict the analyte concentration of an unknown sample in the second, prediction step. This step is demonstrated in Figure 3.3.

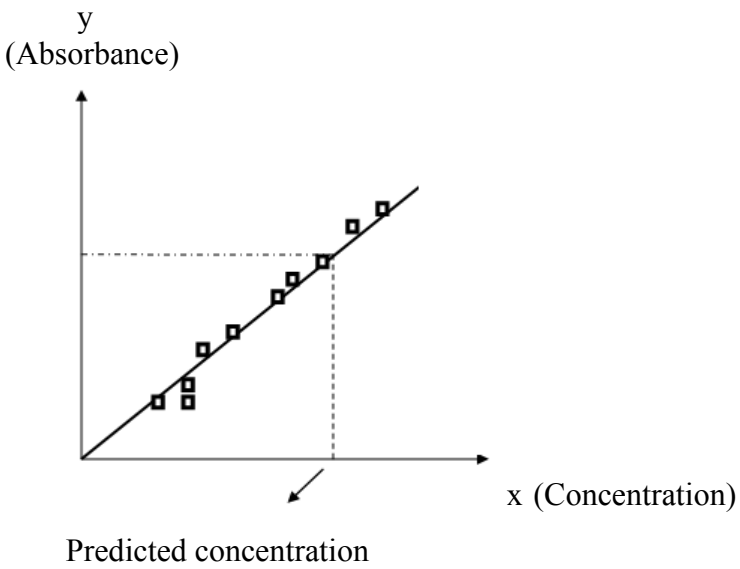

Figure 3.3. A calibration graph for a set of unknown sample

For a univariate calibration model, the instrumental response, s, at a constant frequency is related to the analyte concentration, c, can be represented by the calibration function as:

$$s = f(c) + e_s \tag{3.3}$$

where e_s is the error associated with the instrument response.

In spectroscopic analysis, the relationship between instrument response and analyte concentration described by $f(c)$ is assumed to be linear according to Beer's law which states that the measured absorbance value for a given spectral band of an analyte of interest is proportional to the analyte concentration at maximum absorbing wavelength or wavelengths. The most common univariate calibration methods are the classical and inverse univariate calibrations according to this linear relationship between the instrument response and analyte concentration (Massart, et al. 1988, Brereton 2003).

3.1.1. Classical Univariate Calibration

The classical univariate calibration method uses the statistical model, which assumes Beer's law. This model can be demonstrated by the equation as:

$$a_i = b_0 + b_1 c_i + e_i \tag{3.4}$$

where the instrumental response and analyte concentration for i^{th} sample of m calibration samples are represented by a_i and c_i, and the error associated with instrumental response for each analyte is represented by e_i, which is often called as residual. In ideal case, there is no error and this equation gives a straight line where b_0 and b_1 are the intercept and slop. In fact, there is no such an ideal case in practice and there is always some kind of error associated with instrumental response. Hence, the common practice in quantitative spectroscopic analysis is plotting the instrumental responses against the analyte concentration for a set of m calibration sample and producing a straight line that is best fit to the plotted data by estimating b_0 and b_1 using the method of least squares. The least squares estimates of b_0 and b_1 are found by minimizing the sum of the squares (SS) of the residuals. The sum of the squares (SS) function can be defined by rearranging Equation 3.4 for m calibration sample as:

$$SS = \sum_{i=1}^{m} e_i^2 = \sum_{i=1}^{m} (a_i - b_0 - b_1 \cdot c_i)^2 \tag{3.5}$$

Minimizing SS, which means producing the least possible value of SS needs to be taken its partial derivatives with respect to b_0 and b_1 that are being estimated and the results are set to zero as:

$$\frac{\partial SS}{\partial b_1} = -2 \sum_{i=1}^{m} (a_i - b_0 - b_1 \cdot c_i) = 0 \tag{3.6}$$

and

$$\frac{\partial SS}{\partial b_1} = -2 \sum_{i=1}^{m} c_i (a_i - b_0 - b_1 \cdot c_i) = 0 \tag{3.7}$$

Thus, the estimates of b_0 and b_1 can be solved by these two equations:

$$\sum_{i=1}^{m} (a_i - b_0 - b_1 \cdot c_i) = 0 \tag{3.8}$$

and

$$\sum_{i=1}^{m} c_i (a_i - b_0 - b_1 \cdot c_i) = 0 \tag{3.9}$$

Then from these equations we receive:

$$\sum_{i=1}^{m} a_i - mb_0 - b_1 \sum_{i=1}^{m} c_i = 0 \tag{3.10}$$

and

$$\sum_{i=1}^{m} c_i a_i - b_0 \sum_{i=1}^{m} c_i - b_1 \sum_{i=1}^{m} c_i^2 = 0 \tag{3.11}$$

or

$$mb_0 + b_1 \sum_{i=1}^{m} c_i = \sum_{i=1}^{m} a_i \tag{3.12}$$

and

$$b_0 \sum_{i=1}^{m} c_i + b_1 \sum_{i=1}^{m} c_i^2 = \sum_{i=1}^{m} c_i a_i \tag{3.13}$$

These equations are defined as normal equations and solution of them gives the least-squares estimates of b_0 and b_1:

$$\hat{b}_1 = \frac{\sum_{i=1}^{m} c_i a_i - (\sum_{i=1}^{m} c_i)(\sum_{i=1}^{m} a_i)/m}{\sum_{i=1}^{m} c_i^2 - (\sum_{i=1}^{m} c_i)^2/m} \tag{3.14}$$

and

$$\hat{b}_0 = \bar{a} - \hat{b}_1 \bar{c} \tag{3.15}$$

where $\bar{a}$ is the mean value of instrumental response and $\bar{c}$ is the mean value of analyte concentration for m calibration samples. Then the estimated calibration equation can be written as:

$$\hat{a} = \hat{b}_0 + \hat{b}_1 \bar{c} \tag{3.16}$$

and the concentration of an unknown sample can be defined as:

$$c_u = \frac{a_u - \hat{b}_0}{\hat{b}_1} \tag{3.17}$$

where c_u is the unknown analyte concentration and a_u is the instrument response of this analyte. The correlation coefficient (R^2) is a numerical measure, which usually called as the multiple correlation coefficient. R^2 expresses the strength of the linear relationship between c and a, and can be calculated as:

$$R^2 = \frac{\sum_{i=1}^{m} (\hat{a} - \bar{a})}{\sum_{i=1}^{m} (a - \bar{a})} \tag{3.18}$$

From this equation a unit free number can be defined. The values for R^2 range from 0 to 1 should as close as 1 for the best fitted straight line. The method of least-squares can also be described using matrix algebra. The model represented for classical univariate calibration can be written as matrix equation by:

$$\mathbf{a} = \mathbf{C}\ \boldsymbol{\beta} + \mathbf{e}_{\mathbf{a}} \tag{3.19}$$

where a is the *m x 1* vector of instrument responses, C is the *m x 2* matrix of analyte concentrations, β is the *2 x 1* vector of regression parameters (b_0 and b_1) and e_a is the *m x 1* matrix of the errors or residuals not fit by the model. The first column of the C matrix is a vector of ones, which is necessary to estimate b_0 when the multiplication is performed. The two normal equations given in Equation 3.6 can be described in matrix notation by:

$$(\mathbf{C}' \cdot \mathbf{C}) \cdot \boldsymbol{\beta} = \mathbf{C}' \cdot \mathbf{a} \tag{3.20}$$

Then the least-squares solution of this equation is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{C}' \cdot \mathbf{C})^{-1} \cdot \mathbf{C}' \cdot \mathbf{a} \tag{3.21}$$

where $\hat{\boldsymbol{\beta}}$ is the *2 x 1* vector of least squares estimates, b_0 and b_1, which are solved earlier in Equation 3.5 using the sum of squared residuals being minimized (Massart, et al. 1988, Brereton 2003).

3.1.2. Inverse Univariate Calibration

Although classical univariate calibration is the most widely used method in analytical chemistry, it is not always the most suitable approach in terms of two reasons. First reason is that in the classical univariate calibration, though, the concentration is predicted from the instrumental response, inverse of this approach is impossible. Classical univariate or direct calibration relates to the regression of the responses on the concentrations, whereas inverse calibration relates to the regression of the concentrations on the responses. The second reason depends on the error distributions. The errors not fit by the model are usually due to the independent variable, often concentration. Classical calibration constructs a model where all errors are in the response Figure 3.4.a. However, after the developments in instrumentation, the more appropriate assumption indicates that errors are primarily in the measurement of concentration Figure 3.4.b.

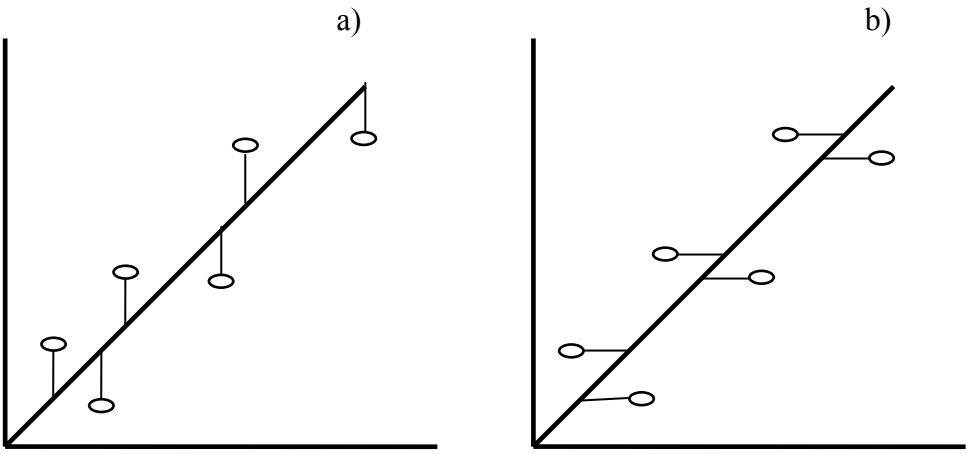

Figure 3.4. Errors in classical (a) and inverse (b) calibrations

All of these reasons arise the need for inverse univariate calibration method. The inverse calibration method assumes the inverse Beer's law and the statistical model is represented by:

$$c_i = p_o + p_1 \cdot a_i + e_i \tag{3.22}$$

where e_i is assumed to be the error associated with the reference value c_i. In the calibration step, the model parameters (p_o and p_1) are estimated by the method of least-squares described earlier. The following equations represent the estimated model parameters that are the slope and the intercept of the calibration line.

$$\hat{p}_1 = \frac{\sum_{i=1}^{m} a_i c_i - (\sum_{i=1}^{m} a_i)(\sum_{i=1}^{m} c_i)/m}{\sum_{i=1}^{m} a_i^2 - (\sum_{i=1}^{m} a_i)^2/m} \tag{3.23}$$

and

$$\hat{p}_o = \bar{c} - \hat{p}_1 \cdot \bar{a} \tag{3.24}$$

where $\bar{a}$ and $\bar{c}$ are the mean values of instrumental responses and analyte concentrations, respectively for m calibration samples. Now the estimated calibration equation can be written as:

$$\hat{c} = \hat{p}_0 + \hat{p}_1 \cdot a \tag{3.25}$$

In the prediction step, concentration of an unknown sample can be calculated by:

$$c_u = \hat{p}_o + \hat{p}_1 \cdot a_u \tag{3.26}$$

where c_u is unknown analyte concentration and a_u is instrument response for that sample. Although the predictions acquired by classical and inverse calibration methods will be different for a sample of interest, in many cases they are not significant. The selection of appropriate univariate calibration method depends on whether the reference values of known samples in other words calibration standards or the instrument responses which are more precise. Although univariate calibration methods offer simpleness for specific types of applications where selective measurements can be found or where the analyte contains no interferences, their applications are limited due to its disadvantages. A disadvantage of univariate calibration methods is in terms of the interference free systems which are rarely met in real applications and another is in terms of the concentrations of the interfering species which are usually unknown and their amounts in sample are not always the same. Moreover, a problem with the use of univariate calibration methods is lack of the constant baseline for every measurement.

In recent years, calibration methods have been increasingly more effective with today's powerful computing resources give the chance to fast quantitative determination of samples which include more than one compenents with complex spectral data. Several different multivariate calibration methods are available in the analysis of such as complex mixtures which have great deal of interference (Brereton 2003).

3.2. Multivariate Calibration

Spectroscopic techniques require the application of multivariate calibration methods to model instrumental responses of a complex sample of interest due to the complexity of their spectra that makes direct interpretation impossible. In addition, multivariate approaches are necessary in order to obtain all the chemical information contained in the spectral variables. Indeed, multivariate calibration is important because it deals with the data containing instrument responses measured on multiple wavelengths for a sample that usually contains more than one component. In addition, multivariate calibration methods are required when the relationship between the analytical signal and the analyte of interest is nonlinear for the analysis of complex mixtures. Major advantage of multivariate calibration is that the analysis of complex mixtures can be achieved without any separation or extraction. The multivariate calibration methods are the classical least squares (CLS), inverse least squares (ILS), principle component analysis (PCA), principle component regression and partial least squares (PCR and PLS), genetic regression (GR), genetic classical least squares (GCLS), genetic inverse least squares (GILS), and genetic partial least squares (GPLS). The choice of the most suitable calibration method is very important in order to generate the best calibration model (Massart, et al. 1988, Brereton 2003).

3.2.1. Classical Least Squares

Classical least squares (CLS) model is described by the classical Beer's law in matrix form as:

$$\mathbf{A} = \mathbf{CK} + \mathbf{E}_{\mathbf{A}} \tag{3.27}$$

where the absorbance is a function of component concentration and is directly proportional to the component concentration for *m* calibration samples with *n* wavelengths. In this equation, A is the *m x n* matrix of the calibration spectra, C is the *m x l* matrix of component concentrations, K is the *l x n* matrix of absorptivity-pathlength constants and E_A is the *m x n* matrix of the spectral errors or residuals not fit by the model. Model errors are assumed to be in the measurement of the instrument responses as it was in the classical univariate method. The K matrix represents the first order estimates of the pure component spectra at unit concentration and unit pathlength. The method of least-squares can be used to estimate K matrix. The least-squares estimate of the K is defined as:

$$\hat{\mathbf{K}} = (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{A} \tag{3.28}$$

Then the estimated $\hat{\mathbf{K}}$ matrix can be used to obtain the concentrations of an unknown sample from its spectrum by:

$$\hat{\mathbf{c}} = (\hat{\mathbf{K}}\hat{\mathbf{K}}')^{-1}\hat{\mathbf{K}}\mathbf{a} \tag{3.29}$$

where a is the vector of unknown sample spectrum and $\hat{\mathbf{c}}$ is the vector of predicted component concentrations.

The model of classical multivariate calibration points out that in order to construct a good model, all the species present in a given sample need to be known included in the calibration step, in the C matrix. This is the major disadvantage of the CLS method since generally concentrations of all species are not known exactly, so the instrument response due to this interfering species can not be modeled thereby causing a large error. However, when the content of the sample is precisely known, CLS offers several advantages as this method can use the full spectrum to build the calibration model compared to the methods that are restricted to single or a small number of wavelengths. Furthermore, in CLS, estimated pure component spectra along with the residuals and simultaneously fitted spectral base lines made this method preferable. Beside advantages of CLS method, the major disadvantage of it, is that all species in the sample and their concentrations included in the model must be known. This need can be eliminated by using Inverse least squares (ILS) method.

3.2.2. Inverse Least Squares

Inverse least squares (ILS) model as understood from its name, is described by inverse Beer's law for *m* calibration samples with *n* wavelengths as:

$$\mathbf{C} = \mathbf{AP} + \mathbf{E}_{\mathbf{C}} \tag{3.30}$$

where C and A are the same as in CLS, P is the *n x l* matrix of the unknown calibration coefficients relating *l* component concentrations to the spectral intensities and E_C is the *m x l* matrix of errors in the concentrations not fit by the model. As can be seen, according to the model, concentrations of the analyte is a function of absorbance. Since modern spectroscopic instruments provide excellent signal-to-noise (S/N) ratios, it is assumed that the great deal of errors lie in the reference or calibration samples, not in the instrumental measurement. The major advantage of ILS is that the equation 3.24 can be reduced for the analysis of single component because of the analysis based on the ILS model is constant respect to the number of components in the sample. The reduced model is given as:

$$\mathbf{c} = \mathbf{Ap} + \mathbf{e}_{\mathbf{c}} \tag{3.31}$$

where c is the *m x 1* vector of concentrations for the analyte that is being analyzed, p is *n x 1* vector of calibration coefficients and e_c is the *m x 1* vector of concentration residuals not fit by the model. During the calibration step, the least-squares estimates of p also called the estimated calibration coefficients symbolized as $\hat{\mathbf{p}}$ can be calculated as:

$$\hat{\mathbf{p}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \cdot \mathbf{c} \tag{3.32}$$

Then from this equation the concentration of the analyte can be predicted as:

$$\hat{c} = \mathbf{a}' \cdot \hat{\mathbf{p}} \tag{3.33}$$

where $\hat{c}$ is the scalar estimated concentration and a is the spectra of the unknown sample. The ability to predict one component at a time without knowing the concentrations of interfering species has made ILS one of the most preferable calibration method (Özdemir 2006).

The major disadvantage of ILS as illustrated in Equation 3.30 is that the matrix, which must be inverted has the dimensions equal to the number of wavelengths in spectra and this number can not exceeded the number of calibration samples. This is a big restriction since the number of wavelengths in a spectrum will generally be more than the number of calibration samples and selection of wavelengths that provides best fit for the model is an important part of the process. Collinearity of wavelengths that are not independent of each other, is also a problem, as it increases the spectral overlaps. Thus, the full spectrum advantage of CLS, where precision of the analysis is significantly improved is impossible with ILS. Several wavelength selection ways such as stepwise wavelength selection and all possible combinations are available to build an ILS model that fits the data best. However, if there is overfitting of the data, even a good calibration model would not produce reasonable predictions.

In last years, new calibration methods have been developed which use genetic algorithms (GAs). These methods are available to solve the calibration problems in many fields in spectroscopy. One of the new calibration method involving GAs is the Genetic Inverse Least Squares (GILS). Before understanding this new technique it is useful to explain the principle and importance of GAs.

3.3. Genetic Algorithms

The term Genetic Algorithm (GA) describes optimization methods which are effective to solve complex problems such as wavelength selection problems from a large spectrum of data whereas the conventional methods cannot offer a sufficient solution to this problem. As the name suggests, the processes employed by GAs are based on natural evolution and selection. According to this processes a population of possible solutions to a problem is generated. These processes imitate those in nature so that subsequent populations fit better and adapt in their environment and are passing their genetic informations to their offspring as a result of breeding. On the contrary, who are not fit and adapt in their environment will be eliminated from the population.

As a long period of time progresses, generations become better suited to their environment and provide more optimal solutions. Computational approach of a typical GA is quite simple and includes five basic steps as shown in Figure 3.5.

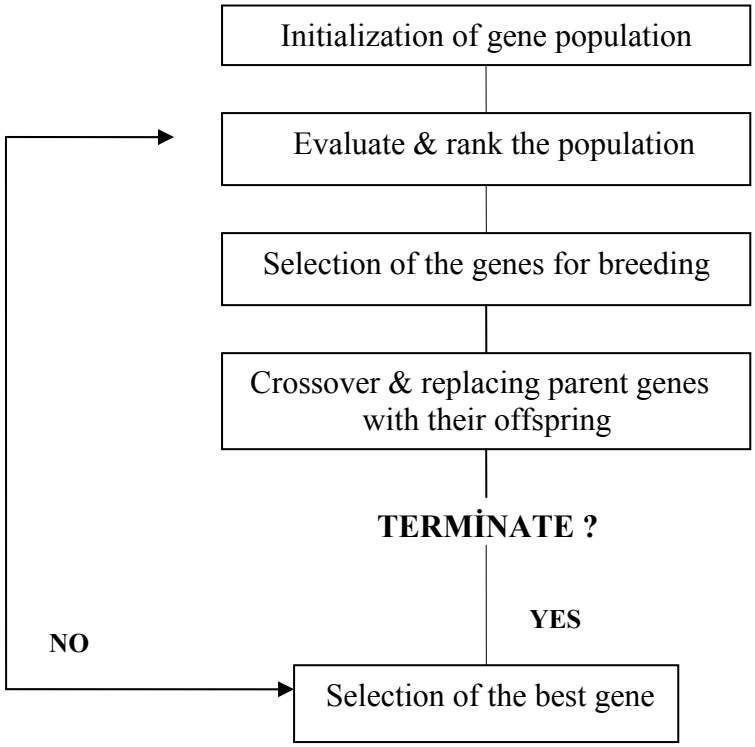

Figure 3.5. Flow chart of a genetic algorithm

These steps consist of initialization of a gene population, evaluation of the population, selection of the parent genes for breading and mating, crossover and replacing parents with their offspring. The name of these steps arise from the biological feature of the genetic algorithm.

3.4. Genetic Inverse Least Squares (GILS)

GILS is a modified version of the ILS multivariate calibration model. As mentioned before GILS method uses GAs in the selection of the wavelengths to create a calibration model with reduced data set. A gene is a potential solution of a given problem which changes from application to application. In the GILS method, the term ‘gene’ is referred as the collection of instrumental response at the wavelength range of the data set. Each

gene produces a subspectrum at a few wavelengths of the full spectrum, which relates to the component concentration. The term 'population' is referred as the collection of individual genes in the current generation.

GILS, which is unique in the encoding genes, depends on the same basic GAs steps of initialization, evaluation, selection, crossover and replacing.

3.4.1. Initialization

The initialization step is the first step where the first generation of genes are randomly created with a fixed population size. The number of the gene pool size is a user optional parameter which allows breeding of each gene in the population. Hence, as the population size is larger, the computation time is longer since the large number of genes result in lower computation speed. The number of instrumental responses found in a gene is determined randomly between a fixed low limit and high limit. The lower limit was set to 2 to allow single point crossover whereas the higher limit was set to eliminate overfitting problems and reduce the computation time.

3.4.2. Evaluate and Rank the Population

After the initilization of a gene population the next step is to evaluate and rank these genes with the use of a fitness function. The value of the fitness function of the each gene also show their success for the calibration model. The value of the fitness function is obtained by the inverse of the standard error of calibration (SEC):

$$\text{Fitness} = 1/\text{SEC} \tag{3.34}$$

The SEC is calculated from the derivative of the standart error (SE) as:

$$\text{SE} = \sqrt{\frac{\sum_{i=1}^{m} (\hat{c}_i - c_i)^2}{df}} \tag{3.35}$$

where $\hat{c}_i$ and c_i are the predicted and known analyte concentrations for m samples and df is the degrees of freedom, which is given by:

$$df = m\text{-}k \tag{3.36}$$

where k is the number of parameters extracted from the data set. In a calibration data set, for a linear model it is assumed that there are two parameters to be extracted. These parameters are the slop of the line and the intercept. In this case, the degrees of freedom would be equal to m-2. Thus, the standard error of calibration (SEC) is represented as:

$$\text{SEC} = \sqrt{\frac{\sum_{i=1}^{m} (\hat{c}_i - c_i)^2}{m - 2}} \tag{3.37}$$

if m-2 is replaced in the Equation 3.35 for the term df.

3.4.3. Selection of the Genes for Breeding

The third step depends on the selection of the parent genes from the current population for breeding. The selection is made by using a selection method according to their fitness values. The aim of a selection method is to give the genes higher chance to breed with higher fitness values so that the best performing genes of the population will survive and will generate better offspring to pass their information to the next generations. Therefore, the genes with the low fitness values will have lower chance to breed and as a consequence, most of them will be unable to survive.

There are many selection methods that can be used for the selection of parent genes. The simplest selection method is the top down selection where the genes are allowed to mate after they ranked in the current population in a way that the first gene mates with the second gene, the third gene with the forth one and so on until all genes of the current population got a chance to breed. Another selection method, which is used in the GILS is the Roulette wheel selection method. In this method an imaginary roulette wheel is constructed with different size of segments where each segment illustrates a gene as shown in the Figure 3.6. The selection is made randomly by rotating the wheel a number of times that is equal to the population size. It is expected that the gene with the highest fitness value has the biggest segment, on the contrary, the lowest value has the

smallest segment. Thus, when the wheel is rotated there is higher chance of being selected for a gene with higher fitness value than for a gene with a low fitness value. When the wheel stops the segment of a gene which the selection point shows is selected. In this segment there will be also the genes, which are selected multiple times while some of them will not be selected at all and will be removed from the gene pool.

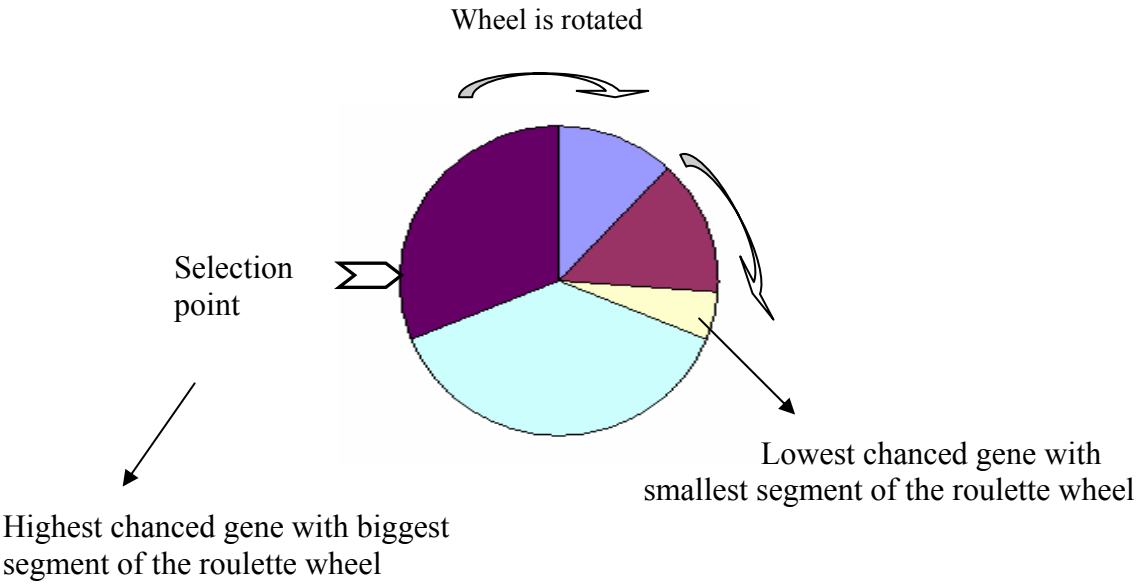

Figure 3.6. Roulette wheel selection

3.4.4. Crossover and Replacing Parent Genes by Their Offspring

After the selection of parent genes is completed, all of them are mating to produce their offspring by crossing over until there is no more rest. For example, if two parent genes of S_1 and S_2 are to be selected, the first part of S_1 is combined with the second part of S_2 likewise the second part of S_1 is combined with the first part of the S_2 to produce the offspring S_3 and S_4 as illustrated in Figure 3.7.

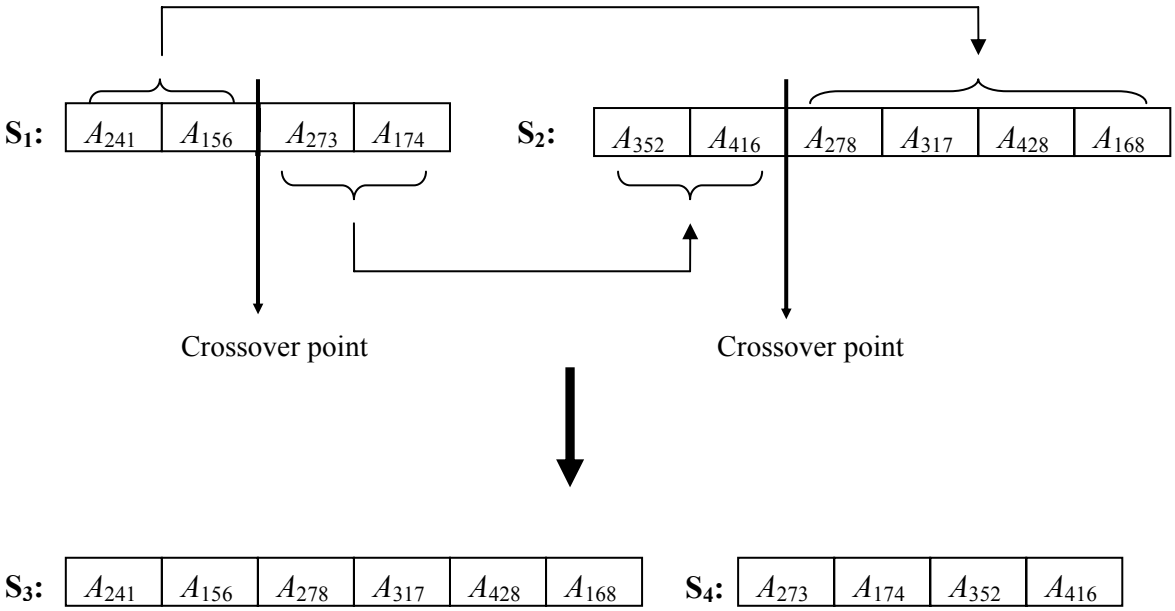

Figure 3.7. Mating of the parent genes

Here A represents the instrument response at the given wavelength which is represented with the subscript of A. After crossover procedure, the parent genes are replaced by their offspring and the offspring are evaluated. Following the evaluation step they are ranked according to their fitness values and then the circle of selection for breeding and mating starts all over again. This is repeated until the number of the predefined iteration value is finished.

Finally, the gene with the lowest SEC, in other words with the highest fitness value is selected to build a calibration model in order to predict the concentrations of components being analyzed in the prediction sets (validation sets). The success of the model in the prediction sets is evaluated according to the standard error of prediction (SEP) values, which is calculated as:

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^{m} (\hat{c}_i - c_i)^2}{m}} \tag{3.38}$$

where m is the number of prediction (validation) samples (Özdemir and Dinç 2005).

3.5. Advantages of GILS

GILS is a preferred multivariate calibration method in terms of its some major advantages over the other calibration methods either univariate or multivariate. One of the advantage is its simplicity in the mathematics of the model building and prediction steps. In addition, it has the advantage of the multivariate calibration methods related to the selection of a small data set of wavelengths from a full spectral data set and solve the wavelength selection problems. Nevertheless, selecting reduced data set of instrument responses makes it possible to eliminate nonlinearities that might be present in the full spectral region (Özdemir and Betül 2007).

CHAPTER 4

DETERMINATION OF DIESEL ADULTERATION BY NIR SPECTROSCOPY COMBINED TO MULTIVARIATE CALIBRATION

NIR spectroscopy has found a wide application area in fuel analysis. The chemical stability of NIR spectroscopy in different solvents and the high light throughput provided by its optical fibers is the main factor for its large application in fuel analysis (Kelly, et al. 1989, Parisi, et al. 1990, Zetter, et al. 1993, Cooper et al. 1996, Breitkreitz, et al. 2003, Mendes et al. 2003) including on line monitoring of fuels (Parisi, et al. 1990, Zetter, et al. 1993, Kim, et al. 2000).

This spectroscopic method when combined to multivariate calibration have been demonstrated to be an effective technique in the analysis of fuel samples such as diesel (Santos 2005, Breitkreitz 2003, Lima and Borges 2002) and kerosene (Chung, et al. 1999, Garrigues, et al. 1995) since it presents good accuracy and precision besides being faster than the usual methods such as filter test (Roy 1999), density test (Yadav, et al. 2005), and American Standarts for Testing Materials (ASTM) methods (Oliveira et al. 2007, Roy 1999).

Since the use of NIR spectroscopy combined to multivariate calibration in the quantitatively determination of biodiesel and vegetable oils that are not converted to biodiesel in petroleum diesel is very recent, several reports can be found in the literature. Santiago et al. and Aliske et al. (2007) reported that the determination of biodiesel-diesel mixtures concentrations is possible by using NIR spectroscopy. Pimentel et al. (2006) determined the concentration of biodiesel in blends with petroleum diesel, in the presence of raw vegetable oil by NIR spectroscopy combined to multivariate calibration methods. Oliveira et al. (2007) used FT-NIR spectroscopy combined to multivariate calibration methods to determine the presence of vegetable oils and biodiesel in petroleum diesel. The aim of this study was to determine the content of different vegetable oils, used frying oil, biodiesel, and kerosene in petroleum diesel by NIR spectroscopy combined to multivariate calibration.

4.1. Instrumentation and Data Pre-Processing

In this study near-infrared spectroscopic analyses were performed by a FTS-3000 NIR spectrometer (Bio-Rad, Excalibur, Cambridge, MA). Spectra were taken between 4,000 and 10,000 cm^{-1} wavenumber range with a wavelength interval of 4 cm^{-1}. This spectrometer was equipped with 250 W tungsten-halogen lamp as a source, calcium fluoride (CaF_2) as a beam splitter, and lead selenide (PbSe) as a detector. Resolution was optimized to the 16 cm^{-1} and 128 scans were done. Samples were held in a 1.00 cm pathlength infrasil quartz sample holder from Starna (Atascadero, CA). Backround was taken with an empty infrasil quartz sample holder. Triple measurements have done for each sample and the means of the measurements were used in multivariate analyses. All spectra collected from the instrument were transfered to a computer to develop a calibration model for the prediction. Microsoft Excel (MS Office 2003, Microsoft Corporation) program was used to prepare the text files for calibration and validation sets, which are required to employ and test the multivariate calibration method used in this study. The genetic algorithms based genetic inverse least squares (GILS) multivariate calibration method was written in MATLAB programming language using Matlab 5.3 (MathWorks Inc., Natick, MA) and employed to set up the calibration models. The first step in the development of a calibration model is the design of calibration set. In the design of calibration set it is important to chose the samples that have maximum and minimum concentration values. In addition, the success of model in prediction can be tested by validation (prediction) set. The samples in calibration and validation sets were selected randomly. The unit of concentration values was taken as mass percent.

4.2. Sample Preparation and Design of the Data Sets

Different diesel adulteration with sunflower, canola oil, used frying oil, biodiesel, and kerosene were prepared to determine their content in petroleum diesel blends. Sunflower, and canola oils were bought from a local supermarket. Used frying oil was acquired from a restaurant in Izmir. Biodiesel was obtained from a company named Ege Biotechnology Ltd. (Izmir Institute of Technology, Technopark, TR). Euro diesel, diesel, and kerosene were purchased from a service station in Izmir, where the public

provides its fuel. In order to construct calibration models four different sets were prepared including binary mixture of diesel with sunflower oil and ternary mixtures of diesel with sunflower oil, used frying oil, canola oil, biodiesel, and kerosene.

The first set involves 30 binary mixtures of sunflower oil and diesel. The concentration data corresponding to each component of each set were divided into two sets which are called calibration and validation (prediction) sets. Table 4.1 to Table 4.8 illustrate the concentrations of each component for the calibration and validation sets. All concentrations on the tables are given in the mass percentage. The calibration set composed of 20 samples, and validation set composed of 10 samples. The concentrations of sunflower oil in diesel were in the mass range between 0.65% and 39.56% and diesel were in the mass range between 60.43% and 99.35%.

Table 4.1. Concentration profile of the calibration samples in the first set

Sample No	Sunflower oil (w/w%)	Diesel (w/w%)	Sample No	Sunflower oil (w/w%)	Diesel (w/w%)
1	1.50	98.50	11	20.85	79.15
2	26.23	73.76	12	26.56	73.43
3	7.15	92.85	13	22.81	77.18
4	39.56	60.43	14	24.25	75.75
5	21.18	78.81	15	22.82	77.17
6	7.10	92.90	16	0.65	99.35
7	11.20	88.79	17	19.95	80.05
8	28.76	71.23	18	27.40	72.60
9	26.95	73.05	19	31.55	68.45
10	15.20	84.79	20	38.23	61.76

Table 4.2. Concentration profile of the validation samples in the first set

Sample No	Sunflower oil (w/w%)	Diesel (w/w%)	Sample No	Sunflower oil (w/w%)	Diesel (w/w%)
1	2.80	97.20	6	35.91	64.08
2	18.09	81.90	7	7.35	92.65
3	24.05	75.95	8	30.10	69.90
4	35.81	64.18	9	23.48	76.51
5	2.95	97.04	10	6.90	93.09

The second set contains 30 ternary mixtures of sunflower oil, used frying oil, and diesel. This set includes differently from the first set used frying oil since diesel fuel may have possible adulteration with used frying oil that was not converted into biodiesel. The calibration set composed of 20 samples, and validation set 10 samples. The concentrations of sunflower oil in diesel were in the mass range between 2.15% and 24.71%, used frying oil were in the mass range between 0.54% and 23.56%, and diesel were in the mass range between 52.02% and 93.79%.

Table 4.3. Concentration profile of the calibration samples in the second set

Sample No	Sunflower oil (w/w%)	Used frying oil (w/w%)	Diesel (w/w%)	Sample No	Sunflower oil (w/w%)	Used frying oil (w/w%)	Diesel (w/w%)
1	2.89	12.09	85.00	11	7.05	23.56	69.38
2	14.60	5.60	79.80	12	18.15	8.30	73.55
3	16.24	3.94	79.81	13	11.90	19.00	69.08
4	18.70	15.05	66.25	14	24.71	23.26	52.02
5	11.15	12.65	76.18	15	4.80	1.40	93.79
6	20.32	2.64	77.02	16	8.25	0.55	91.20
7	20.21	8.75	71.03	17	20.53	0.54	78.91
8	9.64	14.74	75.61	18	23.28	18.84	57.87
9	3.49	5.09	91.40	19	2.15	9.40	88.44
10	22.25	16.15	61.60	20	11.45	9.30	79.23

Table 4.4. Concentration profile of the validation samples in the second set

Sample No	Sunflower oil (w/w%)	Used frying oil (w/w%)	Diesel (w/w%)	Sample No	Sunflower oil (w/w%)	Used frying oil (w/w%)	Diesel (w/w%)
1	21.18	17.44	61.36	6	5.30	4.70	90.00
2	9.05	6.90	84.03	7	16.59	3.94	79.46
3	3.15	14.55	82.30	8	4.55	18.55	76.88
4	23.08	6.49	70.41	9	5.60	7.85	86.54
5	22.73	0.49	76.76	10	6.74	5.85	87.39

Then the third set was prepared that containes 30 ternary mixture of kerosene, euro diesel, and normal diesel due to the possible adulteration of diesel with kerosene. The calibration set in the first set composed of 20, validation set 10 samples. The concentrations of kerosene in diesel were in the mass range between 3.99% and 49.00%, euro diesel were in the mass range between 1.00% and 49.02% and normal diesel were in the mass range between 3.90% and 92.00%.

Table 4.5. Concentration profile of the calibration samples in the third set

Sample No	Kerosene (w/w%)	Euro Diesel (w/w%)	Normal Diesel (w/w%)	Sample No	Kerosene (w/w%)	Euro Diesel (w/w%)	Normal Diesel (w/w%)
1	45.97	1.00	53.02	11	49.00	32.01	18.98
2	25.98	17.99	56.02	12	30.96	17.95	51.07
3	10.99	4.04	84.95	13	25.93	49.02	25.03
4	34.03	1.04	64.91	14	37.03	47.02	15.94
5	47.00	48.00	5.00	15	3.99	42.97	53.02
6	25.01	11.00	63.98	16	6.00	12.00	81.99
7	47.00	42.05	10.95	17	44.97	7.99	47.02
8	44.97	13.04	41.97	18	15.99	26.03	57.97
9	45.97	50.07	3.95	19	13.95	18.00	68.05
10	36.96	21.96	41.07	20	35.00	9.95	55.05

Table 4.6. Concentration profile of the validation samples in the third set

Sample No	Kerosene (w/w%)	Euro Diesel (w/w%)	Normal Diesel (w/w%)	Sample No	Kerosene (w/w%)	Euro Diesel (w/w%)	Normal Diesel (w/w%)
1	18.99	5.04	75.96	6	12.00	26.96	61.03
2	38.96	16.03	45.00	7	47.05	19.00	33.95
3	15.04	48.97	35.98	8	23.03	19.94	57.02
4	21.97	13.03	64.98	9	42.95	36.05	21.00
5	2.00	5.95	92.05	10	33.00	36.00	31.00

The fourth set consists of 50 ternary mixtures of canola oil, biodiesel (canola oil methyl ester), and diesel. The concentrations of canola oil, biodiesel, and diesel were in the mass range between 0% and 100%.

Table 4.7. Concentration profile of the calibration samples in the fourth set

Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)	Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)
1	100	0.00	0.00	16	9.90	14.61	75.47
2	0.00	100	0.00	17	10.33	67.17	22.49
3	0.00	0.00	100	18	25.97	26.16	47.86
4	37.20	22.88	39.91	19	8.89	47.05	44.05
5	32.43	57.78	9.78	20	39.40	20.30	40.30
6	31.43	25.77	42.79	21	35.75	22.96	41.28
7	5.13	12.83	82.03	22	9.67	38.62	51.70
8	59.71	9.72	30.55	23	36.46	44.05	19.48
9	38.13	2.38	59.48	24	16.61	78.17	5.20
10	11.02	44.99	43.98	25	27.08	24.30	48.61
11	15.74	83.05	1.20	26	33.46	59.11	7.41
12	4.66	40.42	54.91	27	25.94	45.30	28.74
13	16.68	20.36	62.94	28	52.24	32.36	15.38
14	14.57	16.33	69.09	29	2.10	46.28	51.60
15	42.19	57.00	0.79	30	83.11	1.09	15.78

Table 4.8. Concentration profile of the validation samples in the fourth set

Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)	Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)
1	23.84	6.53	69.61	11	57.92	20.12	21.95
2	56.88	9.71	33.40	12	37.40	10.59	52.00
3	17.34	24.30	58.35	13	34.06	9.01	56.91
4	33.57	60.93	5.49	14	67.61	24.78	7.60
5	44.36	23.23	32.39	15	45.55	8.09	46.35
6	72.28	9.03	18.68	16	14.16	42.49	43.33
7	36.62	43.81	19.55	17	50.60	15.20	34.20
8	6.48	63.21	30.30	18	35.42	39.74	24.82
9	41.45	26.49	32.04	19	29.45	37.20	33.33
10	34.09	27.46	38.43	20	24.12	41.44	34.43

4.3. Results and Discussion

Near infrared spectra of pure sunflower oil, used frying oil, canola oil, biodiesel, kerosene, and diesel samples are shown in Figure 4.1 and 4.2. As it can be seen from the Figure 4.1 the spectral bands of sunflower oil, used frying oil, canola oil, and biodiesel, which was derived from canola oil were overlapped due to their resembling main composition. In addition, the spectral bands of kerosene, normal diesel, and euro diesel in the Figure 4.2 were also much alike since all of them are petroleum products and their main composition is similar. Only small differences exist in some parts of the whole spectrum where the spectral bands of kerosene can be distinguished around 5000-5500 cm^{-1} wavenumber region. The overlapping of spectral bands makes the use of multivariate calibration necessary to resolve the components from the full spectral data which is impossible with univariate calibration. The algorithm used in GILS can select the genes which are the selected wavelengths that have maximum correlation with the components.

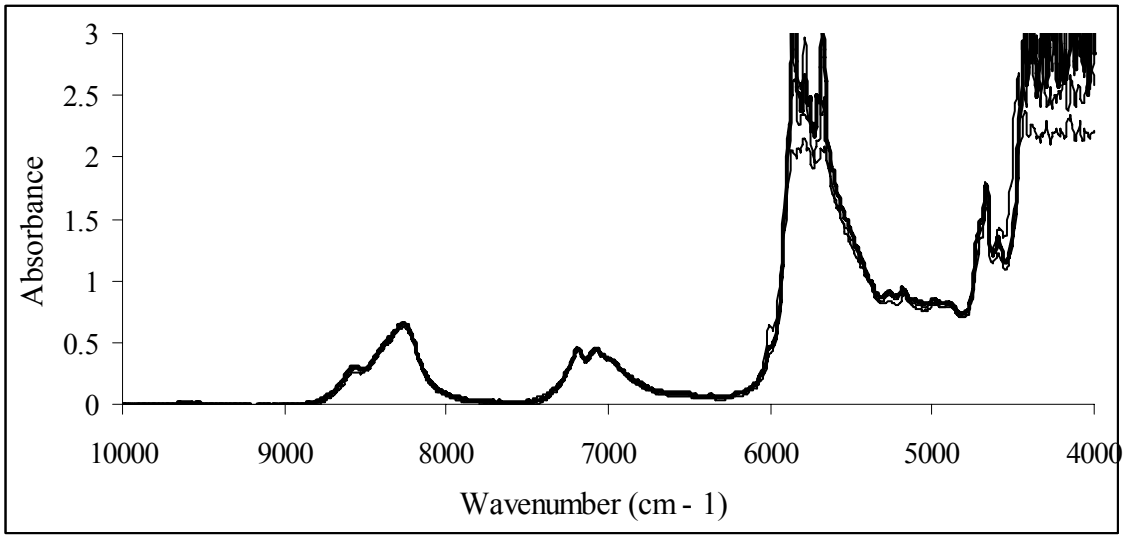

Figure 4.1. NIR spectra of pure sunflower oil, used frying oil, canola oil, and biodiesel

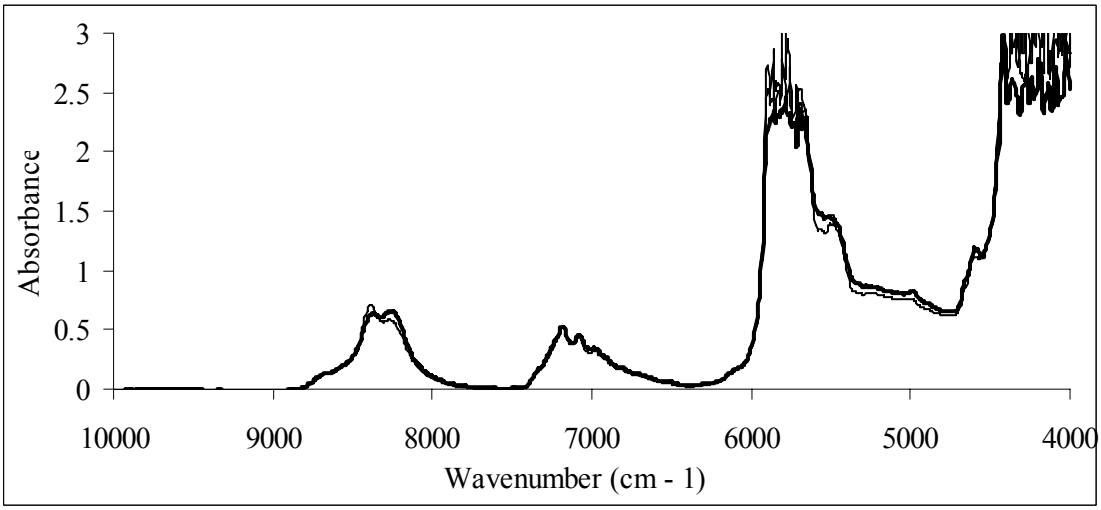

Figure 4.2. NIR spectra of kerosene, diesel, and euro diesel

The standart error of calibration (SEC) values of each component were found between 0.19% (w/w) and 2.31% (w/w) and the standart error of prediction (SEP) values of each component were found between 0.55% (w/w) and 1.98% (w/w) by using GILS method for all sets. Calibration models for sunflower oil determination gave SEC and SEP values as 0.21% (w/w) and 0.95% (w/w) and for diesel determination as 0.19% (w/w) and 0.68% (w/w) for the first data set, respectively. When these SEC and SEP values are examined, it is seen that these values are compatible with each other, which illustrates a good prediction for fast identification of a possible diesel adulteration with

sunflower oil. Figure 4.3 shows the actual sunflower and diesel concentration values versus their GILS predicted concentration values based on NIR spectra for the first data set. The R^2 values of regression lines for sunflower oil and diesel was found 0.999. When the overall calibration performance of the models is examined, it is possible to state that the NIR spectra contain quantitative information of sunflower oil and diesel.

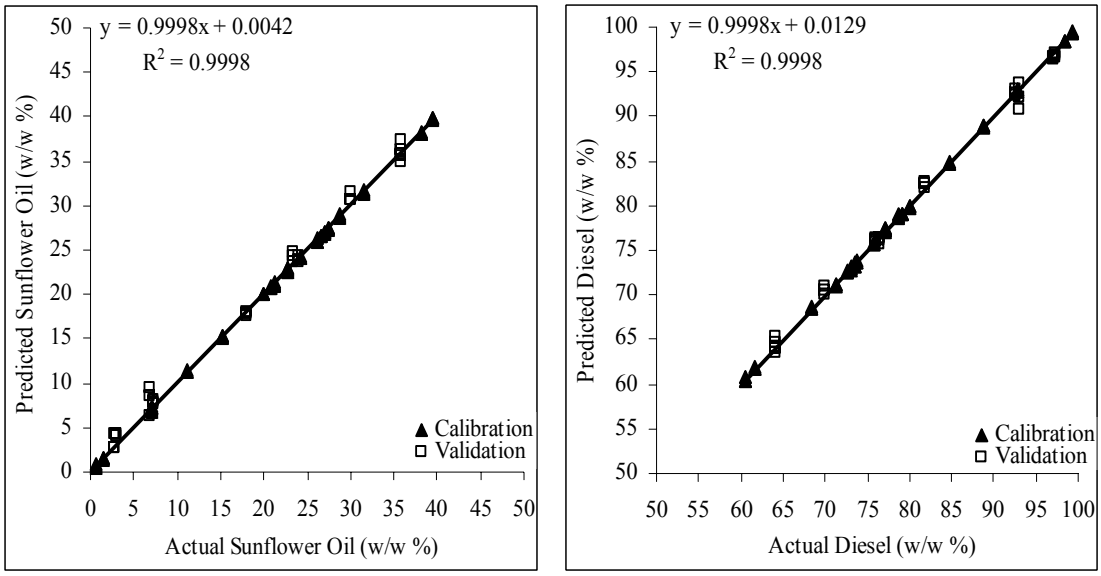

Figure 4.3. Actual vs predicted concentration plots obtained by GILS for the first set

Figure 4.4 shows the actual sunflower oil, used frying oil, and diesel concentration values versus their GILS predicted concentration values based on NIR spectra for the second data set. While the concentrations of sunflower oil were ranging between 0.65% and 39.56% (w/w) for the first data set, for the second data set the concentrations of sunflower oil were between 2.15% and 24.71% (w/w). On the other hand, the concentrations of diesel in the second data set were in a narrower mass range between 52.02% (w/w) and 93.79% (w/w) when compared with the first data set. The SEC values for sunflower oil, used frying oil, and diesel were found as 0.56, 0.77, and 0.40, while the SEP values were found 0.79, 0.92, and 0.55, respectively. The R^2 value of regression line for sunflower oil was 0.996, lower than this obtained for the first data set. Furthermore, the R^2 values of regression lines for used frying oil and diesel were 0.994 and 0.999. When the SEC and SEP values of sunflower oil and diesel for the second data set are examined, it is seen that the agreement between these values are better than those obtained for the first data set. In addition, there is also a good

agreement between the SEC and SEP values of used frying oil for the second data set. When the overall calibration performance of the models is examined, it can be said that the NIR spectra contain quantitative information of sunflower oil, used frying oil, and diesel.

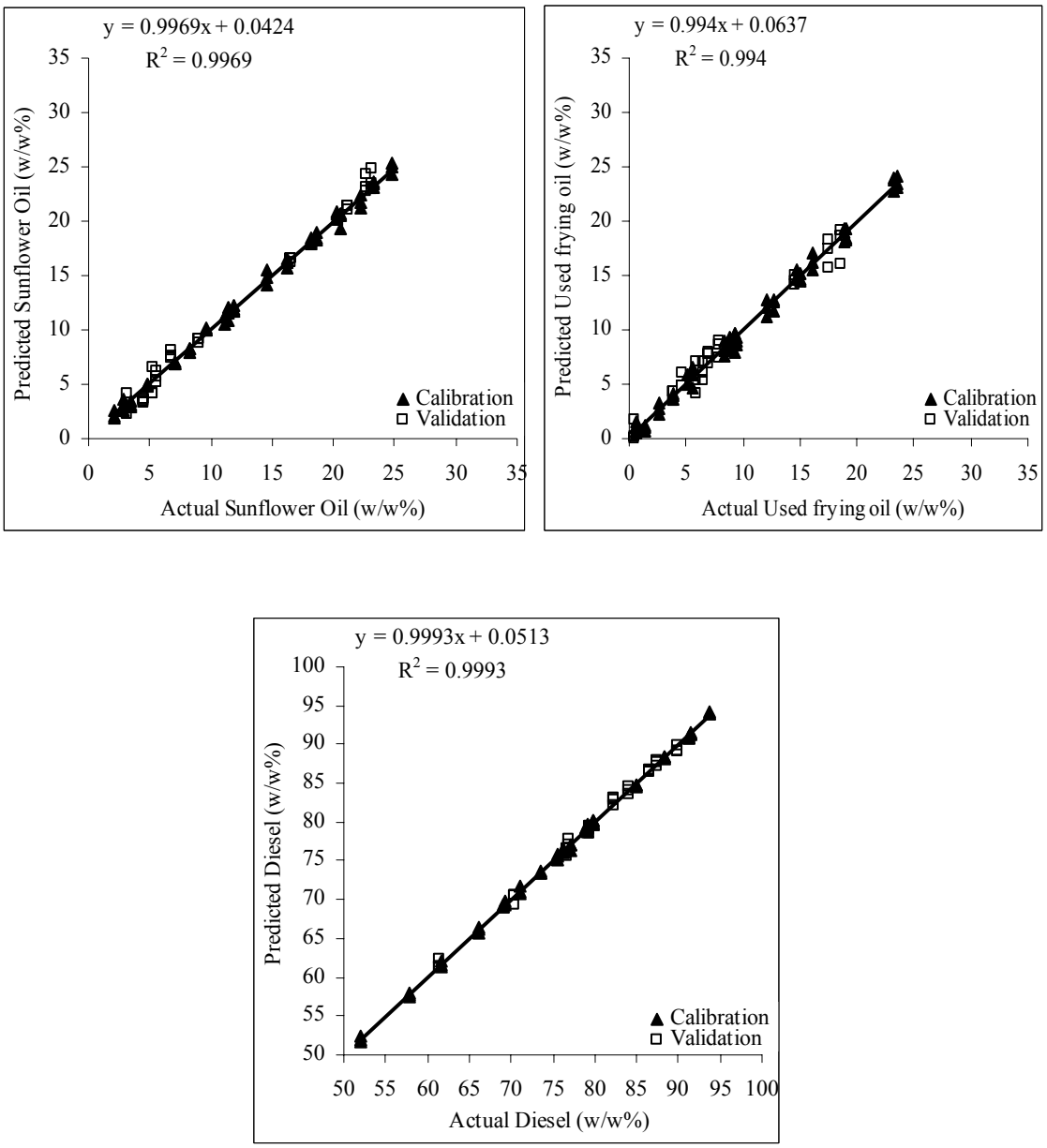

Figure 4.4. Actual vs. predicted concentration plots obtained by GILS for the second set

The calibration plots of kerosene, diesel and euro diesel for the third data set are given in Figure 4.5. The SEC and SEP values for the determination of kerosene content were obtained as 1.66% (w/w) and 1.81% (w/w) and those for the determination of

normal diesel were 2.31% (w/w) and 1.75% (w/w), respectively. In the case of euro diesel determination, the SEC and SEP results were 1.50% (w/w) and 1.98% (w/w), higher than the first and second data sets. On the other hand, R^2 value of euro diesel somewhat went down, while its calibration and prediction results increase. In addition, R^2 values of regression lines for kerosene and normal diesel were 0.993 and 0.995. Similar regression coefficients show that NIR spectra also contain quantitative information of kerosene, normal diesel, and euro diesel.

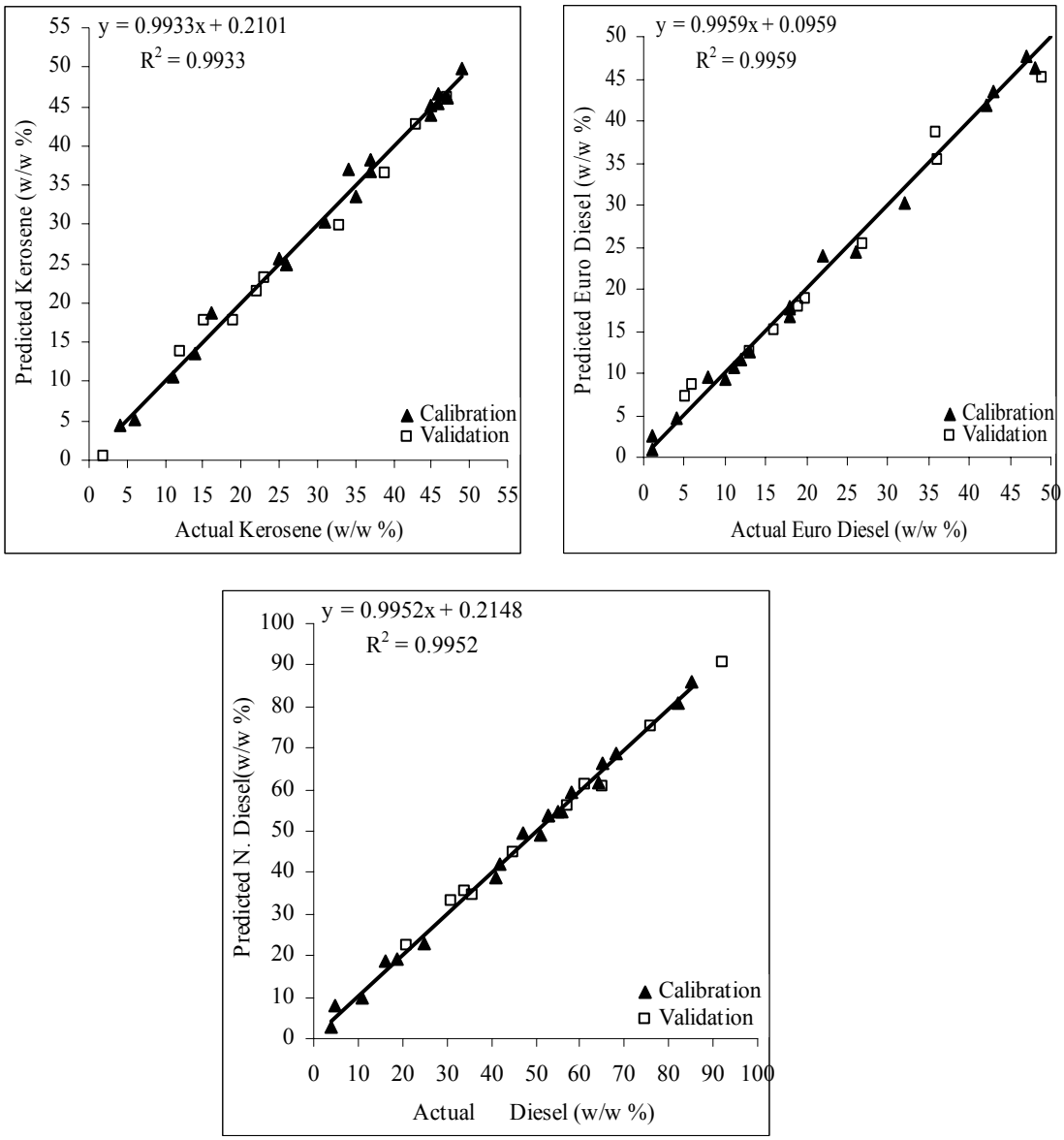

Figure 4.5. Actual vs predicted concentration plots obtained by GILS for the third set

Figure 4.6 shows the actual canola oil, biodiesel, and diesel concentration values versus their GILS predicted concentration values based on NIR spectra for the fourth data set. The R^2 values of regression lines for canola oil, biodiesel, and diesel were found 0.999. Calibration models for the fourth data set, for canola oil determination gave SEC and SEP values as 0.77% (w/w) and 0.73% (w/w), for biodiesel determination 0.69% (w/w) and 0.76% (w/w), and for diesel determination 0.56% (w/w) and 0.64% (w/w), respectively. When these SEC and SEP values are examined, it can be said that these values are compatible with each other, which demonstrates a good prediction for fast identification of a possible diesel adulteration with canola oil and biodiesel. In addition, the overall calibration performance of the models showed that the NIR spectra contain more quantitative information of canola oil, biodiesel, and diesel when compared with other sets.

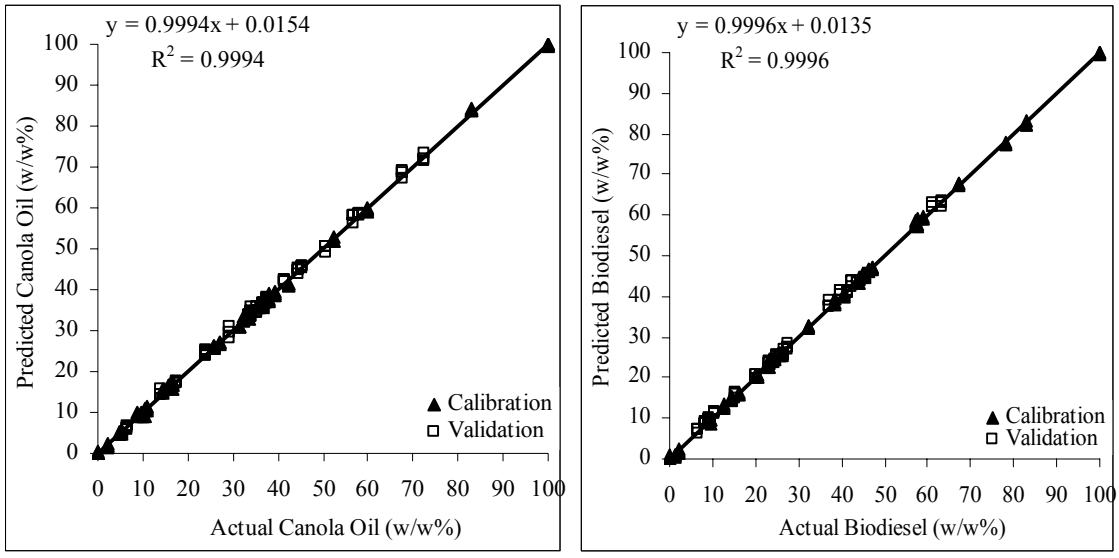

Figure 4.6. Actual vs. predicted concentration plots obtained by GILS for the fourth set

(Cont. on next page)

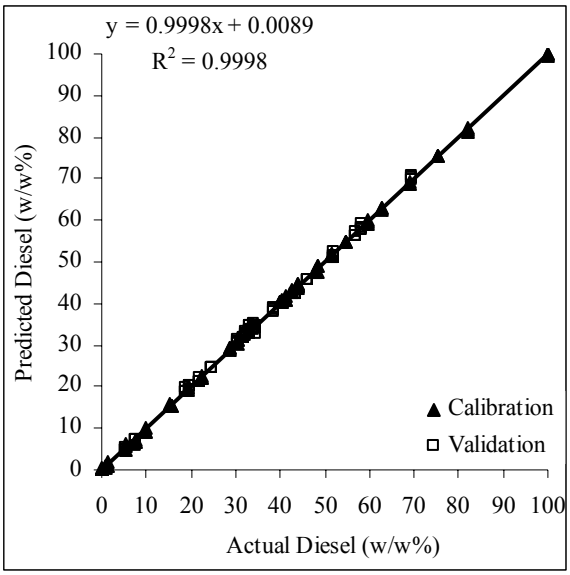

Figure 4.6(cont). Actual vs. predicted concentration plots obtained by GILS for the fourth set

Due to the fact that GILS is a wavelength selection method, the distribution of selected wavelengths in multiple runs over the whole spectrum that correspond to each component can be observed. The frequency distributions of selected wavenumbers in 50 runs with 20 genes and 50 iterations were plotted against wavenumber range for each component in Figure 4.7 to 4.10. As can be seen from the figures, the frequency of the selected wavenumbers is significantly higher around the peak maximum of each component. This indicates that the GILS method selects the wavenumbers, where the most concentration related information is contained. As a result of this, it can be said that the GILS method can be used for the determination of diesel adulteration with sunflower oil, used frying oil, canola oil, kerosene, and biodiesel.

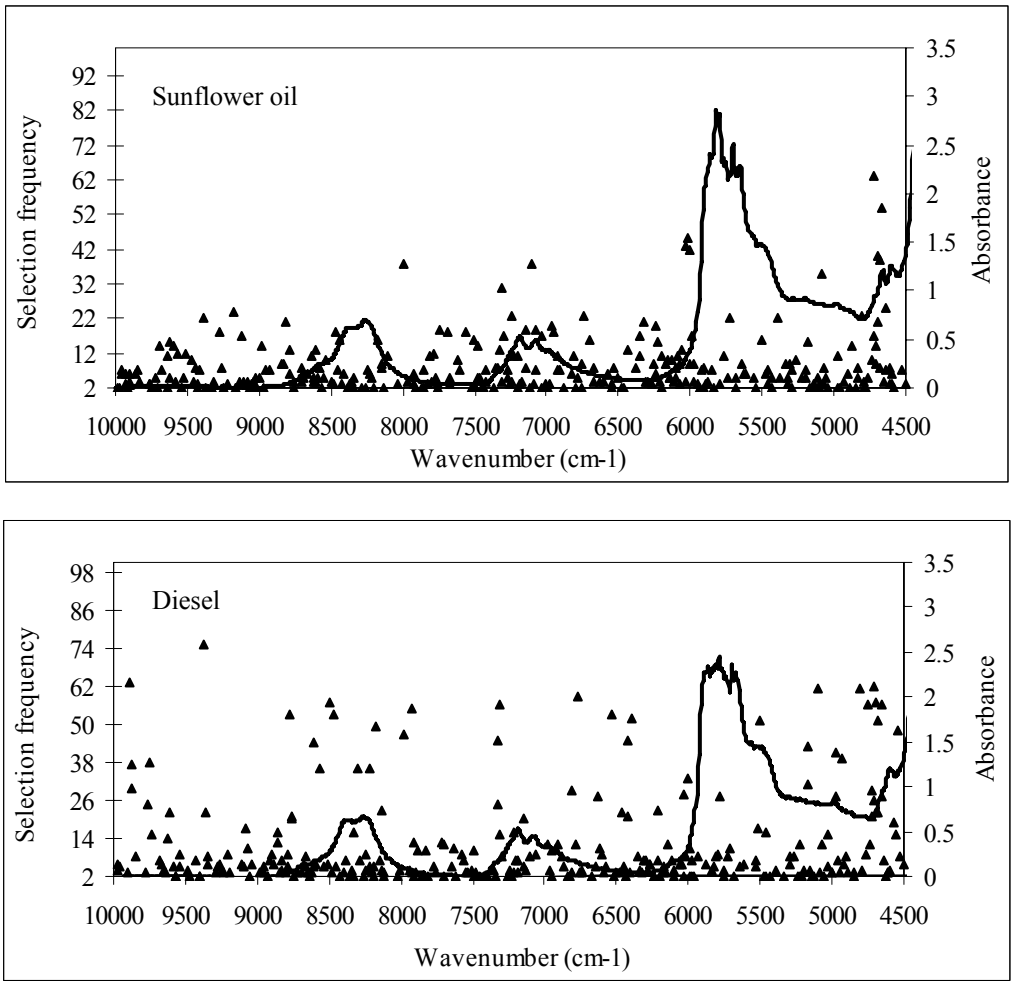

Figure 4.7. Frequency distribution of GILS selected wavelengths using NIR data of first set for sunflower oil and diesel along with its pure component spectrum

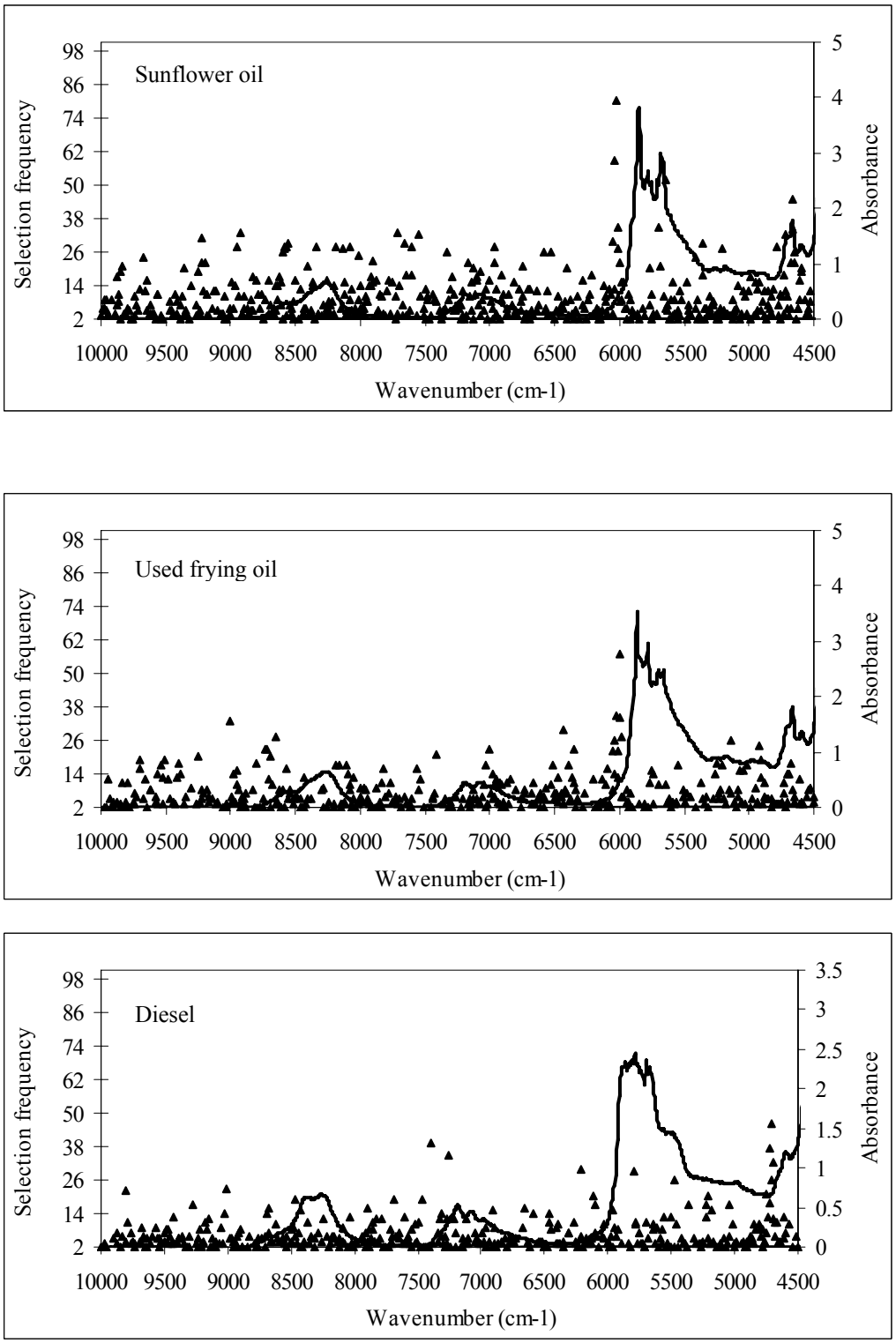

Figure 4.8. Frequency distribution of GILS selected wavelengths using NIR data of second set for sunflower oil, used frying oil, and diesel along with its pure component spectrum

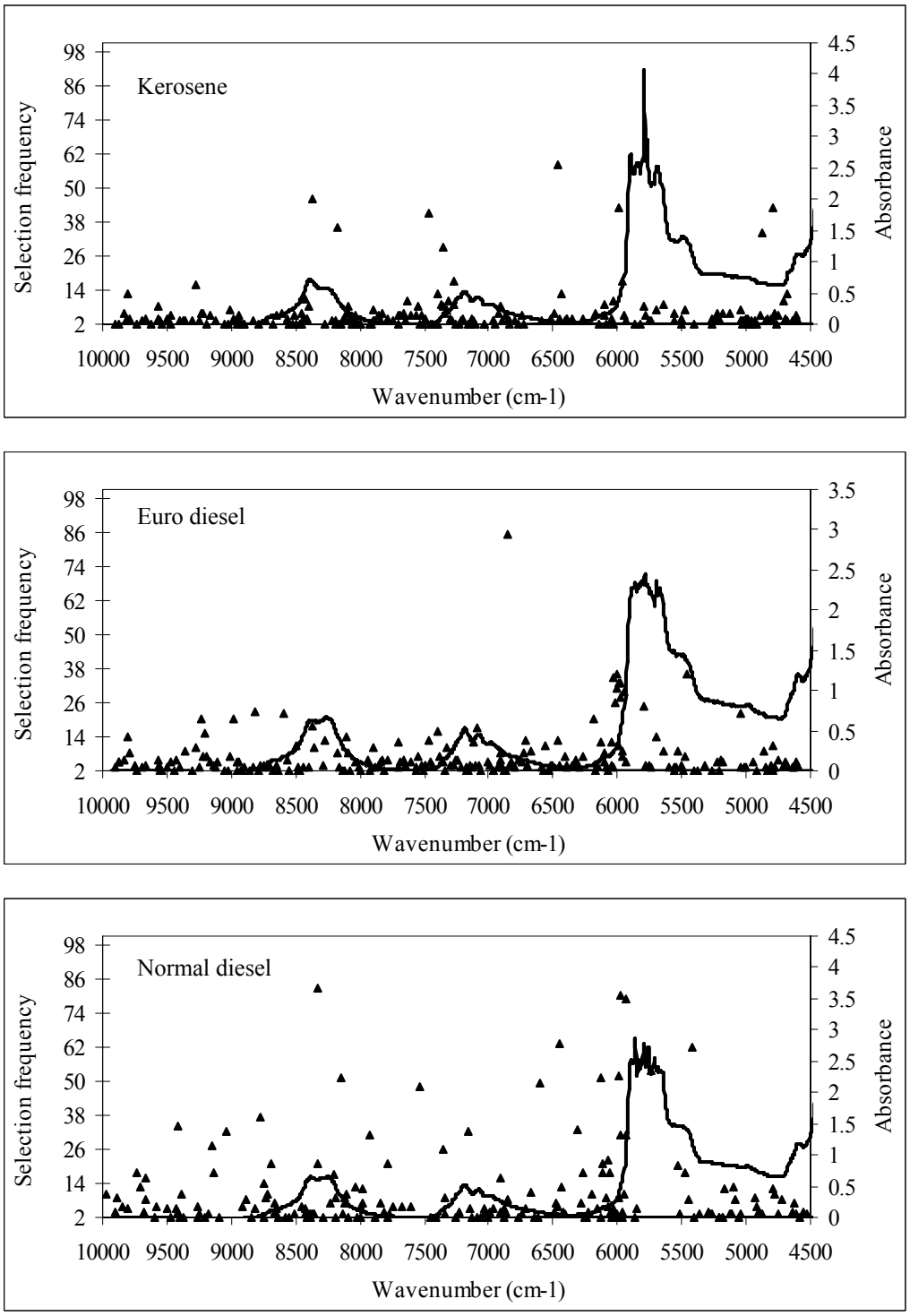

Figure 4.9. Frequency distribution of GILS selected wavelengths using NIR data of third set for kerosene, euro diesel, diesel along with its pure component spectrum

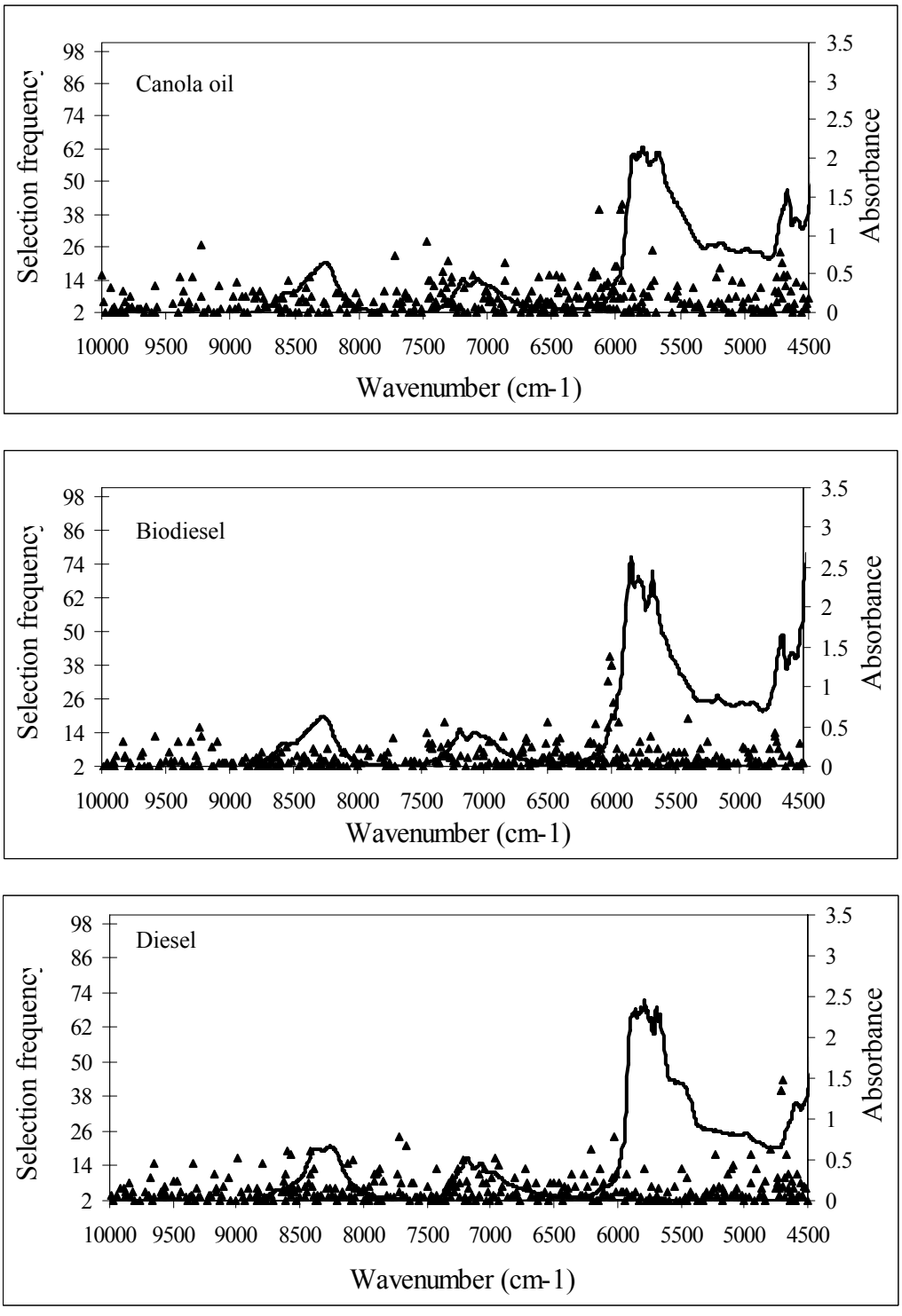

Figure 4.10. Frequency distribution of GILS selected wavelengths using NIR data of fourth set for canola oil, biodiesel, and diesel along with its pure component spectrum

CHAPTER 5

DETERMINATION OF DIESEL ADULTERATION BY FTIR-ATR SPECTROSCOPY COMBINED TO MULTIVARIATE CALIBRATION

In literature several different methods have been employed for determining the amount of biodiesel in petroleum diesel fuel in the presence of vegetable oils which are not converted into biodiesel. Their determination with classic methods including ^{1}H NMR spectroscopy (Knothe 2001) and chromatography (Foglia, et al. 2005) are expensive and time consuming as they require sample preparation. Infrared spectroscopy combined to multivariate calibration has been shown to be an alternative analytical technique to classic methods since it allows low cost, fast, and non-destructive determination without sample preparation (Pimentel, et al. 2006). Among IR measurements, fourier transform infrared (FTIR) spectroscopy has found several applications in the determination of biodiesel content in diesel. For example, Aliske et al. (2007) reported that FTIR spectroscopy is the most promising method in the determination of biodiesel-diesel mixtures concentrations, since in the mid infrared range there are many unmatched peaks between biodiesel and diesel. On the other hand, the literature is very scarce in the use of fourier transform infrared-attenuated total reflectance (FTIR-ATR) spectroscopy combined to multivariate calibration analysis as an analytical tool. Oliveira et al. (2007) has determined the biodiesel contents in petroleum diesel by using FTIR-ATR spectroscopy combined to multivariate calibration methods.

In this study we used the FTIR-ATR spectroscopy combined to multivariate calibration to determine the adulteration of diesel with biodiesel and its corresponding vegetable oil present in diesel.

5.1. Instrumentation and Data Pre-Processing

A Spectrum 100 FTIR spectrometer (Perkin Elmer, Waltham, MA) was employed to obtain the FTIR-ATR spectra between 450 cm^{-1} and 4,000 cm^{-1} wavenumber range. The resolution was set to 4 cm^{-1} and 4 scans were done using air as backround. This spectrometer was equipped with a tungsten lamp as source, extended range KBr as beam splitter, and temperature stabilized fast recovery deuterated triglycine sulfate (FR-DTS) as detector. Also triple measurements were taken by using an ATR accessory (Pike Technologies, ATR diamond KRS 5 Accessory). Then the means of the measurements were used in multivariate analyses. All spectra collected from the instrument were transferred to a computer and the genetic inverse least squares (GILS) was employed to set up the calibration models. Microsoft Excel (MS Office 2003, Microsoft Corporation) program was used to prepare the text files for calibration and validation sets, which are required to build the calibration models and test the GILS method which was written in MATLAB programming language using Matlab 5.3 (MathWorks Inc., Natick, MA). The samples in calibration and validation sets were selected randomly. The unit of concentration values was taken as mass percent.

5.2. Sample Preparation and Design of the Data Sets

Canola oil was purchased from a local supermarket and biodiesel which is produced from canola oil was obtained from a company named Ege Biotechnology Ltd. (Izmir Institute of Technology, Technopark, TR). Diesel was bought from a service station in Izmir.

A set including 50 ternary mixtures of canola oil, biodiesel, and diesel were prepared. The concentrations of canola oil, biodiesel, and diesel were in the mass range between 0% and 100% in the mixture. From the 50 samples, 30 samples were used in the calibration set and the other 20 samples were separated for validation set. Table 5.1 and Table 5.2 illustrate the concentrations of each component for the calibration and validation sets. All concentrations on the tables are given in mass percentage.

Table 5.1. Concentration profile of the calibration samples

Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)	Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)
1	100	0.00	0.00	16	9.90	14.61	75.47
2	0.00	100	0.00	17	10.33	67.17	22.49
3	0.00	0.00	100	18	25.97	26.16	47.86
4	37.20	22.88	39.91	19	8.89	47.05	44.05
5	32.43	57.78	9.78	20	39.40	20.30	40.30
6	31.43	25.77	42.79	21	35.75	22.96	41.28
7	5.13	12.83	82.03	22	9.67	38.62	51.70
8	59.71	9.72	30.55	23	36.46	44.05	19.48
9	38.13	2.38	59.48	24	16.61	78.17	5.20
10	11.02	44.99	43.98	25	27.08	24.30	48.61
11	15.74	83.05	1.20	26	33.46	59.11	7.41
12	4.66	40.42	54.91	27	25.94	45.30	28.74
13	16.68	20.36	62.94	28	52.24	32.36	15.38
14	14.57	16.33	69.09	29	2.10	46.28	51.60
15	42.19	57.00	0.79	30	83.11	1.09	15.78

Table 5.2. Concentration profile of the validation samples

Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)	Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)
1	23.84	6.53	69.61	11	57.92	20.12	21.95
2	56.88	9.71	33.40	12	37.40	10.59	52.00
3	17.34	24.30	58.35	13	34.06	9.01	56.91
4	33.57	60.93	5.49	14	67.61	24.78	7.60
5	44.36	23.23	32.39	15	45.55	8.09	46.35
6	72.28	9.03	18.68	16	14.16	42.49	43.33
7	36.62	43.81	19.55	17	50.60	15.20	34.20
8	6.48	63.21	30.30	18	35.42	39.74	24.82
9	41.45	26.49	32.04	19	29.45	37.20	33.33
10	34.09	27.46	38.43	20	24.12	41.44	34.43

5.3. Results and Discussion

The genetic inverse least squares (GILS) multivariate calibration method combined to FTIR-ATR spectroscopy have been used to determine diesel adulteration with canola oil and biodiesel. Figure 5.1 demonstrates the FTIR-ATR spectra of pure canola oil, biodiesel, and diesel between 450 and 4,000 cm^{-1} wavelength range. Because the spectral characteristics of canola oil and its methyl ester (biodiesel) are much alike, they exhibit overlapped signals. The spectral bands of canola oil, biodiesel, and diesel are only discriminable approximately between 800 and 1200 cm^{-1} wavelength range as seen from the Figure 5.1

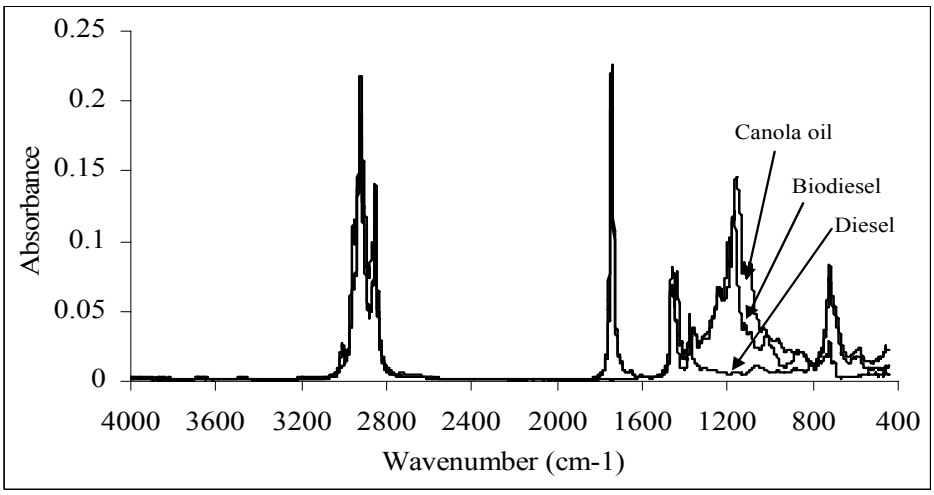

Figure 5.1. FTIR-ATR spectra of pure canola oil, biodiesel, and diesel

Figure 5.2 shows the actual canola oil, biodiesel, and diesel concentration values versus their GILS predicted concentration values based on FTIR-ATR spectra. The SEC values for canola oil, biodiesel, and diesel were found as 0.23% (w/w), 0.11% (w/w), and 0.14% (w/w), while the SEP values were found 0.49% (w/w), 0.37% (w/w), and 0.33% (w/w), respectively. When these SEC and SEP values, which are obtained by GILS using FTIR-ATR spectra are examined, it is seen that the agreement between these values are better than those obtained by GILS using NIR spectra. In addition, the R^2 results of regression lines for canola oil, biodiesel, and diesel were higher than those acquired by NIR. When the overall calibration performance of the FTIR-ATR based

models is examined, it is possible to state that the FTIR-ATR spectra contain more quantitative information of canola oil, biodiesel, and diesel than those contain NIR spectra.

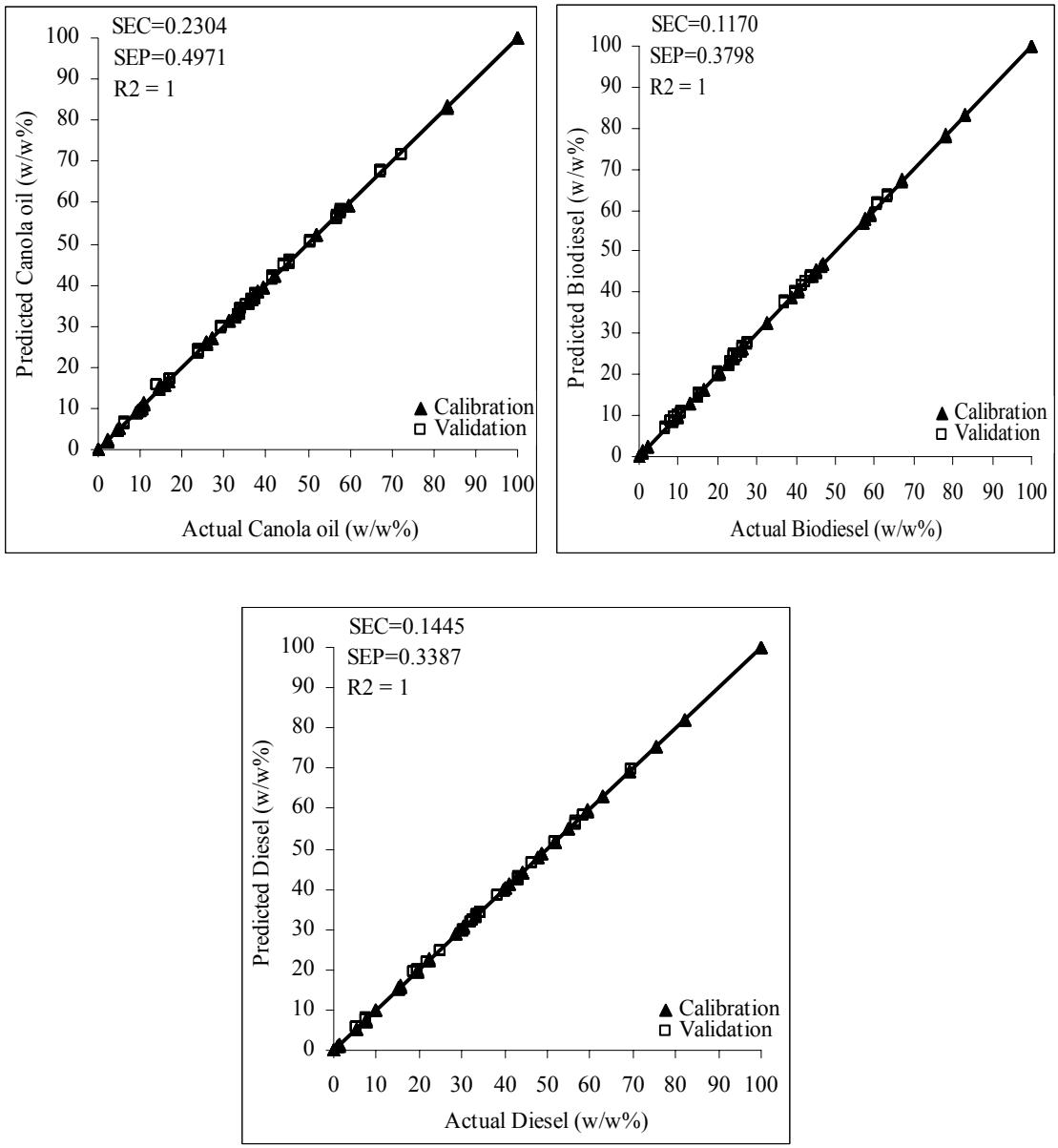

Figure 5.2. Actual vs. predicted concentrations plots obtained by GILS

The frequency distributions of selected wavenumbers in 50 runs with 20 genes and 50 iterations were plotted against wavenumber range for each component in Figure 5.3. As seen from the figure, the frequency of the selected wavenumbers is significantly higher around the maximum peak of each component. This shows that the GILS method

selects the wavenumbers, where only the information related to the particular component is used to construct the model so that the noise in the whole spectrum reduces. As a result, it is possible to state that the GILS method can be used for fast and simultaneous determination of diesel adulteration with canola oil and biodiesel.

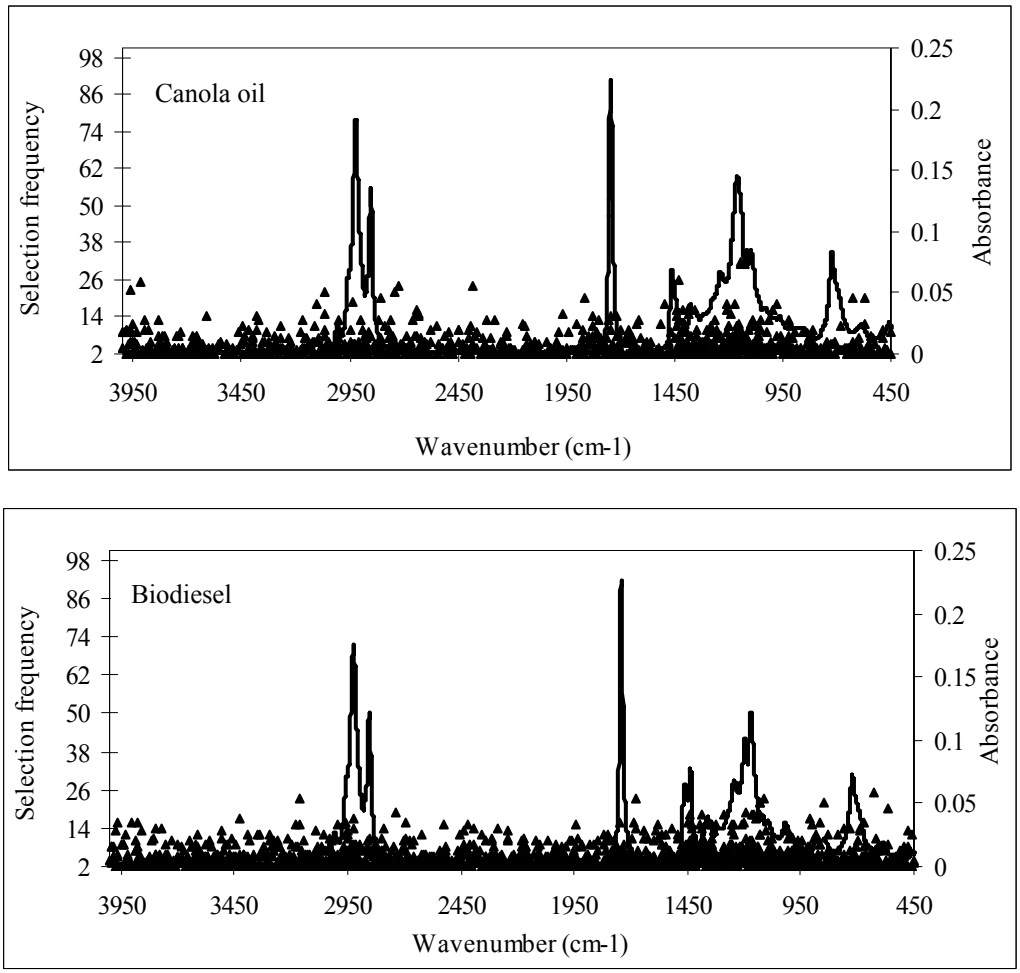

Figure 5.3. Frequency distribution of GILS selected wavelengths using FTIR-ATR data for canola oil, biodiesel, and diesel along with its pure component spectrum

(Cont. on next page)

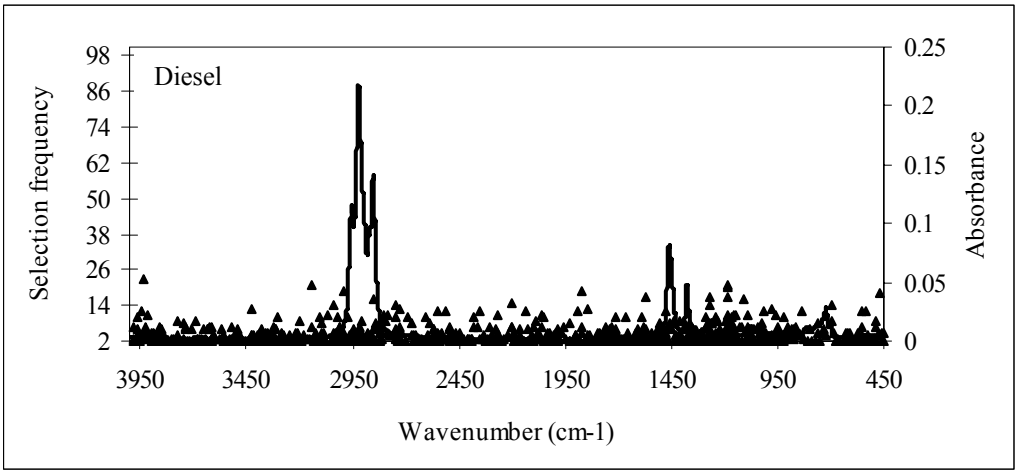

Figure 5.3 (cont). Frequency distribution of GILS selected wavelengths using FTIR-ATR data for canola oil, biodiesel, and diesel along with its pure component spectrum

CHAPTER 6

DETERMINATION OF DIESEL ADULTERATIONS BY MOLECULAR FLUORESCENCE SPECTROSCOPY COMBINED TO MULTIVARIATE CALIBRATION

Fluorescence spectroscopy combined to multivariate analysis which is a fast, sensitive, and inexpensive tool, finds many applications for the analysis of petroleum products (Patra and Mishra 2002, Dudelzak, et al. 1991, Shanahan, et al. 1991, Patra, et al. 2001). Because standart techniques like GC–MS or testing of physical properties are time consuming, costly and often complicated in determining the adulteration level, fluorescence based methods are offered as rapid and inexpensive techniques for the analysis of petroleum products. For example, Patra and Mishra (2001) used EEMF spectroscopy to estimate the amount of kerosene present in diesel. However, it was not found in the literature a fluorescence spectroscopy based method capable of determining diesel adulterations with vegetable oils and biodiesel. In this sence, this study is important to be the first study in the determination of diesel adulterations with vegetable oils, used frying oil, and biodiesel by using molecular fluorescence spectroscopy combined to multivariate calibration.

6.1. Instrumentation and Data Pre-Processing

Fluorescence measurements were carried out with Varian Cary Elipse spectrofluorimeter with a 100 W Xenon lamp as source. For two sets emission spectra were collected. The emission fluorescence spectra of both sets were recorded in the excitation wavelengths between 300 and 400 nm and emission wavelength range between 300 and 650 nm within an interval of 10 nm respectively. Excitation slit width was chosen 5 and emission slit width was chosen 2.5 nm. For the second set additionally the synchronous fluorescence spectra were collected by simultaneously scanning both the excitation and emission monochromators for an excitation wavelength range 300 to 700 nm and $\Delta\lambda$ was chosen 10 nm. Excitation and emission slit widths were chosen 5.

Preprocessing of the data obtained from instrument were performed with Microsoft Excel (MS Office 2003, Microsoft Corporation) program using genetic inverse least squares (GILS) method which was written in MATLAB programming language using Matlab 5.3 (MathWorks Inc., Natick, MA).

6.2. Sample Preparation and Design of the Data Sets

Two sets including ternary mixtures of diesel with sunflower oil, used frying oil, canola oil, and biodiesel were prepared. Sunflower and canola oils were purchased from a local supermarket. Used frying oil was acquired from a restaurant. Biodiesel was obtained from a company named Ege Biotechnology Ltd. (Izmir Institute of Technology, Technopark, TR) and diesel was purchased from a service station in Izmir. The concentration data corresponding to each component of each set were divided into two sets which are called calibration and validation (prediction) sets. Table 6.1 to Table 6.4 illustrate the concentrations of each component for the calibration and validation sets. The samples in calibration and validation sets were selected randomly. All concentration values on the tables are given in mass percentage. The first set contains 30 ternary mixtures of sunflower oil, used frying oil, and diesel. The calibration set composed of 20 samples, and the validation set composed of 10 samples. The concentrations of sunflower oil were in the mass range between 2.15% and 24.71%, used frying oil were in the mass range between 0.54% and 23.56%, and diesel were in the mass range between 52.02% and 93.79%.

Table 6.1. Concentration profile of the calibration samples in the first set

Sample No	Sunflower oil (w/w%)	Used frying oil (w/w%)	Diesel (w/w%)	Sample No	Sunflower oil (w/w%)	Used frying oil (w/w%)	Diesel (w/w%)
1	2.89	12.09	85.00	11	7.05	23.56	69.38
2	14.60	5.60	79.80	12	18.15	8.30	73.55
3	16.24	3.94	79.81	13	11.90	19.00	69.08
4	18.70	15.05	66.25	14	24.71	23.26	52.02
5	11.15	12.65	76.18	15	4.80	1.40	93.79
6	20.32	2.64	77.02	16	8.25	0.55	91.20
7	20.21	8.75	71.03	17	20.53	0.54	78.91
8	9.64	14.74	75.61	18	23.28	18.84	57.87
9	3.49	5.09	91.40	19	2.15	9.40	88.44
10	22.25	16.15	61.60	20	11.45	9.30	79.23

Table 6.2. Concentration profile of the validation samples in the first set

Sample No	Sunflower oil (w/w%)	Used frying oil (w/w%)	Diesel (w/w%)	Sample No	Sunflower oil (w/w%)	Used frying oil (w/w%)	Diesel (w/w%)
1	21.18	17.44	61.36	6	5.30	4.70	90.00
2	9.05	6.90	84.03	7	16.59	3.94	79.46
3	3.15	14.55	82.30	8	4.55	18.55	76.88
4	23.08	6.49	70.41	9	5.60	7.85	86.54
5	22.73	0.49	76.76	10	6.74	5.85	87.39

The second set consists of 50 ternary mixtures of canola oil, biodiesel, and diesel (euro). The concentrations of canola oil, biodiesel, and diesel were in the mass range between 0% and 100%.

Table 6.3. Concentration profile of the calibration samples in the second set

Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)	Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)
1	100	0.00	0.00	16	9.90	14.61	75.47
2	0.00	100	0.00	17	10.33	67.17	22.49
3	0.00	0.00	100	18	25.97	26.16	47.86
4	37.20	22.88	39.91	19	8.89	47.05	44.05
5	32.43	57.78	9.78	20	39.40	20.30	40.30
6	31.43	25.77	42.79	21	35.75	22.96	41.28
7	5.13	12.83	82.03	22	9.67	38.62	51.70
8	59.71	9.72	30.55	23	36.46	44.05	19.48
9	38.13	2.38	59.48	24	16.61	78.17	5.20
10	11.02	44.99	43.98	25	27.08	24.30	48.61
11	15.74	83.05	1.20	26	33.46	59.11	7.41
12	4.66	40.42	54.91	27	25.94	45.30	28.74
13	16.68	20.36	62.94	28	52.24	32.36	15.38
14	14.57	16.33	69.09	29	2.10	46.28	51.60
15	42.19	57.00	0.79	30	83.11	1.09	15.78

Table 6.4. Concentration profile of the validation samples in the second set

Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)	Sample No	Canola oil (w/w%)	Biodiesel (w/w%)	Diesel (w/w%)
1	23.84	6.53	69.61	11	57.92	20.12	21.95
2	56.88	9.71	33.40	12	37.40	10.59	52.00
3	17.34	24.30	58.35	13	34.06	9.01	56.91
4	33.57	60.93	5.49	14	67.61	24.78	7.60
5	44.36	23.23	32.39	15	45.55	8.09	46.35
6	72.28	9.03	18.68	16	14.16	42.49	43.33
7	36.62	43.81	19.55	17	50.60	15.20	34.20
8	6.48	63.21	30.30	18	35.42	39.74	24.82
9	41.45	26.49	32.04	19	29.45	37.20	33.33
10	34.09	27.46	38.43	20	24.12	41.44	34.43

6.3. Results and Discussion

Molecular fluorescence spectroscopy combined to genetic inverse least squares (GILS) multivariate calibration method was used for the determination of diesel adulteration. Figure 6.1 and Figure 6.2 illustrate the excitation-emission and synchronous fluorescence spectra of pure components and their mixtures. According to the Figure 6.1, it is evident that excitation-emission spectral bands of canola oil, biodiesel, and diesel can be distinguished around 350 nm wavelength region. In addition, as it can be seen from the Figure 6.2, while the synchronous fluorescence spectral bands of canola oil, biodiesel, and diesel were overlapped around 350-500 nm wavelength region, the spectral bands of canola oil and diesel were overlapped around 600-700 nm wavelength region.

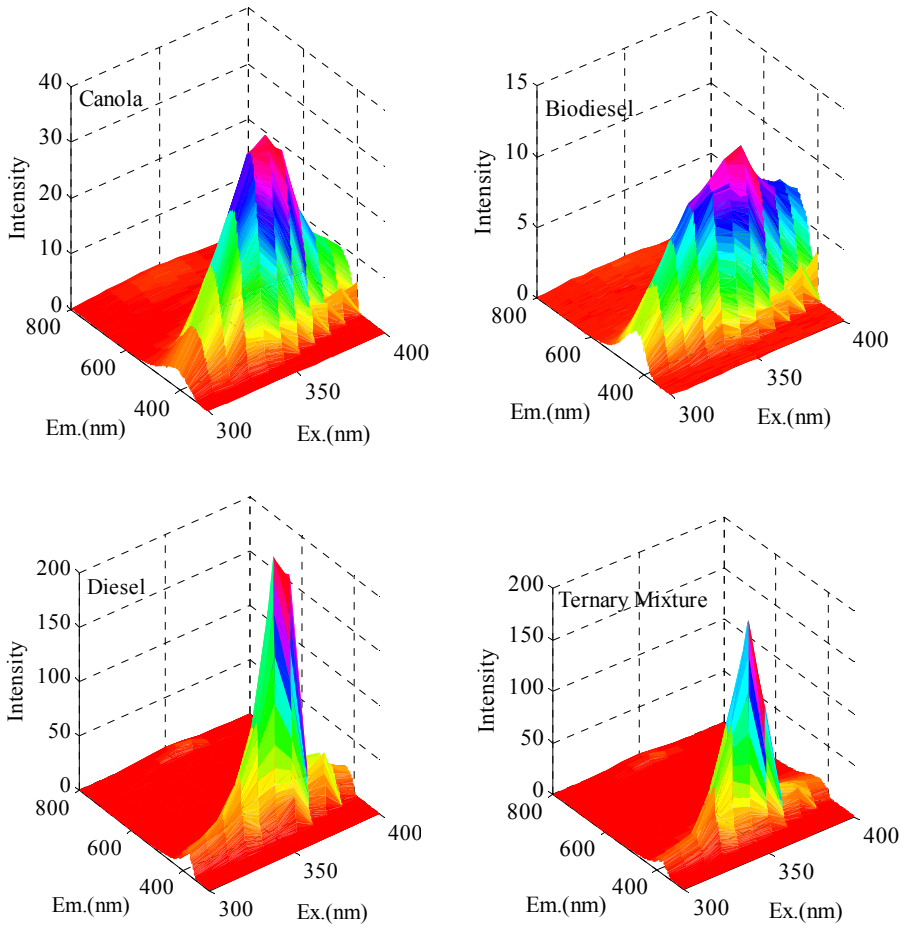

Figure 6.1. Excitation-emission fluorescence spectra of pure canola oil, biodiesel, and diesel along with their ternary mixture

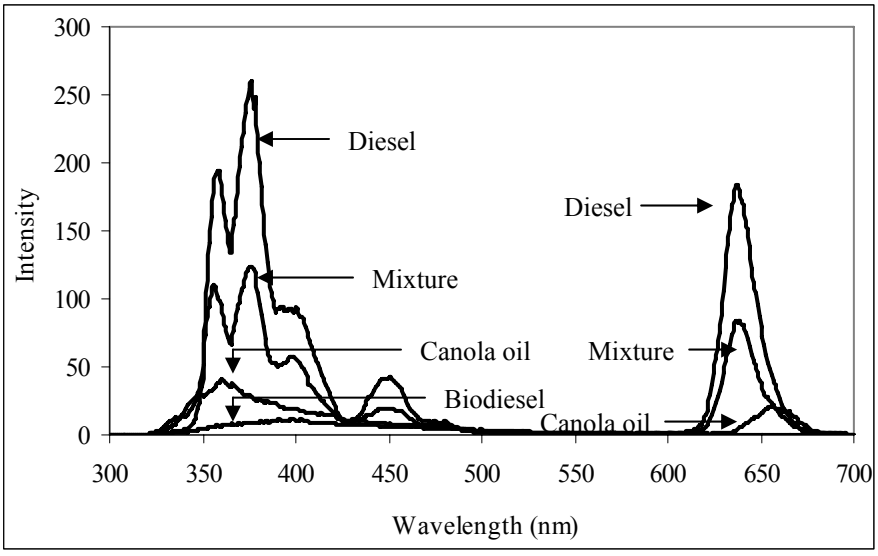

Figure 6.2. Synchronous fluorescence spectra of pure canola oil, biodiesel, and diesel along with their ternary mixture

For the excitation-emission fluorescence measurements the standart error of calibration (SEC) values were found between 0.92% (w/w) and 2.89% (w/w), and the standart error of prediction (SEP) values were found between 0.86% (w/w) and 3.87% (w/w) by using GILS. On the other hand, for the synchronous fluorescence measurements the SEC values were found between 0.83% (w/w) and 1.38% (w/w), and the SEP values were found between 1.46% (w/w) and 2.66% (w/w) by using GILS. Calibration models for sunflower oil determination gave SEC and SEP values as 1.88% (w/w) and 1.57% (w/w), for used frying oil determination 0.92% (w/w) and 0.86% (w/w), and for diesel determination 1.02% (w/w) and 2.56% (w/w) for the first data set, respectively.

When these SEC and SEP values are examined, it is seen that these values are compatible with each other, which illustrates a good prediction for fast identification by excitation-emission fluorescence for a possible diesel adulteration with sunflower oil and used frying oil. Figure 6.3 shows the actual sunflower oil, used frying oil, and diesel concentration values versus their GILS predicted concentration values based on excitation-emission fluorescence spectra for the first data set.

The R^2 values of regression lines for sunflower oil, used frying oil, and diesel was found as 0.964, 0.991, and 0.995. When the overall calibration performance of the models is examined, it is possible to state that the excitation-emission fluorescence spectra contain quantitative information of sunflower oil, used frying oil, and diesel.

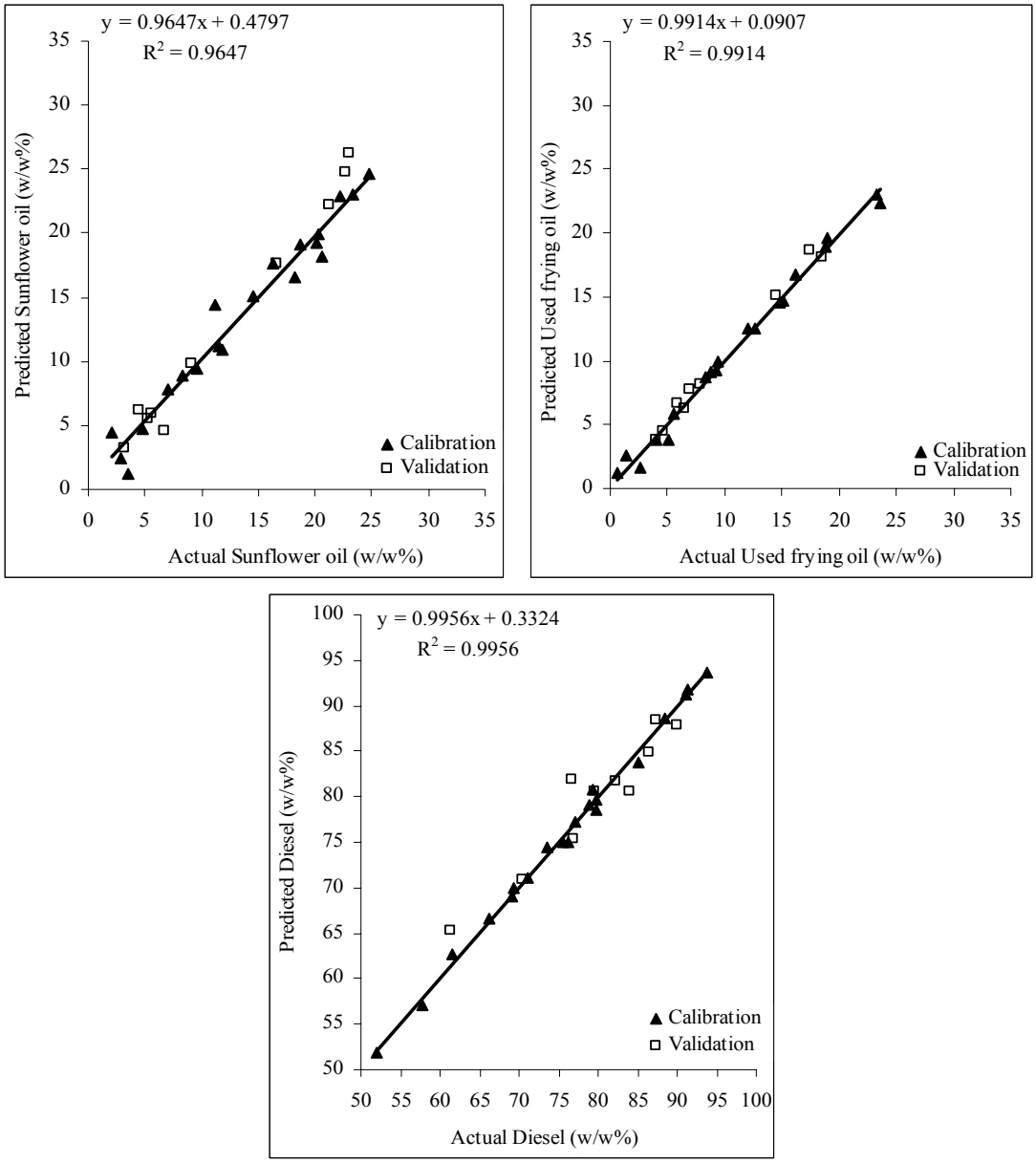

Figure 6.3. Actual vs. predicted concentration plots of sunflower oil, used frying oil, and diesel by GILS using excitation-emission fluorescence data

Figure 6.4 shows the actual canola oil, biodiesel, and diesel concentration values versus their GILS predicted concentration values based on excitation-emission fluorescence spectra. The SEC values for canola oil, biodiesel, and diesel were found as 1.56, 2.89, and 1.72, while the SEP values were found as 2.34% (w/w), 3.87% (w/w), and 2.49% (w/w), respectively. When these SEC and SEP values are examined, it is seen that the agreement between these values are worse than those obtained for

sunflower oil, used frying oil, and diesel in the first data set. The R^2 value of regression line for diesel higher than this obtained for the first data set.

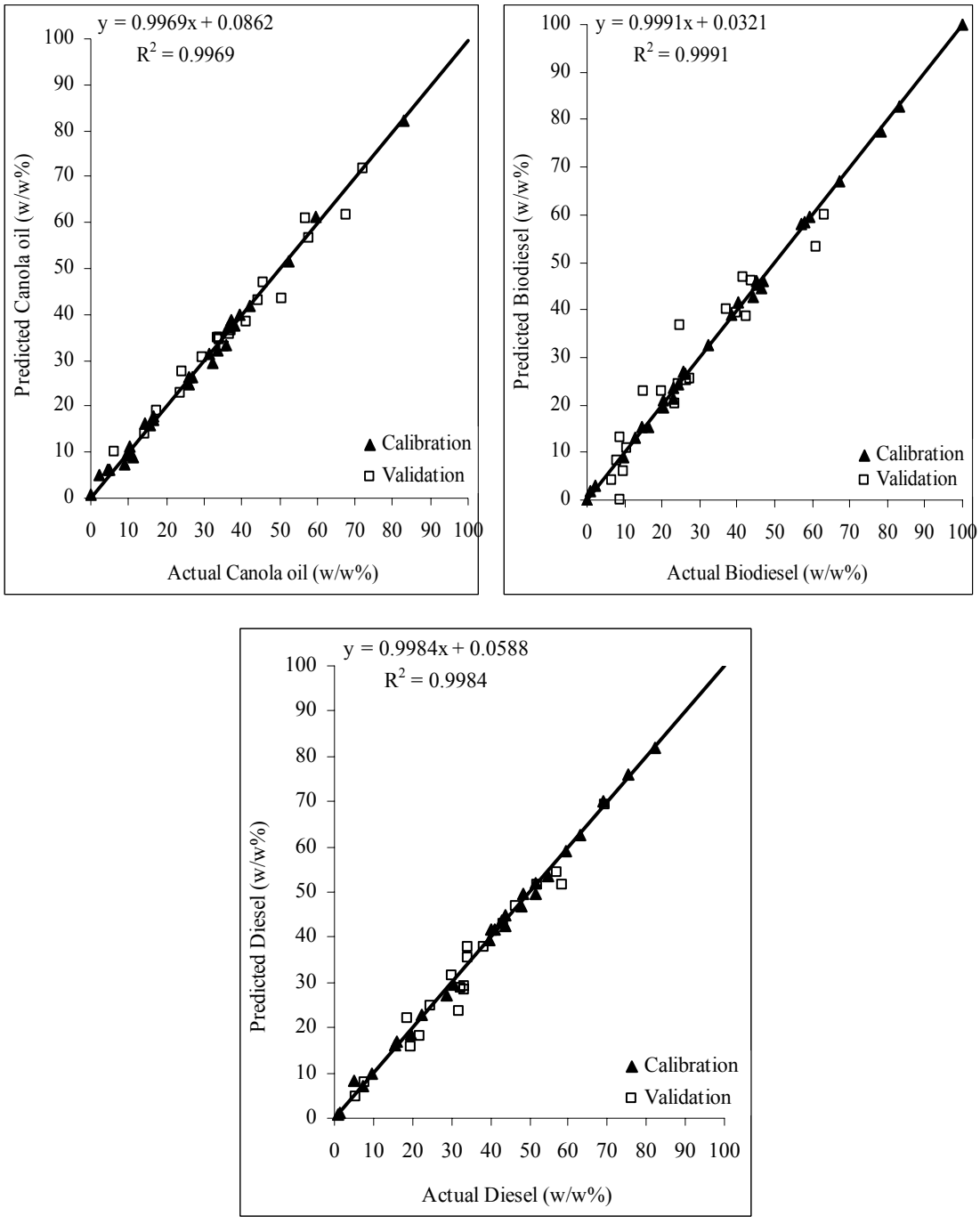

Figure 6.4. Actual vs. predicted concentration plots of canola oil, biodiesel, and diesel by GILS using excitation-emission fluorescence data

In addition, Figure 6.5 shows the actual canola oil, biodiesel, and diesel concentration values versus their GILS predicted concentration values for second set based on syncronous fluorescence spectra. In the case of canola oil, biodiesel, and diesel determination, while the SEC results were found as 1.38% (w/w), 1.10% (w/w), and 0.83% (w/w), the SEP results were found as 2.30% (w/w), 2.66% (w/w), and 1.46% (w/w), respectively. When these SEC and SEP values are examined, it is seen that the agreement between these values are better than those obtained by GILS using excitation-emission fluorescence spectra. On the other hand, R^2 values of regression lines of diesel were somewhat higher. Similar regression coefficients show that syncronous fluorescence spectra also contain quantitative information of canola oil, biodiesel, and diesel.

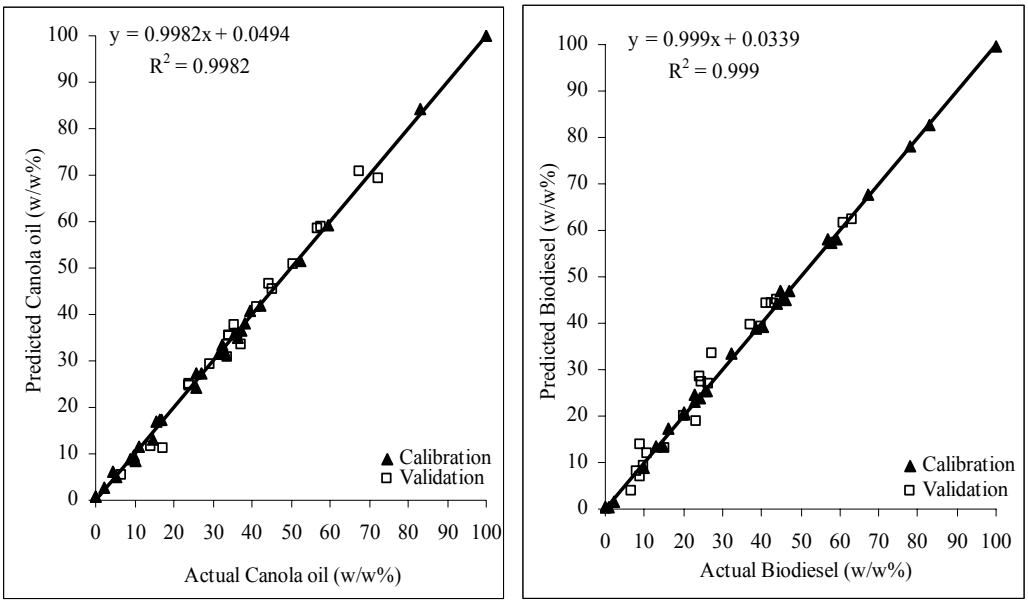

Figure 6.5. Actual vs. predicted concentration plots of canola, biodiesel, and diesel by GILS using synchronous fluorescence data

(cont. on next page)

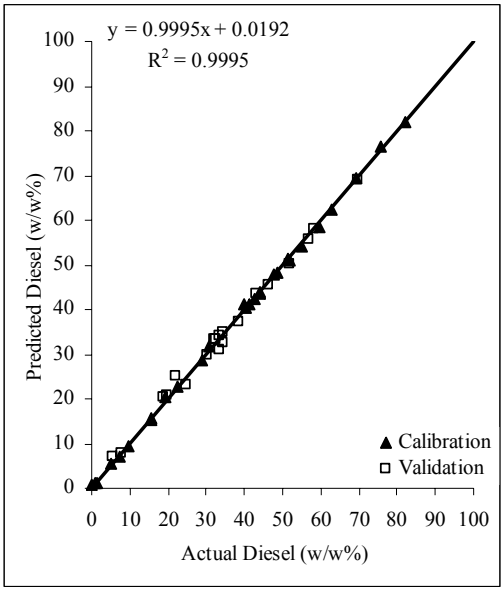

Figure 6.5 (cont). Actual vs. predicted concentration plots of canola, biodiesel, and diesel by GILS using synchronous fluorescence data

In order to construct calibration models, the GILS program was set to run 50 times with 20 genes and 50 iterations. The frequency distributions of selected regions for each component of the first and second data sets were plotted in Figure 6.6 to 6.8. As can be seen from the figures, the frequency of the selected cell numbers correspond to selected wavelengths is significantly higher around the peak maximum of each component. This shows that the GILS method selects the regions, where the most concentration related information is contained.

As a result, it can be said that the GILS method can be used for the determination of diesel adulteration with sunflower oil, used frying oil, canola oil, and biodiesel using excitation-emission and synchoronous fluorescence spectra.

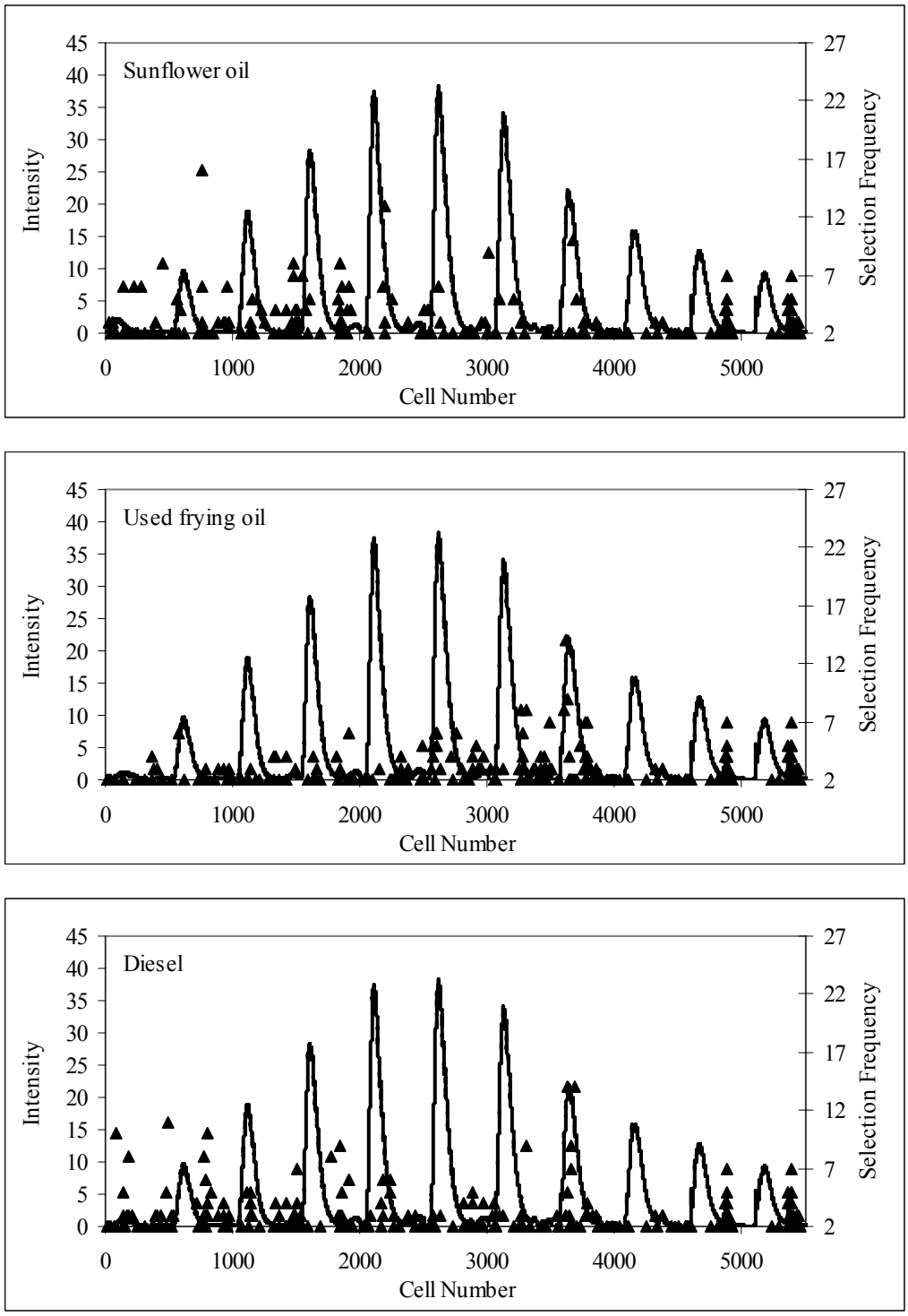

Figure 6.6. Frequency distribution of GILS selected wavelengths using excitation-emission fluoresecence data of the first set for sunflower oil, used frying oil, and diesel along with their concatenated pure component spectrum.

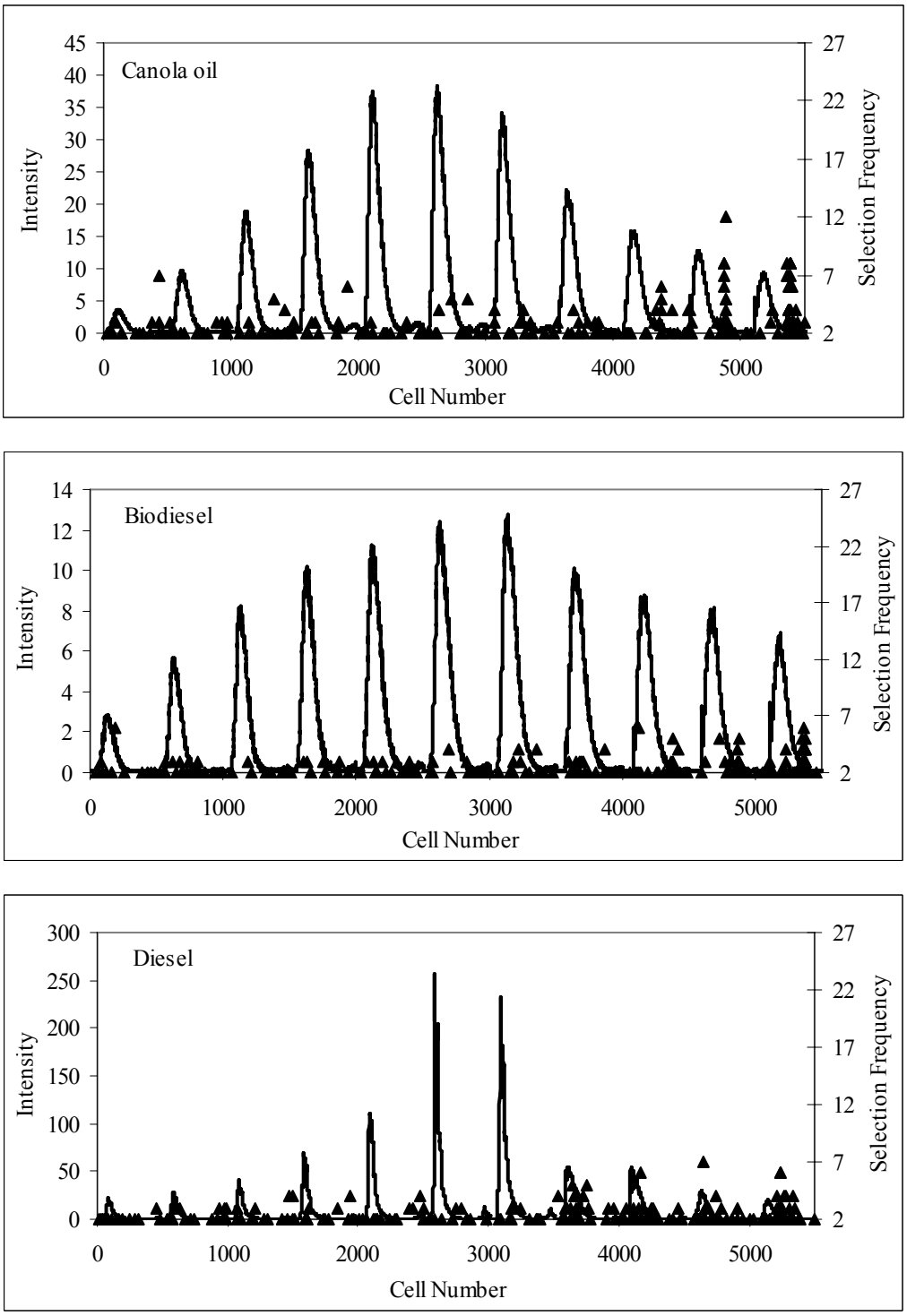

Figure 6.7. Frequency distribution of GILS selected wavelengths using excitation-emission fluoresecence data of the second set for canola oil, biodiesel, and diesel along with their concatenated pure component spectrum

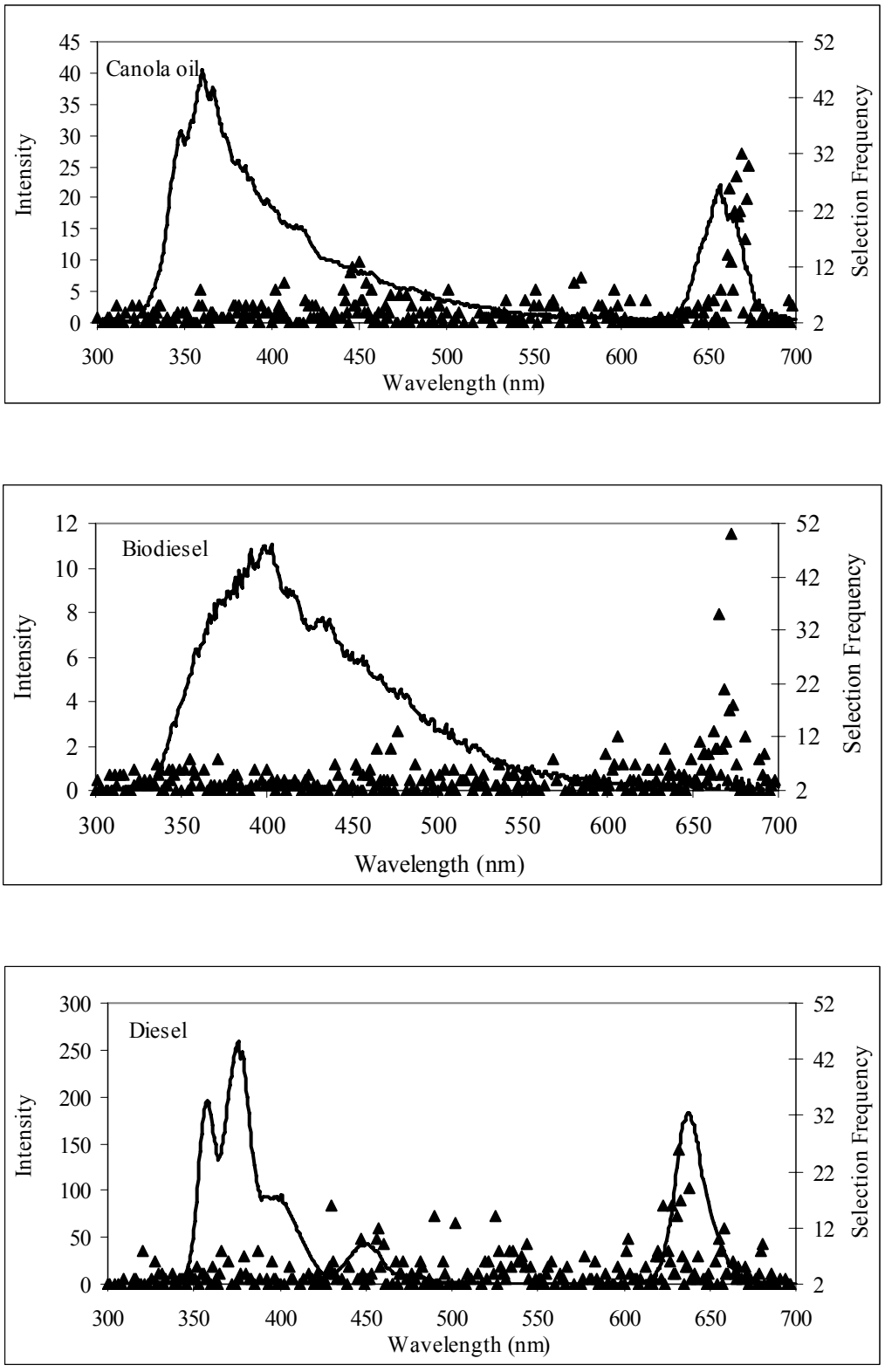

Figure 6.8. Frequency distribution of GILS selected wavelengths using synchronous fluoresecence data of the second set for canola oil, biodiesel, and diesel along with their concatenated pure component spectrum

CHAPTER 7

CONCLUSION

This study has demonstrated that the NIR, FTIR-ATR, and fluorescence spectroscopies with multivariate calibration method can be used for fast and simultaneous determination of diesel adulteration with several vegetable oils, used frying oil, kerosene, and biodiesel. In this sence, this study is very important in the presence of time consuming and expensive standard methods. In order to construct successful calibration models, genetic inverse least squares (GILS) was used as a multivariate calibration method. The validity and the performance of the models was determined by standart error of calibration (SEC) and standart error of prediction (SEP) values. The lowest SEC and SEP values, and also regression coefficient (R^2) values of actual vs. predicted component concentrations plots that are equal or closed to 1 indicated the best calibration model which includes the most quantitative information of the components. According to the results, all spectroscopic methods used in this study can be applied for the fast and non-destructive determination of diesel adulteration. On the other hand, the best results were obtained from FTIR-ATR spectroscopy combined to GILS when compared to the other methods.

REFERENCES

Abajo, M.J., Saiz, J.M. and Pizarro, C. 2006. Prediction of organic acids and other quality parameters of wine vinegar by near infrared spectroscopy. *Food Chemistry* 99:615-621

Agarwal, A.K. 2007. Biofuels applied as fuels. *Progress in Energy and Combustion Science* 33:233-271.

Aliske, M.A., Zagonel, G.F., Costa, B.J., Veiga, W., Saul, C.K. 2007. Measurement of biodiesel concentration in a diesel oil mixture. *Fuel* 86:1461-1464

Altun, Ş., Bulut, H. and Öner, C. 2008. The comparison of engine performance and exhaust emission characteristics of sesame oil-diesel fuel mixture with diesel fuel in a direct injection diesel engine. *Renewable Energy*. 33:1791-1795.

Banwell, C.N. and McCash, E. M. 1994. *Fundamentals of Molecular Spectroscopy*. New York:McGraw–Hill

Barton, F.E. 2002. Theory and Principles of Near Infrared Spectroscopy. *Spectroscopy Europe*. 14:12-19

Brereton, R.G. 2003. *Chemometrics:Data Analysis for the Laboratory and Chemical Plant*. England: John Wiley and Sons..

Burns, D.A. and Ciurczak, E.W. 2001. *Handbook of Near-infrared Analysis*. New York:CRC Press.

Carolei, L. and Gutz, I.G.R. 2005. Simultaneous determination of three surfactants an water in shampoo and liquid soap by ATR-FTIR. *Talanta*. 66:118-124.

Demirbas, A. 2007. Biodiesel from sunflower oil in supercritical methanol with calcium oxide. *Energy Conversion and Management*. 48:937-941

Divya, O. and Mishra, A.K. 2007. Combining synchoronus fluorescence spectroscopy with multivariate methods for the analysis of petrol-kerosene mixtures. *Talanta* 72:43-48.

Divya, O. and Mishra, A.K., 2007. Multivariate methods on the excitation emission matrix fluorescence spectroscopic data of diesel-kerosene mixtures: a comparative study. *Analytica Chimica Acta*. 22:9.

Font, R., Celstino, M. and Bailón, A. 2006. The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Industrial Crops and Products*. 24:307-313

Gayete, V., Guardia, M. and Garrigues, S. 2006. Attenuated total reflectance infrared determination of sodium nitrilotriacetate in alkaline liquid detergents. *Talanta* 70:870-875

Hernando, J., Leton, M.P., Novella, J., Builla, A. 2007. Biodiesel and FAME synthesis assisted by microwaves: Homogenous batch and flow processes. *Fuel*. 86:1641-1644.

Hollas, J.M. 2004. *Modern Spectroscopy*. New York:Wiley

Hsu, P.S. 1997. *Handbook of Instrumental Techniques for Analytical Chemistry*. Upper Saddle River: Prentice Hall

Huzayyin, A.S., Bawady, A.H., Rady, M.A., Dawood, A. 2004. Experimental evaluation of diesel engine performance and emission using blends of jojoba oil and diesel fuel. *Energy Convers Manage*. 45:2093–112

Ingle, J.D. and Crouch, S.R. 1988. *Spectrochemical Analysis*. New Jersey:Prentice Hall.

Knothe, G. 2001. Historical Perspectives on Vegetable Oil-Based Diesel Fuels. *Industrial Oils*. 12:1103-1107.

Knothe, G., Matheaus, C., Thomas, W. 2003. Cetane numbers of branched and straight-chain fatty esters determined in an ignition quality tester. *Fuel*. 82:971-975.

Kramer, R. 1998. *Chemometric Techniques for Quantitative Analysis*. New York:Marcel Dekker

Kwanchareon, P., Luengnaruemitchai, A. and Jai-In, S. 2007. Solubility of a diesel biodiesel ethanol-blend, its fuel properties and its emission characteristics from diesel engine. *Fuel*. 86:1053-1061

Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Kaufman, L. 1988. *Chemometrics: a text book*. New York:Elsevier.

McKelvy, M.L., Britt, T.R., Davis, B.L., Gillie, J.K., Lentz, L.A., Leugers, A., Nyquist, R.A.,and Putzig, C.L. 1996. Infrared Spectroscopy. *Anal. Chem*. 68:93-160

Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Kaufman, L. 1988. *Chemometrics: a text book*. New York:Elsevier.

McKelvy, M.L., Britt, T.R., Davis, B.L., Gillie, J.K., Lentz, L.A., Leugers, A., Nyquist, R.A., and Putzig, C.L. 1996. Infrared Spectroscopy. *Anal. Chem*. 68:93-160

Oliveira, C.C.F, Brandao, R.R.C, Ramalho, F.H., da Costa, A.F.L., Suares, A.Z.P., Rubim, C.J. 2007. Adulteration of diesel/biodiesel blends by vegetable oil as determined by Fourier transform (FT) near infrared spectrometry and FT-Raman spectroscopy. *Analytica Chimica Acta* 587:194-199.

Özdemir D. 2006. Genetic multivariate calibration for near infrared spectroscopic determination of protein, moisture, dry mass, hardness and other residues of wheat. *International Journal of Food Science and Technology* 41:12-21.

Özdemir, D., Dinç, E. 2005. Determination of benazepril HCl and hydrochlorothiazide in pharmaceutical preperations using UV-Visible spectrophotometry and

genetic multivariate calibration methods. *Journal of Food and Drug Analysis* 13:301-311.

Özdemir, D. And Öztürk, B. 2007. Near infrared spectroscopic determination of olive oil adulteration with sunflower and corn oil. *Journal of Food and Drug Analsis*. 15:40-47..

Pimentel, M.F., Riberio, M.G.S.G., Da Cruz, S.R., Stragevitch, L., Filho, J.G.A.P., Teixeira,L.S.G. 2006. Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. *Microchemical Journal* 82:201-206.

Pinto, A.C, Guarieiro, L.N., Rezende, J.C., Ednildo, A., Wilson, A., Pereira, P. and Jailson, B. 2005. Biodiesel: An overview. *J. Braz. Chem. Soc.* 16:1313-1330.

Poulli, K.I., Mousdis, G.A., Constantinos, A.G. 2006. Synchronous fluorescence spectroscopy for quantitative determination of virgin olive oil adulteration with sunflower. *Anal. Bioanal. Chem.* 386:1571-1575.

Pugazhvadivu, M., Jeyachandran, K. 2005. Investigations on the performance exhaust emissions of a diesel engine using preheated waste frying oil as fuel. *Renew Energy*. 30:2189–202.

Rakopoulos, C.D., Antonopoulos, K.A., Rakopoulos, D.C., Hountalas, D.T, Giakoumis, E.G. 2006. Comparative performance and emissions study of a direct injection diesel engine using blends of diesel fuel with vegetable oils or bio-diesels of various origins. *Energy Convers Manage*. 47:3272–87.

Reich, G. 2005. Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications. *Advanced Drug Delivery Rewiews*. 57:1109-1143.

Roy, S. 1999. Fiber optic for determining adulteration of petrol and diesel by kerosene. *Sensors and Actuators*. 55:212-216.

Skoog, D.A. and Leary, J.J. 1992. *Principles of Instrumental Analysis*. Florida:Harcourt Brace College Publishers

Smith, B.C. 1996. *Fundamentals of Fourier Transform Infrared Spectroscopy*. New York:CRC Press

Stavarache, C., Vinatoru, M. and Maeda, Y. 2007. Aspects of ultrasonically assisted transesterification of various vegetable oils with methanol. *Ultrasonics Sonochemistry* 14:380-386

Stuart, B.H. 2004. *Infrared Spectroscopy: Fundamentals and Applications*. England:John Wiley and Sons.

Taksande, A. and Hariharan, C. 2006. Synchronous fluorescence method to check adulteration of petrol and diesel by kerosene. *Spectroscopy Letters*. 39:345-356.

Valeur, B. 2002. *Molecular Fluorescence*. New York: Wiley

Vieira, P.A., Vieria, R.B., França F.P. and Cardoso, V.L. 2007. Biodegradation of effluent contaminated with diesel fuel and gasoline. *Journal of Hazardous Materials*. 140:52-59.

.

Wang, Y.D., Al-Shemmeri, T., Eames, P., McMullan, J., Hewitt, N., Huang,Y., An experimental investigation of the performance and gaseous exhaust emissions of a diesel engine using blends of a vegetable oil. *Appl Therm Eng*. 26:1684–91.

Wang, Y., Ou, S., Liu, P., Xue, F., Tang S. 2006. Comparision of two different process to synthesize biodiesel by waste cooking oil. *Journal of Molecular Catalysis*. 252:107-112.

Xixiong, Z., Jun, L., Zhihong, W. and Zhiyu, W. 2007. *The Application of Chemometric Methods in Near Infrared Spectroscopy Analysis*. Jilin:IEEE Conference Publishing